GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 252

The Role of Experiments for the Development of Economic Theories

Klaus M. Schmidt*

*Klaus M. Schmidt, Department of Economics, University of Munich, Ludwigstr. 28, D-80539 München, Germany,
email: klaus.schmidt@LMU.de

January 2009

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Urs Schweizer. · Department of Economics · University of Bonn · D-53113 Bonn,
Phone: +49(0228)739220 · Fax: +49(0228)739221

The Role of Experiments for the Development of Economic Theories[*]

Klaus M. Schmidt[a)]
University of Munich, CESifo and CEPR

November 27, 2008

ABSTRACT: Economic experiments interact with economic theories in various ways. First of all they are used to test economic theories. However, they can neither confirm nor falsify them in a strict sense. They rather inform us about the range of applicability, the robustness and the predictive power of a theory. Furthermore, economic experiments discover and isolate phenomena and challenge economic theorists to explain them. Finally, many economic experiments are "material" models. They are used to analyse and predict how changes in the environment affect economic outcomes. However, they cannot offer an explanation for what we observe. This has to be provided by economic theory.

KEYWORDS: Economic experiments, economic theories, falsification, confirmation, phenomena, models.

JEL CLASSIFICATION NUMBERS: B41, C90.

[*)] I would like to thank Hartmut Kliemt, Clemens König and seminar participants in Munich for many very helpful comments, even though I did not take up all of their suggestions. Financial support by the German Excellence Initiative and the Deutsche Forschungsgemeinschaft through SFB-TR 15 is gratefully acknowledged.

[a)] Klaus M. Schmidt, Department of Economics, University of Munich, Ludwigstr. 28, D-80539 München, Germany, email: klaus.schmidt@LMU.de .

1 Introduction

What is the role of economic experiments for the development of economic theories? When Simon Gächter and the scientific organizing committee of this conference asked me to discuss this question I was a bit reluctant to accept the invitation. After all, I am not a philosopher of science and there are no new philosophical answers that I can contribute to the ongoing debate on the falsification and confirmation of theories in the social sciences.[1] However, the organizing committee wanted me to discuss the interplay between economic theory and economic experiments as an economist who is working in both of these fields. How should we deal with experimental results that seem to falsify or confirm economic theories? What other roles do experiments play in economics and in particular for economic theorists? It is from the point of view of an economist, not a philosopher, that I will address these questions.

There is a broad consensus among economists that experiments "test" economic theories. Indeed, it is difficult to ignore the claims that this or that experiment falsified neoclassical economics and the model of homo economicus on which it is based. On the other hand, Vernon Smith was awarded the Nobel Prize for his experiments on market games that seem to confirm neoclassical price theory that is also built on homo economicus. For almost any economic theory that has been suggested over the last decades there are some experiments claiming to confirm and some claiming to falsify them. Economic theorists seem unimpressed. Depending on the application many of them (including myself) work with neoclassical theory and behavioral theories - even though these theories are inconsistent with each other. Thus, the first question that I will discuss in this essay is in which sense experiments confirm or falsify economic theories and how we should deal with this.

Most experiments, however, do not test economic theories. Even though many experimental papers have a section in which they discuss whether the results of the experiment are consistent with standard theory or with a subset of alternative models they cannot be considered a rigorous test of any theory. For this the experiments are too complicated and leave too much room for different interpretations. These experiments have a different objective. They try to capture some essential aspects of a complex (social) decision process in the lab. Sometimes they uncover new phenomena, i.e. regularities in behavior that

---

[1] Guala (2005) provides an excellent discussion of many methodological issues in experimental economics from the perspective of modern philosophy of science. For an excellent early discussion of the role of experiments in economics see Roth (1995).

are typical in certain situations, and examine how robust these phenomena are. Or they are used as a "material" model of an aspect of the real world. A "material" model is a model of flesh and blood the exogenous variables of which are controlled by the experimental design in order to see how the endogenous variables react to changes in the treatment variables. These experiments often start out without much connection to economic theory. However, if they are successful, if they demonstrate the existence of robust phenomena or if the observed behavior corresponds well with behavior in the natural world, then they challenge economic theorists to explain them. This is the second role of experiments that I will discuss.

2. Testing Economic Theories

A laboratory experiment is a highly controlled situation in which subjects (mostly students) perform abstract tasks on a computer and interact anonymously with other subjects. How can behavior in such an artificial environment be a test of economic theories? After all, economic theories are supposed to explain human behavior "in the wild", i.e. in natural economic environments.

> *"The logic is as follows. General theories must apply to simple special cases. The laboratory technology can be used to create simple (but real) economies. These simple economies can then be used to test and evaluate the predictive capability of the general theories when they are applied to the special cases. ...*
>
> *A staggeringly large number of theories exist. One purpose of the lab is to reduce the number by determining which do not work in the simple cases."*
>
> Charles Plott (1991, p. 902 and 905)

According to Plott it is the "predictive capability" of a theory that is a measure of its success, i.e., a theory is confirmed if it makes good predictions. Furthermore, he emphasizes that the lab should be used to reduce the large number of theories by refuting those "that do not work", i.e., experiments should be used to falsify economic theories. This methodological position is widely shared by many economists, but I think it does need some important qualifications. I will first discuss in which sense experiments can confirm economic theories and then in which sense they falsify them.

2.1 Confirming Economic Theories

In the late 1940s Edward H. Chamberlin (1948) conducted simple market experiments with his students at Harvard University. He divided the students into buyers and sellers of a fictitious good. Each buyer was assigned a different reservation price and each seller a different cost to deliver the good. Reservation prices and costs were private information. Then he gave his students a few minutes to find a trading partner and to haggle about the price. When he compared the actual trades with the prediction of neoclassical price theory, the typical result was that prices fluctuated widely and that the traded quantity was often larger than the competitive equilibrium quantity. Chamberlin used these experiments not as a rigorous test but mainly as a pedagogical tool to demonstrate the incorrectness of neoclassical price theory.

In the early 1960s Vernon Smith conjectured that the problem with Chamberlin's market experiment is the lack of public information about available bids and offers. Smith (1962, 1964) conducted a series of market experiments that differed from Chamberlin's experiments in that all bids and offers were publicly recorded. Because buyers and sellers can make bids and offers simultaneously, this is called a "double auction". Furthermore, he conducted several sessions, so that his subjects could learn by experience. His experimental set up resembles the trading rules and procedures of the traditional trading floor of most financial markets before the introduction of computerized trading. The experimental results are striking: Prices quickly converge to equilibrium prices and the traded quantity is very close to the efficient quantity predicted by the competitive equilibrium. Do these experiments confirm neoclassical price theory?

There are two important arguments why experiments cannot confirm theories in a strict sense. First, there is the well known problem of induction. The fact that the outcome of an experiment coincided with the prediction of the theory in the past does not imply that it will do so in the future. The fact that the sun rose every morning over the last few billion years does not imply that it will do so again tomorrow. This fact is also consistent with the hypothesis that the sun rose every morning so far, but that it will take a break tomorrow. From a logical point of view, there are infinitely many theories that are consistent with any finite set of past observations. Karl Popper (1934, 1963) concludes that if the outcome of an experiment coincides with the prediction of one particular theory, this particular theory is not more (or more likely) true.

Popper's position is extreme and most philosophers of science do not agree with him on this point. There may be infinitely many theories that are consistent with a finite set of past observations, but they are not all equally convincing. Some theories are simpler, more general or more elegant than others, even if it is often difficult to specify exactly what these terms mean. Furthermore, the fact that a theory is able to predict correctly what is going to happen in the future even under conditions where this seems unlikely a priori is an important indication that this is not just one random theory consistent with the data. But, again, philosophers of science find it difficult to make these arguments precise.[2]

Second, the actual experiment does not reflect the assumptions imposed by neoclassical price theory. The theory assumes a frictionless market with infinitely many buyers and sellers, each of whom is perfectly rational and purely self-interested. It assumes the existence of a Walrasian auctioneer and a tatonnement process to find a market clearing price. It assumes that no trade takes place until the Walrasian auctioneer has found the price at which total demand equals total supply. In contrast, the experimental markets of Vernon Smith had only a few buyers and sellers, there was no Walrasian auctioneer and no centralized mechanism to find a market clearing price. His subjects could trade any time. Thus, the correspondence between the theory and the experiment is rather loose.

Despite these two qualifications, I believe that Smith's experiments impressively confirm neoclassical price theory, but only in a weak sense. They do not tell us that this theory is "true". But they do tell us a lot about its usefulness and applicability: First of all, the theory did very well in predicting market outcomes in hundreds (if not thousands) of double auction experiments with different information structures, different market sizes, and different trading rules. Second, the experimental results match the predictions of the theory even under extreme conditions. For example, Holt, Langan and Villamil (1986) conducted a double auction where all buyers have the same reservation price and all sellers have the same cost. If there is an excess supply, the theory predicts that all surplus goes to the buyers; if there is an excess demand, all surplus goes to the sellers. Many market participants consider these outcomes as very unfair. Nevertheless, after a few trading periods these are exactly the market

---

[2] See Guala (2005, p. 84ff) for a much more detailed discussion of the problem of induction.

outcomes observed in the experiments. Third, the fact that the set-up of the experiments and the assumptions of the theory do not correspond perfectly speaks in favor of the theory. It shows that the theoretical results do not describe a knife-edge case that holds only if all the assumptions of the theory are perfectly met. The experimental results suggest that the theory captures an important basic mechanism that determines market outcomes under much more general conditions. Thus, the theory has been shown to be robust and to be applicable under many different circumstances.

In this sense the double auction experiments do confirm neoclassical price theory. They show that this theory is a powerful predictor of economic outcomes in simple competitive markets. They do not show that neoclassical price theory or its assumptions are an accurate description of the real motivations and interactions on competitive markets. But they do confirm that this theory captures an important effect, and that this effect is so strong that it dominates the situation even if not all assumptions of the theory are satisfied.

The experiments also help us to understand the domain of the theory, i.e. under what conditions it predicts well and when it is less reliable. For example, it has been shown that neoclassical price theory makes accurate predictions when complete contracts are traded and when expectations do not play a role. However, if contracts are incomplete or if market participants have to form expectations about future prices, the experimental results often differ sharply from theoretical predictions.[3]

2.2 Falsifying Economic Theories

There are many economic experiments in which standard economic theories do less well. A famous example is the so called “gift exchange game” introduced by Fehr, Kirchsteiger and Riedl (1993) that considers a stylized labor relationship.[4] There are two players, called “employer” and “employee”. At the first stage of the game the employer offers a fixed wage w, $w \geq \underline{w}$, to the employee. At the second stage, the employee observes the wage and decides which effort level e, $e \in \left[\underline{e}, \overline{e}\right]$, to choose. The higher the effort level the higher is the employee’s monetary cost of effort, $c(e)$, and the higher is the profit $\pi(e,w)$ that accrues to the

---

[3] See e.g. Fehr, Kirchsteiger and Riedl (1993) for markets with incomplete contracts and Smith et al. (1988) for markets with bubbles.

[4] The gift exchange game belongs to the class of social dilemma games in which all players would benefit from cooperation, but each individual player has an incentive to defect. Other examples of social dilemma games include the prisoners’ dilemma game, public good games, trust games, etc.

employer. In the experiment joint surplus is maximized if the employee chooses the maximum effort level $\bar{e}$.

The game-theoretic analysis of this game is straightforward. At the second stage the wage is fixed, so it is a dominant strategy for a self-interested employee to choose the minimum effort level $\underline{e}$. A rational employer anticipates this and will offer the lowest possible wage $\underline{w}$. Thus, there is a unique Nash equilibrium in which cooperation fails and the lowest effort is chosen.

Figure 1 shows what happened in the experiment: For low wages most employees choose the minimum effort level $\underline{e}=0.1$. However, if higher wages are offered, many employees choose higher effort levels. There is a lot of heterogeneity among employees, but the average observed effort level is strongly increasing in wage. Many employers seem to anticipate this and offer high wages.

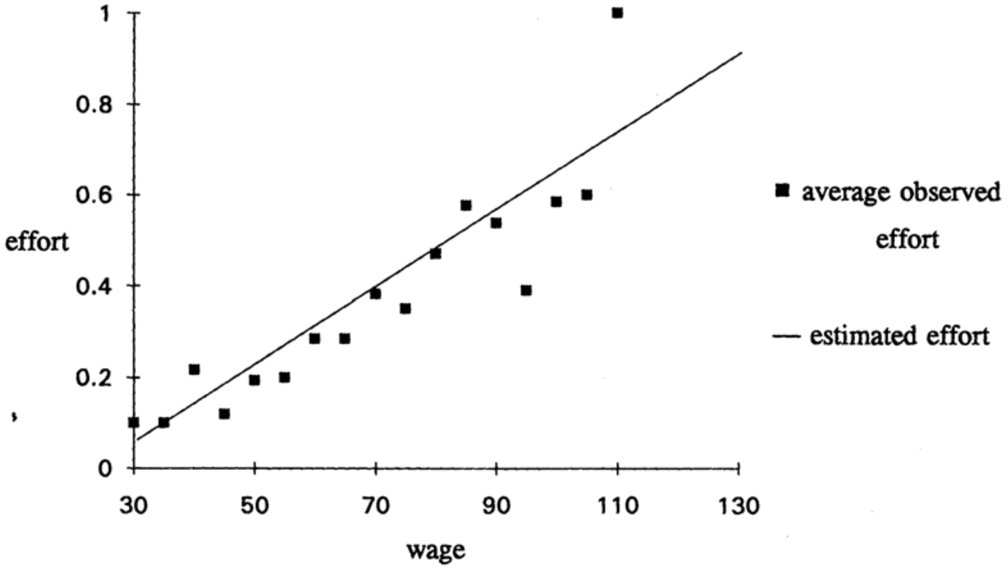

Figure 1: The Effort-Wage Relation in a Gift Exchange Game
Source: Fehr, Kirchsteiger and Riedl (1993, Fig. I)

Does the fact that many employees choose $e>\underline{e}$ falsify game theory? Again, there is an important argument why experiments cannot falsify theories in a strict sense. This is the so called Duhem-Quine problem[5] which states that it is impossible to test a theory in isolation. Any test is a joint test of the theory and several auxiliary assumptions (or bridge hypotheses) that bridge the gap between the theory and the actual experiment.

---

[5] The problem is named after the French physicist Pierre Duhem (1861-1916) and the American philosopher Willard van Orman Quine (1908-2000). For a more detailed discussion of this problem see Guala (2005).

Let me illustrate this point in the context of the gift exchange game. Game theory predicts:

> GT: "A player always chooses a strictly dominant strategy if such a strategy exists."

The hypothesis that was tested in the laboratory is:

> H: "All employees choose the minimum effort level."

The problem is that GT alone does not imply H. Additional auxiliary assumptions are required, for example:

> A1: "All experimental subjects understand the game and believe that this game is being played"

If some subjects are confused and did not fully understand the instructions or believe that a different game is being played, it may be rational to choose high effort levels. For example, it could be the case that the subjects believe that this game is part of a larger game in which they will play against the same employer again and that he will be able to retaliate if they do not provide enough effort. Given these beliefs their behavior could be perfectly rational.

> A2: "All experimental subjects care only about their own monetary payoffs" (self-interest hypothesis)

If this was not the case, e.g. because some subjects care about the payoff of the employer or are motivated by reciprocity, then these subjects may behave rationally by choosing high effort.

There are other auxiliary assumptions, e.g. that the game is common knowledge or that the game describes the situation in the lab as the subjects perceive it, but let us ignore them and focus on A1 and A2 only. Then we have:

$$[\text{GT and A1 and A2}] \rightarrow \text{H}$$

If we now observe that H is false (¬H), we can only conclude (by *modus tollens*):

$$\neg H \rightarrow \neg\,[GT \text{ and } A1 \text{ and } A2].$$

But:

$$\neg\,[GT \text{ and } A1 \text{ and } A2] \iff \neg GT \text{ or } \neg A1 \text{ or } \neg A2.$$

That is, we do not know whether GT itself or one of the auxiliary assumptions is false.

This problem is prevalent in all empirical research. An important advantage of experimental research is that control experiments can be designed that isolate the different auxiliary hypotheses and try to test them separately. For example, to make sure that the subjects understand what they are doing the experiment can be run several times, matching an employee to a different employer in each period. To rule out the possibility that the subjects are misled by the framing of the experiment, the players could be called “buyer” and “seller” or “actor A” and “actor B”. Clearly, this is much easier in the lab than in the field. But, even in the lab, experimental control is never perfect.

Suppose that experimental control solves the Duhem-Quine problem (to a reasonable approximation) and that additional experiments have shown that A1 is satisfied, but that it cannot be ruled out that A2 is false. Can we then conclude that “game theory in conjunction with the self-interest hypothesis” is falsified? Does it imply that we should not apply game theory in conjunction with the self-interest hypothesis to any other situation?

Consider the following analogy: The law of gravity says that “two mass points attract each other with a force that is proportional to their masses and inversely proportional to the square of the distance between them.” To apply the theory we need the auxiliary hypothesis that the objects under consideration are mass points. Suppose that we measure the speed of a feather falling from the tower of Pisa and find that the speed with which the feather hits the ground differs from the speed predicted by the theory. Further inquiries reveal that it cannot be ruled out that neither the earth nor the feather is a mass point. Does this imply that the law of gravity in conjunction with the mass point hypothesis is falsified and should not be applied?

We knew all along that mass points do not exist - as we knew that people are not purely motivated by self-interest. The mass point hypothesis and the self-interest hypothesis are idealizations that are made to keep the model simple and tractable. The question is not whether these assumptions are “true”, but rather whether reducing objects to mass points and human motivation to self-interest is legitimate. This depends on the situation under consideration. If the object has a high density, if it falls in a vacuum tube, or if the objects are planets that move in space, the law of gravity in conjunction with the mass point hypothesis makes highly accurate predictions, even though mass points do not exist. However, if the objects move through an atmosphere and if they have a low density, then other forces such as friction and buoyancy interfere and the law of gravity in conjunction with the mass point hypothesis performs poorly. Similarly, even though the gift exchange experiments (and many other experiments) show that game theory in conjunction with the self-interest hypothesis predicts poorly in social dilemma games, there are other classes of games, like the competitive market games discussed above, in which the predictions are quite good. Thus, experiments do not refute a theory in general. They rather show whether a theory and its auxiliary hypotheses are able to capture the main forces that determine the outcome in a certain class of situations, or whether other forces interfere that cannot be ignored.

Game theory in conjunction with the self-interest hypothesis cannot explain behavior in gift exchange games. There are other motivational forces (in addition to self-interest) that strongly affect behavior. Subjects may be motivated by social norms, by reciprocity, by altruism, by inequality aversion or by other concerns. In recent years several new theories have been proposed that try to capture these forces, e.g. efficiency seeking (Andreoni, 1989, Charness/Rabin, 2002), intention-based reciprocity (Rabin, 1993, Dufwenberg/Kirchsteiger, 2004, Falk/Fischbacher, 2006), type-based reciprocity (Levine, 1998) and inequity aversion (Fehr/Schmidt, 1999, Bolton/Ockenfels 2000). Each of these theories has been “confirmed” in some experimental settings and “falsified” in others.[6]

The Holy Grail of behavioral economics is a universal theory of human behavior that applies to all circumstances. It may be that such a theory exists and that we will eventually find it. However, at present I do not think that we should strive for such a universal theory. The theoretical attempts of the last 15 years have shown that human behavior is too complex to be fully described by simple principles, so a universal theory is likely to be complicated. But if we want to be able to say something about economically interesting and important phenomena such as how to regulate financial markets or how to design social security systems, we need simple and tractable models. It must be possible to integrate them in more complicated models that are needed to understand and deal with the economic processes that govern the performance of markets and organizations in the real world. The success of the model of homo economicus is based on its simplicity and tractability.

---

[6] See e.g. Charness and Rabin (2002), Engelmann and Strobel (2004) for experiments that test different theories against each other. Fehr and Schmidt (2006) offer a survey of the literature on social preferences.

Economic theorists have to develop new models in close collaboration with experimental economists. The models should be simple, robust and make reliable predictions in a broad range of applications. Experiments can be used to discriminate between motives and to evaluate how important different motives are in different situations. They are also used to evaluate theories and to show how they perform in different situations. Because theories have to be simple and tractable, no such theory will be able to predict well in all situations. By testing the theories in different situations experimental economics tells us a lot about the range of applicability of a theory.

It is sometimes argued that the theorist should specify in advance the range of applicability of his theory. I think that such a requirement is not only unrealistic, but that it would also stifle the development of new theories. Scientific progress works differently. Most new theories are proposed without a proper specification of the range to which they apply. By using the theory in different situations we slowly learn how applicable and useful it is and what its boundaries are. Experiments play a crucial role in this process.

3 Phenomena and “Material Models”

Testing economic theories is not the only function of experiments and perhaps not even the most important one. In this section I will discuss two other important roles that experiments play.

3.1 Discovering Phenomena

Phenomena are regularities in behavior or in outcomes that are often not obvious at first glance. They have be discovered and carefully isolated in the data. An experiment is a powerful tool to do this. The experimental researcher can design the experiment so as to establish the phenomenon as clearly as possible. By experimental control of the environment he can try to isolate the driving forces of the phenomenon and to establish causality. Other researchers can design additional experiments to test how robust the phenomenon is and how it interacts with the environment.

An interesting phenomenon is either counterintuitive or it contradicts standard economic theory. For example, the extreme and very unfair prices observed in double auction experiments in which all buyers have the same valuation and all sellers have the same cost and in which there is either an excess supply or an excess demand is an immediate implication of neoclassical price theory, but it is counterintuitive. On the other hand, the reciprocal behavior observed in many gift-exchange or trust games is intuitive, but it contradicts standard economic theory.

Interesting phenomena are puzzles that call for an explanation. They challenge our received wisdom and stimulate economic theorists to develop new theories that make sense of the observed behavior.

3.2 Material Models

Experiments are sometimes used in a way that is quite similar to the way how we use theoretical models. In this section I will argue that many experiments can be interpreted as "material models", i.e. models not composed of theory, but rather composed of elements of the real world.

Before I can explain this in more detail I have to distinguish the terms "theory" and "model" that I have used interchangeably so far. A theory is a semantic system expressed in a formal language. It is based on a number of axioms or assumptions from which logical conclusions are drawn. These conclusions are general statements that do not directly apply to any given phenomenon. Examples for such theories are game theory or general equilibrium theory. They deal with highly abstract concepts. For example, in game theory a "player" can be a human being, but it can also be a firm, a country, an animal, or part of a computer network. Similarly, a payoff can be a monetary payoff, but it can also be a utility level (that may depend on anything) or the reproductional fitness of a species. Furthermore, theories describe an idealized world that does not exist in reality. For example, general equilibrium theory makes statements about frictionless markets populated by a continuum of buyers and sellers. Game theory deals with hyperrational and omniscient decision makers who perfectly predict what other players are going to do in the future.

Economic theorists explore the properties of a theory often without any reference to the real world. For example, they try to figure out the most general assumptions that guarantee existence or uniqueness of an equilibrium, or they characterise other general properties of the theory. The object of these studies is the theory itself, not necessarily how the theory relates to the real world.

In contrast, a "model" is always a "model of something" in the real world. It is created in order to capture the essential forces that affect a set of specific situations. A typical economic model is composed of theory together with empirical regularities and judgements of the model builder. For example, a model of the German electricity market combines game theory with assumptions that are based on empirical knowledge (e.g. about the number of players, their capacity constraints, the regulatory environment, etc.) and on judgements of the model builder (e.g. whether companies compete in prices or quantities, whether the model is static or allows for repeated interaction, etc.). Whether a model is a good or a bad model cannot be judged in isolation, but only in relation to the aspects of the situation it is supposed to capture.

Thus, a model is a tool that is used to predict and evaluate what is going to happen under various circumstances. For this control is crucial. It must be possible to manipulate the model, to change initial conditions or exogenous parameters, and to see how these changes affect the predicted outcome.

Since the 1950s almost all economic models have had a mathematical structure. In this respect they look very similar to theories. However, a model can also have a less formal structure. For example, it can use flow charts or graphical representations. It may be difficult to get this type of models published in economic journals, but they are used and respected in many other sciences (including economics at least until the middle of the last century).

A model can also be a “material” or “physical” object. A famous physical model in economics is MONIAC, the MOnetary National Income Analogue Computer that was built by Bill Philipps at the London School of Economics in 1949.[7] MONIAC measures approximately 2 cubic meters and consists of a series of transparent pipes and plastic tanks. Each tank represents some aspect of the UK economy. At the top of the machine is a large tank called the treasury. Water (representing money) flows from the treasury to other tanks representing various sectors of the economy. For example, there are tanks for health and education. To increase spending on health care a tap can be opened to drain water from the treasury to the tank that stands for health spending. Water then runs further down the model to other tanks, representing other interactions in the economy. The actual flow of the water is automatically controlled through a series of floats, counterweights, electrodes and cords. When the level of water reaches a certain level in a tank, pumps and drains are activated. Phillips managed to calibrate MONIAC to an accuracy of about two percent. His senior colleagues at LSE were so impressed that they gave him tenure for his invention. When I

---

[7] See http://www.sciencemuseum.org.uk/objects/computing_and_data_processing/1995-210.aspx and http://en.wikipedia.org/wiki/MONIAC_Computer

studied at the LSE in 1989, MONIAC was still on display. It can now be seen at the Science Museum in London.

MONIAC has all the characteristics of a proper model. It captures the main features of the "target system" that it is supposed to represent. Most importantly, it can be manipulated. By pumping different quantities of water from the government's budget to some sector the effects on all the other main parts of the economy can be studied.

Figure 2: The MOnetary National Income Analogue Computer (MONIAC)
Source: Science Museum London
Copyright can be obtained at
http://www.scienceandsociety.co.uk/results.asp?image=10303578

From our perspective today, MONIAC looks a bit unusual, but in other sciences material models are very common. For example, a map is a material model of a geographical area. Engineers use plywood models of an airplane to predict how a real airplane behaves under different airstream conditions. In medical research animals are used as models for human beings to see how certain drugs affect the human organism. The idea is that the essential organic mechanisms of, say, a guinea pig are sufficiently similar to those of a human being that the drug has similar effects. The guinea pig is used in a similar way as we use a mathematical model. In the mathematical model we keep parameter values and initial conditions fixed and study how the change of one exogenous variable (say a tax rate) affects the other variables in the model. Similarly, we keep the environment of the guinea pig fixed and study how the change of one variable, namely the quantity of the administered drug, affects the organism of the guinea pig.

Many economic experiments can also be seen as material models in this sense.[8] For example, the gift exchange game by Fehr, Kirchsteiger and Riedl (1993) discussed above shows how unvoluntary unemployment can result from contractual incompleteness. Other experiments model the effects of minimum wages or job protection laws in the lab[9] or they show under what contractual conditions voluntary cooperation can be sustained.[10] The lab has also been used as a "wind tunnel" to test the design of real world incentive mechanisms. For example, before the spectrum rights were auctioned off in the US and in Europe, economic experiments were used to test the design of the auctions. The experiments used simplified auction rules, students as subjects, and financial incentives that were many orders of magnitude smaller than the money at stake in the real auction. Nevertheless, the insights from the auctions in the lab were very useful to improve the design of the real auctions.[11]

Lab experiments and mathematical models have many similarities. Both ask how changes of the environment affect economic decision making and economic outcomes. They are based on and motivated by theories, prior empirical knowledge and the judgement of the economist who designs the model or the experiment. A crucial element of both, mathematical models and lab experiments, is control. We want to study the effects of the change of one variable in isolation. For this we need to control the variable that is being changed and we need to control the environment (the background conditions) in which the change takes place. In a mathematical model control is achieved by the assumptions and initial conditions of the model and by imposing ceteris paribus conditions. In a lab experiment control is achieved through the experimental design and the physical restrictions imposed on the experimental

---

[8] For a similar view of models and experiments see Morrison and Morgan (1999) and Morgan (2005).

[9] See e.g. Falk, Fehr and Zehnder (2006)

[10] See e.g. Fehr, Klein and Schmidt (2007) and Bartling, Fehr and Schmidt (2008).

[11] See Plott (1997) and Abbink et al. (2005).

subjects. As Uskali Mäki (2005) puts it, a model is a "thought" experiment, while a lab experiment is a "material" model. Mäki goes even further and says: "Experiments are models, models are experiments".

However, equating mathematical models and lab experiments blurs some important differences between them. The former is a formal system built on assumptions and initial conditions from which predictions are derived deductively. The latter is an artificially created part of the real world. Predictions are derived by experimentation, i.e. by changing some conditions in the lab and seeing what comes out. Thus, an experiment is made of the same material as the real world, while a model is made of a formal language, merely representing some aspects of the real world.

Furthermore, mathematical models and lab experiments face different problems. The problem of a mathematical model is its "realism". Does it capture the relevant aspects of a situation sufficiently well to come up with accurate predictions? The problem of an experiment is its internal and external validity. Internal validity refers to the question whether we are drawing the right inferences within the experiment. Does the change of the treatment variable X cause the observed change of the variable Y, or do we observe a mere correlation? External validity refers to the question whether we are drawing the right conclusions from the experiment about the real world. Does the situation in the lab capture all relevant factors that affect behavior in the real world situation that we want to explain? Can we transfer the experimental results to the real world?

Finally, lab experiments and formal models have a different potential for generating discovery and understanding. The analysis of a theoretical model helps us to uncover all the implications of the assumptions that we started from. Some of the implications may be non-obvious and come at first glance as a surprise. However, if we think long enough about the model, we eventually see that all the results are present in the assumptions already. We get out of the model only what we put in with the assumptions we imposed. Thus, a model does not generate anything "new". In contrast, an experiment may uncover a phenomenon that is not predicted by any available theory and that nobody could have predicted. We may observe a truly unexpected phenomenon in an experiment.

On the other hand, an experiment can only demonstrate that a phenomenon exists, but it cannot explain it. A mathematical model may offer an explanation by showing how the phenomenon can be derived from more basic assumptions. The challenge for economic theory is to show that very different phenomena and very different mathematical models of these phenomena can be derived from common principles. The deeper and more general these principles are, the deeper is our understanding of the phenomena.

Let me illustrate this with an example. The double auction experiments of Vernon Smith and others have shown that neoclassical price theory can predict economic behavior on competitive markets fairly well. Neoclassical price theory is based on the assumptions that all people are perfectly rational and purely self-interested. However, the gift exchange experiments by Ernst Fehr and others are inconsistent with these assumptions. They show that fairness and reciprocity are strong motivational forces that affect behavior. Thus, there is a puzzle: If fairness is such a strong motivational force if subjects interact in small groups, why do the same subjects accept very unfair outcomes in market experiments?

The theory of inequity aversion by Fehr and Schmidt (1999) offers a first answer to this question. The theory assumes that people are not only interested in their own material payoff but also in how their payoff compares to the payoff of other people in their reference group. They dislike being worse off and they also dislike being better off than other members of their reference group. However, the latter inequality aversion may be less pronounced than the former. Furthermore, people are assumed to be heterogeneous: Some people are more inequity averse than others. Fehr and Schmidt (1999) show that this very simple model can explain why the same people behave very fairly in some games and very unfairly in others. In particular, they show that a highly inequity averse player who rejects any unequal proposal in the ultimatum game will accept very low offers in an ultimatum game with responder or proposer competition. The reason is that competition makes it impossible for any subject to enforce an equal outcome unilaterally. Thus, people are forced by competition to behave as if they were completely selfish.

More recently Dufwenberg et al. (2008) have generalized this result considerably. They consider a general equilibrium model and allow for a large class of social preferences. They show that the set of equilibria is independent of whether people are purely self-interested or whether they have social preferences. In competitive markets, self-interest and social preferences are observationally equivalent. Thus, this theoretical result helps us to better understand why the neoclassical model of homo economicus is doing fairly well in some circumstances but less so in others.

4 Conclusions

Economic experiments interact with economic theories in various ways. First of all they can be used to test economic theories. However, they can neither confirm nor falsify economic theories in a strict sense. They rather inform us about the range of applicability, the robustness and the predictive power of a theory. Furthermore, economic experiments discover and isolate phenomena and challenge economic theorists to explain them. Finally, many economic experiments are material models. They are used to analyse and predict how changes in the environment affect behavior and economic outcomes. However, economic experiments cannot offer an explanation for what we observe. This has to be done by economic theory. The challenge for economic theorists is to explain the observed behavior with tractable models and to derive these models from a set of simple common principles.

References

Abbing, K., Irlenbusch, B., Pezanis-Christou, P., Rockenbach, B., Sadrieh, A., and Selten, R. (2005), „An Experimental Test of Design Alternatives for the British 3G/UMTS Auction," European Economic Review, Vol. 49(2), 505-530.

Andreoni, J. (1989), "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy* 97, 1447-1458.

Bartling, B., Fehr, E. and Schmidt, K. M. (2008), "Reputation and Contract Design," *mimeo*, University of Munich.

Bolton, G. E. and Ockenfels, A. (2000), A Theory of Equity, Reciprocity and Competition. *American Economic Revie,w* Vol. 100, 166-193.

Chamberlin, E.H. (1948), "An Experimental Imperfect Market," *Journal of Political Economy*, Vol. 56(2), 95-108.

Charness, G. and Rabin, M. (2002), "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics, 117*, 817-869.

Dufwenberg, M. and Kirchsteiger, G. (2004), "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, Vol. 47, 268-298.

Dufwenberg, M., Heidues, P., Kirchsteiger, G., Riedel, F., Sobel, J. (2008), "Other Regarding Preferences in General Equilibrium," *CEPR Discussion Paper* No. 6815.

Engelmann, D. and Strobel M. (2004), "Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments", *American Economic Review* 94 (4), 857-869.

Falk, A., Fehr, E., and Zehnder, C. (2006), "Fairness Perceptions and Reservation Wages – The Behavioral Effects of Minimum Wage Laws," *Quarterly Journal of Economics*, Vol. 121, 1347-1381.

Falk, A. and Fischbacher, U. (2006), "A Theory of Reciprocity," Games and Economic Behavior, Vol. 54(2), 293-315.

Fehr, E., Kirchsteiger, G. and Riedl, A. (1993), "Does Fairness prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics,* Vol. 108, 437-460.

Fehr, E., Klein, A. and Schmidt, K. M. (2007), "Fairness and Contract Design," *Econometrica*, Vol. 75(1), 121-154

Fehr, E. and Schmidt, K. M. (1999), "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, Vol. 114, 817-68

Fehr, E. and Schmidt, K. M. (2006), "The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories," in : S.-C. Kolm and J. M. Ythier (eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism*, Elsevier/North-Holland, Amsterdam.

Guala, F. (2005), *The Methodology of Experimental Economics*, Cambridge: Cambridge University Press.

Holt, C. A., Langan, L., and Villamil A. P. (1986), “Market Power in Oral Double Auctions”, Economic Inquiry, Vol. 24, 107-123.

Levine, D. (1998), “Modeling Altruism and Spitefulness in Experiments,” *Review of Economic Dynamics*, Vol. 1, 593-622.

Mäki, U. (2005), “Models are Experiments, Experiments are Models,” *Journal of Economic Methodology*, Vol. 12(2), 303-315.

Morgan, M. S. (2005), “Experiments versus Models: New phenomena, inference and surprise,” *Journal of Economic Methodology*, Vol. 12(2), 317-329.

Morrison, M. C. and Morgan, M. (1999), “Models as Mediating Instruments,” in: M. S. Morgan and M. C. Morrison (eds.), *Models as Mediators*, Cambridge: Cambridge University Press, 10-37.

Plott, C. (1991), Will Economics Become an Experimental Science?” *Southern Economic Journal*, Vol. 57, 901-19.

Popper, K. (1934), *Logik der Forschung*, Wien: Springer Verlag.

Popper, K. (1963), *Conjectures and Refutations*, London: Routledge.

Rabin, M. (1993), “Incorporating Fairness into Game Theory and Economics,” *American Economic Review*, 83(5), 1281-1302.

Roth, A. E. (1995), “Introduction to Experimental Economic,” in: J. H. Kagel and A. E. Roth (eds.) *The Handbook of Experimental Economics*, Princeton: Princeton University Press, pp. 3-109.

Smith, V. L. (1962), “An Experimental Study of Competitive Market Behavior,” *Journal of Political Economy*, Vol. 70 (2), 111-137.

Smith, V. L. (1964), “Effect of Market Organization on Competitive Equilibrium,” *Quarterly Journal of Economics*, Vol. 78(2), 181-201.

Smith, V. L., Suchanek, G. L., Williams, A. A. (1988), “Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets,” *Econometrica* 56, 1119-1151.