

Diversity of Experimental Methods in Economics: An Introduction



Toshiji Kawagoe and Hirokazu Takizawa

1 Economics and Experiments

Experimental methods have long been thought of as irrelevant and useless to economic research. To the best of our knowledge, this idea was explicitly stated for the first time by John Stuart Mill in the middle of the nineteenth century. In his treatise on economic methodology, he clearly referred to the impossibility of *experimentum crucis* in economics (Mill 2007/1836). For him, it is this fact that differentiates the methodology of economics from that of natural sciences. Thus, he proposed method a priori for economics that is mainly based on deduction from basic premises. The edifice of economics should be deductively built upon the premise that humans desire to possess wealth, and are capable of thinking of efficient means to that end, which itself is obtained through introspections.

In 1985 when plenty of experimental works had already been in currency, the same idea was expressed in the most celebrated textbook as follows:

Economists ... cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe (Samuelson and Nordhaus 1985, p. 8).

In retrospect, it is somewhat surprising to see the similarity of those ideas after a lapse in time of more than 100 years. In fact, similar ideas have been in currency, such as in Friedman (1953)'s methodology of positive economics.

The recent development of experimental methods in economics, however, shows that there has turned out to be some room for the improvement in the control of "other factors," and that economics has something, actually a lot, to learn from experiments.

T. Kawagoe
Future University Hakodate, Hokkaido, Japan
e-mail: kawagoe@fun.ac.jp

H. Takizawa (✉)
Chuo University, Hachioji, Tokyo, Japan
e-mail: hirokazu.takizawa@gmail.com

Mill, Samuelson and Nordhaus, among others, failed to see these points. It may be that the ideas on the nature of economics as well as of experiments have undergone a major revision among economists in recent years.

By now, experiments are accepted as a legitimate method in economics as exemplified by the awarding of Nobel Prize in economic sciences 2002 to Vernon Smith who “developed methods for laboratory experiments in economics, which has helped our understanding of economic behavior.” Nevertheless, the question of what the experiment is and what and how we learn from experiments are not yet so obvious, even if confined to the realm of natural sciences, and still invite new insights.

Our stereotypical image of experiments seems to have been formed by the experiment conducted by Galileo Galilei that led him to the discovery of the law of falling body (the experiment conducted at the Tower of Pisa seems now to be regarded as fictional, but the experiment with falling bodies along a slope as authentic). This stereotypical image arouses a feeling that there exists a strong tie between experiments and laws; laws are found inductively from experimental results. However, as far as current economic experiments are concerned, they do not seem to purport to discover laws. After all, in economics, we have few laws that are comparable with those in natural sciences, and it seems that economists today are more interested in uncovering and understanding mechanisms behind economic phenomena.

In spite of the wide recognition of the validity of experiments, how they are relevant to economic reasoning is not so straightforward. In fact, there seems to be a diversity in the ways that experimental methods are relevant to economics. With the main focus on laboratory experiments, one of the most notable practitioners in this field attempts to classify the purposes of economic experiments into the following categories (Roth 1986):

- (1) speaking to theorists (testing and modifying formal theories);
- (2) whispering into the ears of princes (providing information for the policy-making process); and
- (3) searching for facts (finding interesting phenomena).

Roth made this classification mainly with laboratory experiments in mind, but this classification seems to apply beyond the laboratory experiments. Thus, we will also rely on this classification as needed in what follows.

This book aims to show that the use of experimental methods depends deeply upon the reasoning researchers employ in their respective research. It also attempts to show the limits of experimental methods by examining in wider perspectives what experimenters are actually doing.

That said, this chapter aims to introduce readers to the subsequent chapters by delineating the overall picture concerning the experimental practices in economics, and then to provide information concerning each chapter so that the readers with specific interests may go directly to the relevant parts.

2 Vernon Smith's Market Experiment

It is difficult to identify the first attempt at the economic experiment, but it is increasingly becoming common to cite Edward Chamberlin's market experiment in 1948 as the one that has triggered the ensuing rise of experimental economics (Chamberlin 1948).¹ In fact, it is the participation in this experiment that provided a spark for Smith to embark on serious experimental research along his own research interests. Chamberlin's experiments were market experiments in a classroom, with which he tried to criticize the validity of the theory of perfectly competitive markets.

Chamberlin's innovation in this experiment was that he successfully created in the classroom an economic environment similar to the one supposed by the market theory, in this case, the partial equilibrium theory with discrete goods or service. For this, it would suffice that we can create buyers and sellers with arbitrary amounts of reservation value. For example, a buyer with value v for the good would be created, if his/her reward is set at $v - p$ when he/she buys the good at the price of p . A seller with the reservation value of c would be created if his/her reward is $p - c$ when he/she sells the good at the price of p . Thus, in an experiment, each participant individually receives information *ex ante* as to whether he/she is a buyer or seller and the amount of his/her own reservation value, and then instructed how the reward is determined. With buyers and sellers thus created, we can draw the demand and supply curves for the market and compute its competitive equilibrium price and quantity. We can then compare the theoretical prediction with experimental outcomes.

Vernon Smith's early experimental study also focused on markets.² However, unlike Chamberlin, his interest was not in refuting the theory of competitive market, but in exploring issues that cannot be dealt with by the mathematical models for this theory. The models of competitive equilibrium theory do not necessarily provide the description of how the competitive equilibrium is attained. However, an influential version presupposes a fictitious auctioneer, who works to search for prices that equate demand and supply for goods and services, a process known as *tatonnement*. Of course, there is usually no such a person in the real market. Therefore, it is worthwhile to explore the conditions that ensure rapid convergence to the theoretical equilibrium with actual human subject learning through time in experimental markets. Thus, the objective of Smith's experiments was not simply to test a theoretical hypothesis in the sense of confirming or refuting it, but to discover the way of organizing markets that achieves theoretically predicted outcomes in an efficient way. He then discovered the "double auction" method, now commonly used in classroom market experiments, as the way of organizing markets that closely simulates the prediction of the competitive market theory. This is a method of transaction where *both* sellers and buyers with their own reservation value, unbeknownst to other market parameters however, quote their asking prices to one another.

¹See Roth (1986) as well as Roth (1993). He cites E. Chamberlin, A. Hoggatt, H. Sauermann and R. Selten, M. Shubik, S. Siegel and L. Fouraker, and J. Friedman as independent early contributors to economic experiment. See also Svorenčik and Maas (2016).

²Smith's first supply and demand experiment was done in January 1956 (Smith 1989).

In general, any market outcome can be regarded as jointly produced by human subjects' behavior and the specific market institution surrounding them. Since the focus of Smith's experiment was on the performance of various market organizations, he wanted to make the behavioral aspect as constant as possible. Thus, with the competitive market theory in mind, he needed to make the preferences of experimental subjects as close to those supposed in the theory as possible. It was at this specific point that the traditional problem of controlling experimental environments arises.

Smith deliberately contrived his *induced value theory* in this context. Note that, in times when the axiomatic/deductive methodology was the orthodoxy, he had to fend off any possible criticism from those unfamiliar with experimentation in economics. The induced value theory aims to identify the right ways to create incentives the theory assumes in the laboratory, but usually cited as the following precepts for experimentation (Smith 1976, 1982):

- (1) non-satiation: making subjects always choose the alternative with the most reward;
- (2) saliency: relating rewards appropriately to experimental outcomes;
- (3) dominance: making the reward structure dominate any other subjective value/cost that may affect a subject's choice;
- (4) privacy: subjects are only informed of their own payoffs; and
- (5) parallelism: the choice tendency observed in the experiment also holds in outside environments.

According to this theory, reward system has to be structured so that subjects have clear-cut and sufficient incentives in making their choices in the experiment, as presupposed by the theory to be tested. This is ensured by the arrangement where the points they obtained in the experiment are converted in monetary units and paid in cash immediately after the experiment. This is regarded as the main reason that the experiments reported in published papers have usually followed this practice.³

3 The Rise of Behavioral Economics

It is indispensable to mention, at this point, the rise of behavioral economics since the 1970s, which also makes use of experimental methods, and to touch upon its relation to the experimental economics advanced along the line of Vernon Smith.

Behavioral economics analyzes actual human behavior and explores its implications for understanding economic phenomena. This area of study has already produced two Nobel laureates, Daniel Kahneman in 2001 (jointly with Vernon Smith) and Richard Thaler in 2017. Given that traditional economics generally used to

³It is interesting to know that historically this practice comes from a laboratory experiment for psychological learning by Siegel (1953). See Svorenčik and Maas (2016, p. 87) for this.

assume that human agents are rational and usually selfish as well, the rise of behavioral economics may be regarded as marking a fundamental transformation of the nature of economic sciences in the recent decades.

It is important to note that the research agenda of behavioral economics greatly differ from experimental economics as aforementioned. As the adjective “behavioral” indicates, it is focused not only on actions but rather also on behaviors in general.⁴ This means that it is closely connected to cognitive sciences and/or experimental psychology. In fact, Kahneman was a psychologist when he started his academic career in the 1960s. Recall that Smith was focused on the performance of the market institutions. The difference in focus also implied the difference in the specific methods used. In fact, it was survey questionnaires that Kahneman and Tversky used when they elaborated upon the theory of heuristics and bias, and its subsequent elaboration of the prospect theory (Kahnemann and Tversky 1979).

Heukelom (2014) describes the process, where psychological research by Kahneman and Tversky was gradually acquired by economists such as Richard Thaler in the 1980s. The process required a deliberative “marketing” strategy to establish the field within economics that had had very conservative methodological thinking. Thaler, together with Kahneman and Eric Wanner, set up research programs funded by Alfred P. Sloan Foundation as well as Russel Sage Foundation, carefully organizing the members so that the projects include almost equal numbers of economists and psychologists. In the setting-up process, Smith was also invited, but he did not participate in the project, possibly due to the difference of philosophy.⁵

By now the differences seem to be fading, and the border between the two fields blurring. This may be partly because game theoretic experiments have become increasingly common since the 1990s, and they tend to be more focused on behavioral anomalies observed in social interactions. However, it should be noted here that different research motivations have developed different experimental methods.

4 Behavioral Economics, Neuroeconomics and Naturalism

The rise of behavioral economics has had enormous impacts on economics. For one thing, it revolutionized economics in that it extended the traditional scope of economics to the analysis of the real human behavior. For another, it has made economics an interdisciplinary field of research by introducing research methods that had been adopted outside economics such as psychology. Let us explain these points in turn.

⁴According to Heukelom (2014), it was in the US in the early 20th century that the word “behavior” began to be used for comprehensive activities not only of human beings but also of animals.

⁵Smith seems to be somewhat critical against behavioural economists’ search for anomalies “in the tails of distributions.” See Smith (2008, p. 22, ff. 6).

The structure of traditional economics is basically deductively constructed. Setting several plausible axioms on the behavior of economic agents, economists usually deduce theorems/propositions useful for understanding the working of an economy. As we saw in Mill's statement at the beginning of this chapter, this fundamental character of economics partly followed from the presupposition that conclusive experiments are impossible in economics. The elegant mathematical edifice of competitive equilibrium theory built in this approach has long made most economists refrain from doubting the very foundation of this theoretical construct. From this point of view, the rise of behavioral economics was revolutionary in that it began to focus on the human behavior as such, which constituted the foundation of the whole system of economics.

The rise of behavioral economics also means that economics has now imported different methodological strategies from other disciplines. It is well known that the mainstream economics opted for separating itself from psychology as it establishes itself as an autonomous scientific endeavor in the first half of the twentieth century. This is best exemplified by the statement in Robbins (1932) that economics will not get involved in psychological issues of human agents. This has enabled economics to exclusively engage in the study of resource allocations mainly through the market mechanisms.

Viewed from the historical perspective, however, there have been two distinct approaches to the explanation of human mind and behavior. One approach tries to understand and explain human behavior in terms of such intentional states as preference and belief.⁶ This approach to human behavior is not so greatly different from the folk psychology we use in our daily life; we usually justify our behavior by explaining our own preference and belief, and attribute them to understand someone else's behavioral choice.⁷ The traditional economics has also adopted this approach. The other approach to the human behavior is naturalistic in the sense that it attempts to understand it in terms of the causal relationship as usually practiced in physical sciences, explicitly excluding teleological elements from the explanation (von Wright 2004).

Human mind and behavior are unique in that they can be subject to analysis under both approaches. If one of the two approaches were reducible to the other, there is no essentially difficult problem. However, there is an influential argument by Davidson that there cannot be laws connecting the mental states and physical states, implying the fundamental incommensurability of the two approaches (Davidson 1980).

⁶Here the term "intentionality" should be taken as a philosophical terminology referring to the characteristic of some mental states that they are always "about" something, and intentional states are mental states with such a property. Intentional states include such mental states as perceiving, believing, wanting, hoping and so on.

⁷However, this way of understanding human behaviour has been philosophically sophisticated by some philosophers and social scientists including Max Weber. In their discourse, an action is distinguished from a sheer behaviour such as yawning, because it is based on intentional states and thus regarded as rational choice of an agent. Weber (1978) restricted the object of sociology to the study of actions with rational motives, because we can only understand actions with rationality. Of course, Weber's sociology includes the discipline of economics in our modern parlance.

It seems that both approaches are dispersed within the field of behavioral economics. For example, the prospect theory proposed by Kahneman and Tversky explains the choice behavior under risk by combining preference and belief, although both preference and belief used there are different from such standard ones in the expected utility theory; in the prospect theory, they are both ridden with considerable biases. Despite these differences in the formulation of preferences and beliefs, however, we may be able to regard the prospect theory in line with the intentional approach. In contrast, most of the economic experiments presented in Ariely (2008) utilize “priming,” a technique frequently used in psychological experiments, where exposure to one stimulus influences a response to a subsequent stimulus without consciousness, to find out a causal relationship without reference to any mental states.

The naturalistic approach to human mind/behavior is even more obvious by the rise of neuroeconomics since the 1990s. This period saw a rapid development of technologies to measure brain activities in non-invasive ways, such as the functional magnetic resonance imaging (fMRI). With this, some researchers began to investigate which parts of the subjects’ brain are activated when they make economic decisions. Most, if not all, of such research has shared the question formulated in behavioral economics. In this sense, neuroeconomics can be regarded as part of behavioral economics, pursuing the same subject matter with different tools.

The emergence of the naturalistic approach gave rise to a controversy in the methodology of economics in the early 2000s. The debate was ignited by Gul and Pesendorfer’s (2008) argument that neuro-scientific evidence is irrelevant to economic research. Economics is, they assert, essentially a science of making choices that enables us to understand how people make choices, for which such categories as preference, belief and constraint are relevant and neuro-scientific data does not provide any useful information. In the same volume, Schotter (2008), in contrast, argues that neuro-physiological data can be useful to economic modelling in some instances. As an example, he raises the interpretation problem regarding the outcomes observed in laboratory experiments of first-price, sealed-bid auction. It has been known that, in such experiments, subjects tend to submit higher bids than the theoretical prediction. To explain the phenomenon, theorists may have two options. One strategy is to adopt a behavioral model where human subjects want to feel the “joy of winning.” The other is to assume that humans want to avoid the “fear of losing.” The physiological data might help us identify which is the right underlying cause for the observed behavior.

The meaning of the rise of naturalism into economics is a very interesting question as such. However, back to our current context, it is interesting to observe that experimental methods are not only related to the so-called “naturalistic” investigation of the causal relationship. It has been also useful in research programs that assume that human decision-making is based on intentional states.

5 Behavioral Game Theory

It is well known that game theoretic experiments were already conducted in 1950, soon after the theory had been launched by Von Neumann and Morgenstern; Melvin Dresher and Meryl Flood experimented with a game that would be later known as “prisoner’s dilemma” at RAND Corporation (Flood 1958). The aim of the experiment was to compare the performance of alternative solution concepts for non-cooperative games, including Nash equilibrium. They reported that the experimental results did not necessarily support the prediction of Nash equilibrium. However, John Nash himself gave a critical comment that their experimental design was such that subjects had been actually playing a repeated game, rather than a one-shot game. This is an interesting episode that tells us about the importance of experimental designs.

As game theory became increasingly common among economists in the 1980s, economists soon became interested in conducting game theoretic experiments. This may be partly because it is relatively easy to conduct a game experiment, if casually, in a classroom. It should be noted, however, that game theoretic experiments require a somewhat different consideration than the market experiments. Since, in game theory, common knowledge of the rule of the game and rationality of players is usually assumed, game experiments need to realize that assumption in a laboratory. Recall that the above-described induced value theory included “privacy” as one of the sufficient, if not necessary, conditions for a good experiment. Obviously game experiments, if they are to test a theory, have to ignore this maxim.

Nevertheless, except the privacy requirement, game theoretic experiments have been conducted mostly following the norms of induced value theory. However, their role in game experiments seems to be somewhat different from that in the market experiments. Whereas the market experiments were mainly focused on the performance of various ways of organizing markets, game experiments came to be increasingly driven by the discovery of various anomalous behaviors in game-theoretic social interaction. This is the reason why Colin Camerer coined the word “behavioral game theory” (Camerer 2003). In this context, strict compliance with the standard norm was necessary for researchers to clearly identify deviation from theories.

Examples of anomalies are abundant, but let it suffice to mention one-shot prisoner’s dilemma, public goods game and ultimatum game among others. In these games, subjects seem to be affected by consideration of social contexts in which they usually live, even though they are given strong appropriate incentive in the laboratory environments. This naturally led to the proposal of a variety of social preferences or other-regarding preferences, such as altruism, fairness concern, inequality aversion, as well as behavioral model based on bounded rationality such as quantal response equilibrium and level-k model. To date, however, the dispute over which has the most explanatory power has not been resolved. Rather, models with most explanatory power seem to depend on specific contexts, such as whether the situation considered involves a distribution problem. Researchers in this subfield seem to rely on the model comparison method developed in statistics.

Game theory also launched a subfield called market design, which aims to design artificial markets that have not existed in history. Laboratory experiments are also utilized in this ambitious enterprise, as best exemplified by the process, where FCC introduced auctions for the allocation of the microwave spectrum. Newly proposed market mechanisms are usually examined in laboratory experiments before being implemented in reality. For example, experimental results here can offer information as to whether players can “game” the rules in an unexpected manner. The purpose of experimentation here is “whispering into the ears of princes.”

6 From Testing Theories to Informing Policies

So far we have mostly seen laboratory experiments, which have the obvious advantage that researchers can make experimental environments close to those supposed by theoretical models. Actually, laboratory experiments, in general, do not necessarily aim to predict what will happen in the real world. Rather they aim to find answers to the question posed by theoretical models. Recall that Smith’s question was under what conditions human behaviors well simulate the competitive equilibrium theory, and game theoretic experiments are meaningful only with regard to theoretical predictions. Even if we know that subjects divide pies on the 50–50 basis in the ultimatum game, it will not attract people’s interest unless we also know that theory predicts the proposer’s taking almost all the pie. Thus, laboratory experiments are deeply involved with theories, and aim to “speaking to theorists,” or “search for facts” to be explored by theories, to use Roth’s classification.

Then how can we “whisper into the ears of princes”? Intuitively speaking, for this kind of purposes, it seems that we need to get out of laboratories and get closer to the real-world environments where policies are implemented. Furthermore, it does not usually suffice that we only have information about the correlation between relevant variables. It may be that increasing money supply does not cause inflation, even if we know the correlation between them is very strong. Thus, in order to make an effective policy, we need information regarding causality rather than correlation.

In the early 1990s, there emerged the movement of “evidence based medicine” in the UK, which asserts that medical treatments should become more systematic and scientific, relying less on “soft” evidence such as doctors’ experiences and gut feelings and more on “hard” evidence. This idea was soon generalized to “evidenced based policies” in the UK and USA, in such areas as social care (the UK Sure Start program) and education policy (No Child Left Behind Policy, USA). It also permeated economics very soon, especially in development economics. The movement regards randomized controlled trial (RCT) as the “gold standard” for obtaining “hard” evidence.

In a typical RCT, all subjects are randomly divided into two groups. The reason that the assignment is random is to make the two groups have identical characteristics from a statistical point of view. One group, called “treatment group,” receives treatment, whereas the other group, called “control group,” either remains untreated,

receives alternative treatment, or is given a placebo, depending on the purpose of the experiment. In order to make the experiment more rigorous, a double-blind method may be adopted, where not only the subjects but also the experimental administrators are given no information as to which is the control group and which the treatment group. This is to avoid the “experimenter effect,” whereby the expectation of experimenters regarding the effect of treatment may unknowingly affect the behavior of subjects.

This method dates back to Fisher (1935), but is obviously reminiscent of the “Canon of Inductive Method” that John Stuart Mill crystalized as reliable forms of induction, especially “Method of Difference.” Suppose that there are two situations, one of which has some particular phenomenon and the other of which does not. If the two situations are alike except in some other specific factor, then it is judged as an effect or a cause of the phenomenon. The RCT today differs from this method only in that it tries to identify the difference in the effect of a treatment by using modern statistical inferences.

One of the greatest advantages of an ideally implemented RCT seems to lie in the general applicability of its causal inference. Researchers conducting RCTs do not need to have detailed knowledge on the mechanism that causally connects treatments and their effect, although they will have to have specific hypotheses *ex ante* to design the experiment. In contrast, the use of economic models usually requires us to have detailed knowledge about corresponding mechanisms, but the models as such often rely on dubious assumptions. There are no such worries involved in the implementation of RCTs, because the mechanism is left as a black box as it were.

This methodology, brought into development economics by Abhijit Banerjee, Esther Duflo and others, has achieved significant results in policy-making in developing countries. The introduction of RCT turned the nature of research in development economics from the armchair study that examines the effectiveness of development aid dealing with data econometrically to a kind of research that collects data in the field to investigate the effectiveness of specific policies.

However, even such a powerful method is not without criticism. First, there is the problem of “external validity” (Guala 2005), which questions whether the results obtained in some specific experimental environments also hold in other environments. As this is almost equal to what Smith called “parallelism” in his precepts, external validity is not limited to RCTs, but also problematic in laboratory experiments. However, the problem may become more serious in the context of RCTs, because we might have almost no information as to the causal mechanisms that are working to bring about the obtained results. Thus, we may not be left with enough clues for inferring the possibility that the same results obtain in other contexts.

The second criticism concerns the possible existence of “general equilibrium effects.” They may arise when coverage of the policy considered is extended to wider public on a greater scale so that the general implementation of the policy may greatly change the use of other resources in society. The policy might have been effective in a small-scale experiment, because of the *ceteris paribus* conditions that did not affect greatly the use of other resources. As an example, suppose the famous experiment conducted in the USA that examined the effect of a class size

on the effectiveness of education. The result was that a smaller class size has better educational outcomes. Would this result continue to hold if we made all the classes in the country smaller? The problem is that we may not be able to keep the quality of teachers constant, if we extend the program to the whole country.

7 Field Experiments and the Identification of Causality

The idea of exploiting high-quality causal information from the RCTs was not confined in development economics, but today is extended to wider areas in economics. As one can easily presume, it is not so easy to conduct RCTs in advanced countries. Nevertheless, there are some situations where we can conduct field experiments even in advanced countries. Furthermore, we sometimes find some contingent situations comparable to ideal experiments in the real-world settings, which is usually called natural experiments. This extension of the basic idea has been made possible by the innovation that occurred in statistics and is today called “statistical causal inference.”

Readers might have been taught in an introductory statistics lecture that correlation does not usually imply causal relations, presented with a graph with the values of height and weight taken along horizontal and vertical axes, respectively, in which each subject’s data is plotted. Here, obviously the height of a person does not causally determine his/her weight. There is a famous dictum that, in order to examine causality with respect to any system, it is not sufficient to passively observe the data occurring in the system, but it is necessary to intervene into the system to observe the outcomes. The former type of data is called “naturally occurring data,” whereas the latter “experimental data.” We used to think that it was difficult to grasp causality dealing with statistical data, because we used to have naturally occurring data in mind. However, in the 1970s, innovative statistical research was launched that aims to develop methods for identifying causal relations even using naturally occurring data by regarding experimental data as a benchmark for the identification of causality, Rubin’s causal model (Rubin 1974).

In general, to extract causality, we need to compare the outcome when a treatment was applied to an individual/group with the other outcome when it was not. In reality, however, we can only observe either of these. This is the conundrum for identifying causality, called “fundamental problem of causal inference” (Holland 1986). By focusing on causal effects, RCT makes it possible to extract causality by an experimental intervention that randomly assigns subjects to the treatment group, to which a treatment is applied, and the control group, to which it is not, making both groups have the same characteristics from the statistical viewpoint. The point is that, viewed as a random variable, the group assignment is independent of the treatment. This procedure can be regarded as a way to create the counterfactual to measure the causal effect. Although we do not delve into the details here, there are also ways to create the counterfactuals using naturally occurring data. Examples include instrumental variable method, propensity score matching, difference in differences method among others.

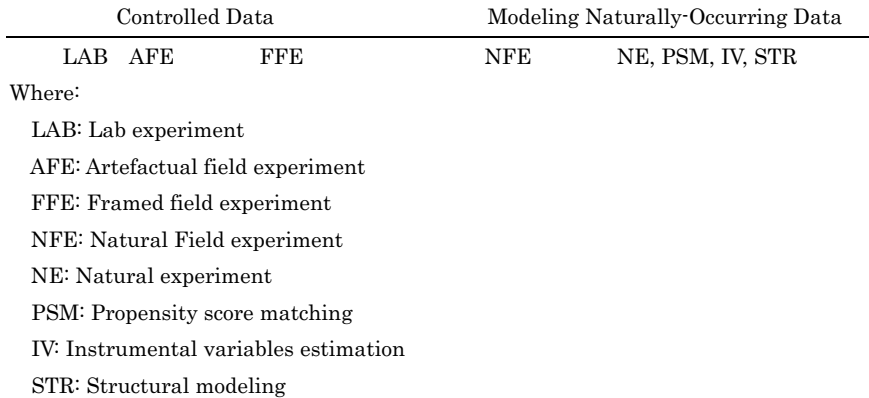


Fig. 1 A field experiment bridge (List 2006, p. 7)

Saying that “the goal of any evaluation method is to construct the proper counterfactual,” List (2006) thus presents a unified framework that enables us to understand almost all experiments in economics with different degree of control. See Fig. 1.

The horizontal segment in the diagram indicates the degree to which the environments are controlled, with the laboratory experiment at the left endpoint, and natural experiment, propensity score matching, instrumental variable estimation, structural modelling, which deal with naturally occurring data, at the right endpoint. Various kinds of field experiments are located between them. The artefactual field experiment is almost the same as the laboratory experiment, but the group of experimental subjects is set closer to people that the researcher is interested in. The framed field experiment is controlled to the same degree as the laboratory experiment, but the goods traded and/or the information available to the subjects in the experiment are closer to the real world. The experimental environments in the natural experiment are real world, but the subjects are randomly assigned as in the RCT. According to this definition, the RCTs in development economics goes into this category.

According to List, field experiments can function as a bridge between laboratory experiments and the real world, with regard to external validity. That is, the field experiments enable us to check whether the result obtained in laboratory experiments also hold in settings closer to the real world, because they have intermediate characters. For example, the well-known “endowment effects,” the phenomenon that people’s Willingness to Accept usually exceeds their Willing to Pay (Kahneman et al. 1990), have been shown to dissipate as the exposure of subjects to market experience increases, by a series of field experiments.

8 Looking at Various Experiments as They Arise

Thus, the framework List proposes is very ambitious in that it tries to grasp the very nature of various economic experiments in a unified approach. However, it may not be the whole story. The picture submitted there induces an image that all experimental research in economics share the same motivation, that is, the motivation to identify causality. However, reflecting back on the various ways that experimentation has brought forth new knowledge in social sciences, we may also say that there are diverse motivations for adopting experimental methods.

The nature of a specific problem at hand seems to mostly determine the experimental method to be used. As aforementioned, the laboratory experiment seems to be more related to the examination of specific mechanisms, which we usually express by means of economic models. In contrast, the RCTs are more concerned with identifying causal relationships that are useful for policy-making issues (“whispering to the ear of princes”) such as in development economics.⁸ Of course, both problems are interdependent on one another, because the knowledge of economics comprises of a complicated network of related hypotheses and models; theories that greatly affect our view of the real world would constrain the hypotheses to be tested.

We will not delve into this issue in this introduction. We hope that the above explanation of the development of various experimental methods sets the ground for the readers to proceed to the subsequent chapters.

9 The Structure of the Book

The book is divided into two parts. Part I: *Diversity in Experimental Methods* and Part II: *Critical Viewpoints*. Let us now give a brief overview of each chapter so that you can go directly to the chapter that you are interested in.

Part I: Diversity in Experimental Methods

Part I mainly describes how various experimental methods have been practiced in such diverse fields as market theory, game theory, development economics, political science, behavioral analysis and evolutionary psychology. These chapters also suggest how it would become possible to have a constructive dialog between diverse fields based on the common language of experimentation.

Chapter 1: Laboratory Experiment in Game Theory

This chapter introduces readers to laboratory experiments in game theory. It begins by briefly reviewing its history. Although one may think of Vernon Smith to hear experimental economics, the experimental game theory has its own origins different

⁸This assertion may not be immediately applicable when we consider designing the actual economic environment by simulating the environment in laboratory experiments.

from his market experiments. With the permeation of game theory into economics in 1990s, experiments on game theory began to increase in number, providing fertile bedrock for economic experimentation.

In this process, experimenters had to slightly change the basic precepts provided by Smith's induced value theory; They had to "ignore" privacy precept, which requires subjects be only informed of their own payoffs, because common knowledge of the rule of the game is usually assumed in game theory. Furthermore, controlling the preferences of subjects gradually came to take on different meanings in game experiments. As the experimental results accumulate, it turned out that subjects do not necessarily behave rationally. In this context however, it was important for experimenters to show that anomaly occurs *in spite of* rigorous controlling procedure of the experimental environments.

With the accumulation of anomalies, the game theory experiment has transformed itself from a tool for testing theory to that for identifying subjects' behavioral model. The behavioral models so far proposed include models based on other-regarding preferences and bounded rationality. The latter half of this chapter explains those models and proceed to ways to compare competing models by using experimental data. One key concept is that of "statistical model" which involves at least one parameter to maximize likelihood function. Experimenters often need to transform a specific theory that provide only point prediction to a statistical model. The other is the criterion of model comparison. AIC and its variants are introduced as a useful tool for model comparison. Unlike other statistical methods, they do not presuppose which hypothesis is true.

Chapter 2: The Field Experiment Revolution in Development Economics

Krugman, a Nobel laureate, stated a quarter century ago that the field of development economics "no longer exists." Beginning with this shocking statement, the chapter describes how the field has been resuscitated with the use of experimental methodology since the turn of the twenty-first century. The key factor to this process has been the use of randomized controlled trials (RCT) in the field, which enable us to examine the causal effect of a policy intervention. The authors not only explain the very basic ideas of RCT and its fruitful applications, but also go further to describe how RCT-based research is enriched by combination with other experimental and/or statistical methods, such as the lab-in-the-field experiment and the structural estimation approach. The reader should feel that a great transformation of economic science is happening now, something that might be properly called "empirical turn" of economics. The authors are aware of this and associate the current development in this field to the general tendency of current economic research, "empiricalization," "scientification" and "team-orientation." The chapter is also rife with informative cases where new experimental methods have come to be used in response to research questions that arise based on the previous experimental research.

Chapter 3: Experimental Research in Political Science

Unsurprisingly, the history of experimentation in political science is almost as old as that of economics, dating back to Gosnell's experiment in the stimulation of voting

in 1926. However, it is not until the 1990s that an increasing number of experimental studies began to be published in major American journals. The experimental methodology in politics seems to have invited criticism from those who question how helpful it can be in understanding complex political phenomena. However, the trend has been changing, with the widespread recognition of the usefulness of experimentation.

The author stresses the importance of employing appropriate experimental methods for appropriate research topics. This belief is reflected in the first part of the paper, which surveys advantages and disadvantages of various experimental methods, ranging from field, laboratory, survey and natural experiments and simulation. Thus, this part will serve as a fine introduction of the experimental toolbox to anyone interested in experimental methods. In response to the complex character of the political phenomena, the author advises researchers to break the phenomena down to make them amenable to experimentation, and to integrate the results later.

The chapter also contains the author's own experimental research on people's ideas on distributive justice. This example illustrates how illuminating it can be to combine different experimental methods as well as the importance of deliberately choosing the experimental subjects so as to represent the whole society in the survey experiment.

Chapter 4: Experiments in Psychology: Current Issues in Irrational Choice Behavior

This chapter constitutes, as such, a complete guide to experimental methods in psychology. Readers who consider conducting experiments in this field would find the content of the chapter very useful and informative. The chapter also presents the state-of-the-art research results in behavioral analytic studies, and clearly identifies several research themes that economists and psychologists may profitably pursue in collaboration with each other.

The experiments in behavioral analysis usually use animals as experimental subjects. This means that, unlike economic experiments, the experimenter cannot rely on intentional concepts such as preferences and belief. This might be the fundamental difference between economic experiments and psychological experiments. Thus, there seems to be a hiatus between two subject fields. However, several findings in behavioral analysis, such as in the study of time discounting, have been successfully transferred to behavioral economics. It is interesting to note that the findings are strikingly similar in the two fields in spite of the difference of basic conceptual structure. Knowing how this was made possible would enable us to make further advancement not only in economics and psychology, but also in social sciences in general.

Chapter 5: Evolutionary Psychology and Economic Game Experiments

Chapter 5 examines the relationship between evolutionary psychology and experiments. Evolutionary psychology is a discipline devoted to understanding the human mind in terms of the theory of natural selection. Regarding human mind as a product of adaptation, evolutionary biologists assume that people behave so as to maximize fitness (survival and reproduction) under environmental constraints. Such environmental constraints are often modelled using economic games, such as prisoner's

dilemma, enabling evolutionary biologists to use evolutionary game theory, a branch of game theory developed by evolutionary biologists. This creates some similarity in reasoning between economics and evolutionary psychology.

The authors, however, note that there are also important differences. Evolutionary psychologists do not assume that people rely on rational deliberations, whereas economists usually do so. As a result, they may disagree in their interpretation of the same behavioral observation. Evolutionary psychologists are more concerned about the underlying mechanisms yielding particular behavioral responses. Simply observing behavior is not sufficient. Examining both cases where economic game experiments are informative and uninformative, the authors suggest that experiments can often select from among competing hypotheses about the underlying mechanism when they yield the same observed behavior.

Part II: Critical Viewpoints

Part II goes deeper into the experimental methodology in social sciences. It deals with such questions as follows. Is it possible to design a rewarding system that meets all the conditions of the induced value theory? Are RCTs really worth huge investment? Do economic experiments provide us with objective properties of human behavior? Does the performativity of economic experiments make them as useless as they were once supposed to be? How is the “experimental turn” related to the current overall changes in economics?

Chapter 6: Reconsidering Induced Value Theory

As described in Chap. 1, the *induced value theory* developed by Vernon Smith has had enormous impacts on the design of economic experiments until now. It has long been supposed that complying with the precepts in the theory is mandatory, if experimenters want to control experimental subjects’ preferences. However, there are several reasons to reconsider the utility of the theory.

Chapter 6 deals with this issue. In the first half of the chapter, the author identifies three approaches to economic experimentation that differ in their aim in conducting experiments, and considers each approach’s relation to induced value theory. The approaches identified there are neoclassical approach, “old school” of behavioral economics, and “new school” of behavioral economics.

The argument proceeds by formulating the basic common structure of an economic model, which can be thought of as a mapping that takes the agent’s (1) utility function and (2) decision environment as *explanans* to yield (3) observed behavior as *explanandum*. The functional form of this mapping is also important, and may be called a “principle of decision making.” The neoclassical economists assume that the principle of decision-making is utility maximization and the utility is the canonical expected utility. Therefore, for them, observed differences in behavior should be explained by some differences in environments. New school in behavioral economics shares with the neoclassical approach the assumption that the principle of decision-making is maximization, but it tries to explain the observed differences in behavior with reference to the differences in the utility function. Finally, the old

school of behavioral economics tries to attribute the observed behavioral difference to different principles of decision-making. From this perspective, the author asserts that controlling preference in the sense of the induced value only matters for the neoclassical approach and old school of behavioral economics, because these two approaches assume the subjects' utility is fixed. For the new school of behavioral economics, the important thing is not to control the preferences, but to identify them.

The simple belief in the induced value theory has recently been shaken by debates among experimental economists as well. The "all-pay system," where all subjects are rewarded for all tasks they fulfilled in an experiment, has long been regarded as the gold standard. However, this system is vulnerable to "wealth effect," i.e., subjects who have eared enough in the early rounds of an experiment might lose interest in later tasks. As a more robust rewarding system, some experimentalists proposed to adopt "random payment system," where payment is made for randomly selected tasks and/or subjects. However, this system is also known to be problematic in some experimental context. Then, we are naturally led to asking if there is any rewarding system that incentivizes subjects in an undistorted manner in any context. Citing recent works on this subject, the author says the answer is "No," as long as the experiment involves more than one round of decision-making.

Chapter 7: Billions of Dollars Worth of Experiments: Calibrating Clinical Trial Investments

As Chap. 2 indicates, in spite of its known limitations, RCTs still provide the strongest method for experimentation, especially in the context of identifying causal factors. Some may even say that RCTs are "gold standard" of experimentation. However, conducting RCT, especially in clinical trials, requires a huge investment. Thus, we may ask whether this social movement, the so-called evidence-based policy, is really worthwhile to continue to pursue. In this chapter, the author tries to calibrate the total clinical trial investments in North America, relying on the most recent studies that estimate the average cost of a single clinical trial. The results, presented with four scenarios, are a stunningly huge amount. Based on this research, the author is going to embark on a more challenging task to put these figures in the context of comprehensive cost-benefit analysis.

Chapter 8: New Wines into Old Wineskins? *Methodenstreit*, Agency and Structure in the Philosophy of Experimental Economics

Admitting that experiments have become part and parcel of today's economics, the author tries to put this trend in the context of the history of economics. It is well known that the methodological controversies on economic science have revolved around the question, whether it should basically be a deductive or an inductive science. The most famous example is the *Methodenstreit* that took place at the turn of the 19th and 20th century between the German Historical school and the Austrian school. While the latter emphasized the universal character of economics, the former regarded economics as a historical science dealing with particular values and cultures. The author argues that current economics is becoming more local and ad hoc, and more

empirical, implying some affinity with the idea of Historical school. He also notes that the current rise of experimental economics fits into this general picture, as he stresses the local and context-dependent nature of the knowledge obtained by means of economic experimentation.

Chapter 9: Creating Social Ontology: On the Performative Nature of Economic Experiments

The conventional view on economic experiments has been that the use of experimental methods enables experimenters to identify objective properties of human behavior, independent from the experimental settings. However, this view is wrong, because economic experiments are *performative* in the sense that subjects, experimenter and experimental designs are inextricably entangled with one another. The author especially focuses on the use of monetary incentives in the laboratory experiments, which most economists have regarded as the only useful device for controlling preference and as contributing to the de-contextualizing of experimental results. He argues that the use of money, in fact, constrains the universalization of laboratory results, because it is a very strong “priming” procedure from the perspective of psychology. However, his insight into the performative nature of experiments does not lead to a negative view of experimentation in economics.

At this juncture, relying on Karen Barad, he turns to Niels Bohr’s view on the controversy about the Copenhagen interpretation of quantum mechanics. Bohr suggested that the fundamental ontological units are phenomena that realize in an experimental setting. According to this view, there is nothing such as an immutable and fully autonomous object that could be separated from the phenomena emerging in the laboratory. The author argues that this also applies to economic experiments. Viewed in this manner, the results obtained in economic experiments are regarded as capturing reality in the sense that under similar environmental conditions, people will manifest similar property. Thus, it is meaningless to ask whether experiments can finally provide evidence on which kind of preferences human beings have in general. Economic experiments can identify certain performative mechanisms that generate a particular kind of preference in a particular context.

References

- Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York: HarperCollins.
- Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Chamberlin, E. H. (1948). An experimental imperfect market. *The Journal of Political Economy*, 56, 95–108.
- Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.
- Fisher, R. (1935/1971). *The design of experiments* (9th ed.). London: Macmillan.
- Flood, M. (1958). Some experimental games. *Management Science*, 5, 5–26.

- Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics* (pp. 3–43). Chicago, IL: Chicago University Press.
- Guala, F. (2005). *The methodology of experimental economics*. Cambridge; New York: Cambridge University Press.
- Gul, F., & Pesendorfer, W. (2008). The case for mindless economics. In A. Caplin & Andrew Schotter (Eds.), *The foundations of positive and normative economics* (pp. 3–41). Oxford: Oxford University Press.
- Heukelom, F. (2014). *Behavioral economics: a history*. New York: Cambridge University Press.
- Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960.
- Kahneman, D., Knetsch, J., & Thaler, R. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 98, 1325–1348.
- Kahnemann, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- List, J. (2006). Field experiments: A bridge between lab and naturally occurring data. *Advances in Economic Analysis & Policy*, 6, Article 8.
- Mill, J. S. (2007/1836). On the definition of Political Economy; and on the Method of ‘Investigation Proper to It. In *Collected works of John Stuart Mill* (Vol. 4). Indianapolis, IN: Liberty Fund.
- Robbins, L. (1932). *An essay on the nature and significance of economic science*. London, UK: McMillan & Co.
- Roth, A. (1986). Laboratory experimentation in economics. *Economics and Philosophy*, 2, 245–273.
- Roth, A. (1993). The early history of experimental economics. *Journal of the History of Economic Thought*, 15, 184–209.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688–701.
- Samuelson, P., & Nordhaus, W. (1985). *Economics*. New York: McGraw-Hill.
- Schotter, A. (2008). What’s so informative about choice? In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 70–94). Oxford, UK: Oxford University Press.
- Siegel, S. (1953). *Certain determinants and correlates of authoritarianism*. Department of Psychology, Stanford University.
- Smith, V. (1976). Experimental economics: Induced value theory. In *The American Economic Review, Paper and Proceedings of the Eighty-eighth Annual Meeting of the American Economic Association* (Vol. 66, pp. 274–279).
- Smith, V. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, 72, 923–955.
- Smith, V. (1989). Theory, experiment and economics. *Journal of Economic Perspectives*, 3, 151–169.
- Smith, V. (2008). *Rationality in economics: Constructivist and ecological forms*. Cambridge, UK: Cambridge University Press.
- Svorenčik, A., & Maas, H. (2016). *The making of experimental economics: Witness seminar on the emergence of a field*. Cham: Springer.
- von Wright, G. (2004). *Explanation and understanding*. Ithaca, NY: Cornell University Press.
- Weber, M. (1978). Economy and society: An outline of interpretive sociology. In G. Roth & C. Wittich (Eds.), *University of California Press*. Berkeley: CA.