# Experimental Economics: Some Methodological Notes

Fiore, Annamaria

January 2009

# Experimental Economics: Some Methodological Notes

ANNAMARIA FIORE[*]

Preliminary draft
This version: January 2009

**Abstract**

The aim of this work is presenting in a self-contained paper some methodological aspects as they are received in the current experimental literature. The purpose has been to make a critical review of some very influential papers dealing with methodological issues. In other words, the idea is to have a single paper where people first approaching experimental economics can find summarised (some) of the most important methodological issues. In particular, the focus is on some methodological practises still debated in experimental literature, such as attainment of control in experimental settings, subject pool, incentive mechanisms, repeated trials and learning. The hope is that increasing awareness on some sharing methodologies will improve the robustness of results in this research field.

*Keywords:* Experimental Economics, Methodology, Control, Incentives, Learning, Deception.

[*] Corresponding author. Annamaria Fiore, Via Marengo 11/A, 70033 Corato (Bari), Italy. Tel.: +39-(0)80-3720036.
*E-mail address:* afiore@dse.uniba.it (Annamaria Fiore).

## 1. Introduction

Traditionally, Economics has been established as science in 1776, coinciding with the publication of "Inquiry into Nature and Causes of the Wealth of The Nations" by Adam Smith. Until very recently, however, Economics was not considered an *experimental* science, but rather an observational science akin to astronomy. For a long time, J. S. Mill's "On the Definition of Political Economy" (1844; 1874) pulled its rank on this concept. Mill was a radical empiricist, and persuaded that the primary source of knowledge was obtained in an inductive way from the sense experience. Notwithstanding this, he accorded a dispensation to social sciences in general for the impossibility of using experimental method, and a dispensation to political economy in particular, for its special subject of inquiry. This view is also echoed in M. Friedman's words (1953)[1]:

> "Unfortunately, we can seldom test particular predictions in the social sciences by experiments explicitly designed to eliminated what are judged to be the most important disturbing influences. Generally, we must rely on evidence cast up by the "experiments" that happen to occur […] The necessity of relying on uncontrolled experience rather than on controlled experiment makes it difficult to produce dramatic and clear-cut evidence to justify the acceptance of tentative hypothesis".

While he admits the importance of collection of data under *ceteris paribus* conditions for the development of a science, at the same time Friedman generally denies the *status* of experimental science to economics, partly because he does not consider this aspect relevant for the distinction between natural and social sciences, partly because he concedes a particular degree of difficulty to economics, given its proper contents:

> "… the fact that economics deals with the interrelations of human beings, and that the investigator is himself part of the subject matter being investigated in a more intimate sense than in physical sciences, raises special difficulties in achieving objectivity at the same time that it provides the social scientist with a class of data not available to the physical scientist".

We can trace exactly the same ideas even more recently, bearing witness to a common view of considering economics not to be among the experimental disciplines: "One possible way of figuring out economic laws … is by controlled experiment… Economists [unfortunately]… 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, as cited by D. Friedman and Sunder, 1994)[2].

---

[1] To get a general idea of the conventional wisdom about the experimental status of economics at the outset of the discipline, we can also refer to Chamberlin (1948, p.95).

[2] Also note Lipsey's pessimistic note regarding experiments in economics, *An Introduction to Positive Economics*, 1979.

But, as D. Friedman and Sunder (1994) point out, once some methodological procedures and techniques have been *repeatedly used and firmly established* in conducting experiments in a particular discipline, this discipline begins to be considered also an experimental science. Based on this, we may now say with some confidence that economics is an experimental science and that economics experiments have became an useful and irreplaceable research tool. And there have been even some authors that have not boggled at stating that "experimental economics has been the protagonist of one of the most stunning methodological revolutions in the history of science" (Guala, unpublished paper).

## 2. A Historical Perspective

First experiment run in economic field dates back to 1931 (if we do not take into account the St. Petersburg paradox, 1738), when Thurstone tried to determine empirically individual indifference curves. His experimental design was rather naïve and based on hypothetical questions, and it attracted critics soon afterwards. It is indicative that the first critical review of this work was given by M. Friedman, in conjunction with Wallis, 1942, about the reliability or not of experimental data, the so called "artificiality critique" (see Section 7.1).

What is relevant here it is that at this criticism a similar experiment followed up, constructed by Rousseas and Hart (1951), in which they tried to meet objections. Hence, even in the very first steps, a feature emerges that then will become a constant in experimental economics: the knowledge is being cumulated and experimental works build upon one another. In this respect, it is not possible consider any experiment in isolation, but rather as being part of a *cumulative process*, a system of experiments related to each other (Roth, 1988) or, as Kagel and Roth (1992, p.1390) note: "experimental methods allow investigators to re-examine one another's conclusion relatively easily, and this is an important factor in the vitality of the experimental enterprise".

This probably may have concurred in establishing a not complete and systematic methodology for some particular aspects, that is complained in some parts. To some extent and from a given perspective, experimental economics seems to be still a discipline "under construction", subjected to continuous refinements and improvements, even if gradually, on some issues, a tacit consensus at the beginning, and a more explicit consensus afterwards, has grown up (Friedman, 1988). Two remarks are necessary. First, the fact that experimental economics appears to be not completely definite from a methodological point of view is not detrimental to itself. On the contrary, being a not satisfactorily explored field yet, it is fascinating and worthy of a closer examination. In this connection, it is noticeable that, in the very last years, there have been some examples of interesting *meta–experiments*, in the sense that some experimentalists have used the methods developed in experimentation to test the robustness of the methods themselves. For an example, see Section 5.3.

Second, this does not mean that no main point has been reached in this domain, as we will see in detail below.

Going back to the historical development of the discipline in the first decades, we should note that the amount of works related to experiments in the economics was not particularly large and systematic, even if there were very important exceptions. For further details, see Roth (1995) and Guala (unpublished paper). It was from the seventies that such a trend was destined to grow steadily. This occurred for a series of reasons, with the very first research programmes fully devoted to experimental investigation in economics (for examples in the public goods domain, see Ledyard, 1995).

The last decades have witnessed to an ever increasing interest related to experiments in economics. We should list some indications for this, as the introduction of specific codes of classification in the Journal of Economic Literature in the late 80s, the foundation of the Economic Science Association (ESA[3]) in 1986, the publication of specific textbooks (Hey, 1991; Davis and Holt, 1993; D. Friedman and Sunder, 1994; Kagel and Roth, 1995, D. Friedman and Cassar, 2004), the appearance of a specific journal devoted to it in 1998, while experimental works continue being published in the most important economic journals as well[4], and, finally, the awarding of the Nobel Memorial Prize in 2002 to two prominent experimentalists, V. Smith[5] and D. Kahneman[6]. This awarding can be considered as the official acknowledgement of experimental turn in economics. Hence, the question raised by Plott in the Sixtieth Annual Meeting of the Southern Economic Association in 1990 (Plott, 1991), "will economics become an experimental science?", may have found a definitive and positive answer.

As a consequence, nowadays it seems rather hard to deny that economics has reached the experimental status among sciences (with regard to this, it could be significative to note that Samuelson and Nordhaus have removed the sentence reported above in the following editions of their book).

However, some experimental economists oppose the establishment of strict rules in experiments ("it is important to avoid establishing rigid orthodoxies on question of methodology", Roth, 1995, p.86), since they attach more importance to the possibility to balance costs and benefits in the design of any single experiment, in

---

[3] From the ESA statement of purpose and bylaws: "The Purpose of the Association. I. To advance, enhance, or further economics as an observational science through use of laboratory and field methods of observation and data collection under the control and responsibility of the research investigator, and the development of economic theory and statistical or econometric methods based on such direct observations and data. The Association seeks to foster replicable, clearly documented, empirical work in all subdisciplines of economics, and recognition of the important tasks of data creation, data quality evaluation and empirical description, as well as theory development and testing".

[4] See Palfrey and Porter (1991) for some guidelines for the submission of an experimental work in *Econometrica*, as evidence of interest in experimental economics in economic scientific community.

[5] "for having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms" (from the motivation of the prize).

[6] "for having integrated insight from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty" (from the motivation of the prize).

order to plan each experiment in conformity with its specific goals. In my opinion, however, in this discipline, as in other experimental disciplines as well, in designing an experiment, in deciding what is important and what is not, what to be included and what to be not, it is a question of careful appreciation and experience, a sort of trade-off between flexibility and orthodoxy, between control and degree of freedom. Nevertheless, on a certain number of issues, it seems that a general acceptance and agreement has been reached among experimentalists. If we cannot refer to a well-established methodology yet, formalized in rigid rules, without any doubt we can refer to a "common practice" generally received.

In the following sections, we will try to list some of the uncontroversial issues.

## 3. Control

By definition, an experiment is a carefully planned and fully replicable observation of a phenomenon under *controlled conditions*.

Of preliminary and paramount interest is try to attain as much control as possible in the laboratory ("control is the essence of experimental methodology", Smith, 1976, p. 275). As in other experimental disciplines, experiments in economics are carried out under the *ceteris paribus* condition. Generally, a great number of variables theoretically and practically affects results in the real world as well as in a particular experimental environment, so the first main task of the empirical scientist consists in individuating the most important variables for his/her own goals. After this preliminary analysis, then it is important to try to keep other variables *constant* (maintained at some constant level along all the experiment), in order to isolate the actual effect of the *treatment variable(s)* (variable(s) set at different levels at different points in the same experiment or changed across different treatments) on the phenomenon under examination. Depending on the specific aim of an experiment, the same variable may be considered treatment variable or constant across different experiments or in different tournaments of the same experiment (i.e., to test for interaction between treatments). Obviously, in order to detect the effect of any treatment variable on the phenomenon, the experimenter will choose two or more appropriate 'values' for each treatment variable, including the case in which the treatment variable will assume a 'reference value', treatment usually referred as the *control group*, that it will serve as the benchmark in analysing results. Another way to obtain control indirectly is by *randomisation*. Given that usually it is not possible to control for every possible influential factor, the best practice consists of assigning participants among treatments using a complete random device. At this end, the use of specific recruiting software could prove useful (among the most used ones, we can cite ExperimenTrak, ORSEE (Greiner, 2004), CasselWeb2, and ExLab), since they permit to recruit and assign subjects to experiment in a completely random and anonymous manner.

Classical example has become the experiment to test for willingness to cooperate in a prisoner's dilemma game (Dresher and Flood at Rand Corporation, 1950). Since there is sufficient reason to believe that there may be not negligible differences in behaviour among male and female subjects, also known as *gender effect* (Ortmann and Tichy, 1999), in case the experimenter is not particularly interested in it (or in case of financial constraints, so that the researcher have not the opportunity to isolate each single treatment variable, but she is forced to focus her attention only on a sub-sample of these), she will look to allocate male and female randomly across treatments, such that results will not be vitiated a priori by any not considered gender effect[7].

Another important feature in designing an experiment in economics consists of choosing between a *within design* or a *between design*. In the former, each participant will experience more than one treatment variable, in the latter, each participant will experience just one treatment. Naturally, both of them have merits and flaws. The principal merit in using a within design is the opportunity to generate a greater amount of data at a smaller cost, but often these data are harder to analyse, even if more reliable and meaningful, since more statistically powerful, compared to the same number of observation obtained under between design. Moreover, many typical situations being investigated in economics experiment do not lend themselves to between design (e.g., in the case of a binary outcome variable).

However, data obtained under a within design experiment often need a further test: a test for detecting any *order effect* that could be occurred when participants have been exposed to more than one task in succession.

### *3.1. Controlling for subjects' preferences*

Especially in individual decision making, since most of the theories to be tested are based on given assumptions about the subjects' utilities, it is essential to measure subjects' preferences to test consequent predictions, or, at least, to try for control for them. In experimental literature, a series of attempts have been developed at this aim. We will try to mention the procedures most widely used by experimenters in economics.

#### *3.1.1. Vickrey's auction*

In a very well-known paper, Vickrey (1961) demonstrated that the procedure to obtain people truly revealing their valuation for a commodity is the second-price sealed-bid auction.

In this particular type of auction, where bidders submit written bids without knowing the bids made by other participants, the commodity is allocated to the bidder with the highest bid, but the price to be paid is the second highest bid.

The principal feature of this kind of auction is its self-revelation or incentive compatibility property, whereby each bidder maximizes her expected utility by revealing her true valuation. Moreover, it is ex-post efficient, that is, the winning bidder is the subject with the highest valuation for the commodity to be auctioned.

Despite its large use in economic theoretical and empirical literature (virtually, it is rather difficult to list all the works that actually implement this auction), the fact that it is not particularly common in everyday life, it could require a lot of training

[7] Recently, a broad literature has been developed about the relationship between demographics and economic behaviour. See, for example, Carbone (2005).

for the experimental subjects to acquaint themselves with the procedure, slowing down the opportunity to achieve its theoretical properties.

### *3.1.2. Becker-DeGroot-Marschak (BDM) mechanism*

The authors' primary aim (Becker et al., 1964) to be attained with this paper was the development of a sequential method that provides at each step an estimate of the utility for a commodity.

It is important to note that this procedure is based on the expected utility theory (EUT), the mainstream theory in individual decision making, but it is independent of the specific attitude towards risk.

The method is such that the subjects are endowed with a risky prospect, and then they are required to state their willingness to accept (WTA) to sell this lottery, $s$. If the subject obeys the axioms of the EUT, then it can be shown the dominant strategy be the true revealing of her certainty equivalent for the lottery. At this point, a mechanism is implemented to establish the selling price: a number $b$ is drawn randomly from a uniform distribution. Then, if the case $b \geq s$ occurs, the subject sells the lottery and gets the price $b$, if the case $b < s$ occurs, the subject retains the lottery and its corresponding prize will be determined.

In fact, this mechanism has been very widely used in the literature, especially to categorize subjects according to their attitude to risk, also thanks to its relative simplicity. Nevertheless this mechanism has been proved to have some shortcomings. Just to cite one, the empirical finding about the willingness to pay (WTP)-WTA disparity or endowment effect (Knetsch, 1989) could cause subjects to appear more risk loving than they really are, given that they are required to state their WTA. In fact, it has been demonstrated empirically that WTA is generally higher than the corresponding WTP in many experiments.

Moreover, it is important to note that some experimenters are sceptical in the use of non-iterative price-naming institutions for the elicitation of value, as BDM mechanism actually is. For example, Bohm et al. (1997) have found that, since the mechanism is sensitive to the choice of upper bound of the randomly drawn selling price, $b$, it cannot be considered incentive-compatible in general.

### *3.1.3. Binary Lottery System*

This method was developed by Roth and Malouf (1979) with the explicit goal of controlling for subjects' utilities, provided that people obey to reduction of compound lotteries axiom.

In their experiments, each subject $i$ bargains over the distribution of lottery tickets. Two the possible prizes: a small ($\sigma_i$) and a large ($\lambda_i$) amount of money, and the tickets determine the probabilities of winning the large prize, in the sense that if one agent gets 30 percent of the available tickets, she also has 30 percent of possibility to win the large prize. Given that the utility functions are unique up to a linear transformation, we can put $u(\sigma_i)= 0$ and $u(\lambda_i)= 1$. In this manner, it is simply to note that each agent's expected utility is exactly equivalent to the percentage of lottery tickets obtained.

As authors themselves note, the virtue of this mechanism is the fact that it is quite simple to create the "complete information" condition, since knowing her own utility corresponds to know exactly other bargainer's utility.
Although this procedure has been used in some experimental contexts (auctions, game theory) a recent paper by Abbink et al. (1999) has pointed out some shortcomings of this method.

### *3.2. Controlling for income effect: the Random Lottery Incentive (RLI) system*

Very often, subjects participating in experiments are required to perform more than one task, especially in experiments featuring within-subject design. Since they are generally paid for each task they complete, this fact could produce the so called 'income effect'[8], in the sense that at the starting of a subsequent task, the subjects' endowment has changed during the experiment with respect to the beginning, contaminating the data. For this reason, and also to economize on research costs, allowing experimenters to gather a greater amount of data with the same expenditure, it has become rather established among researchers to use the RLI system as incentive mechanism. According to this procedure, only the reward related to just one task performed during the experiment will be paid at the end. It is essential to note that the task to be paid is *randomly* chosen at the end of the experiment (each task has exactly the same probability to be drawn) and that the subjects are informed about the procedure before the experiment starts. In this way, even if the payoff is gradually changing along with the experiment, nevertheless this procedure should ensure subjects keep their endowment unchanged until the end (with regard to this, it is assumed that people consider each task in isolation).
To the best of my knowledge, it is very likely that Becker et al. (1964) were the first in using also this procedure. Ever since, the mechanism has been largely employed by researchers for its appealing features, until Holt (1986) did not start the debate and subjected the procedure to criticism.
The controversy was brought about by the results about the preference reversals[9] obtained by Grether and Plott (1979). In his paper, Holt affirms that the RLI system is only appropriate if the axioms of the EUT are satisfied.
In a typical preference reversals experiment, people are required to perform three tasks, one choice task and two pricing tasks (one for each lottery), and since the RLI system is implemented, if subjects treat all the tasks in the experiment as a whole, the strategy in the game is equivalent to the choice among two compound lotteries. Therefore, he shows that if people fail to obey the independence axiom, the

[8] In the literature, this effect is also known as 'wealth effects'.

[9] The discovery of this phenomenon was due to two psychologists, Slovic and Lichtenstein (1968) and Lichtenstein and Slovic (1971), but it attracted soon economists' attention because it was against the *procedure invariance* they advocated. In fact, contrary to this, it was discovered that it is possible to construct pairs of lotteries such that, when asked to choose among them, most people prefer one of the two (the P-bet, a lottery with a high probability of winning a small amount of money), but when required to state the price they would selling them at, they put a higher price on the other lottery (on the $-bet, a lottery with a low probability of winning a large amount of money).

preferences elicited in these compound lotteries may not be the same as elicited in a direct choice. Consequently, the use of the RLI system may distort the results whenever the independence axiom, or some other axiom, such as the reduction of compound lotteries, is not satisfied. For example, an explanation could be found in the "isolation effect" (Kahneman and Tversky, 1979): it may be that, even if every decision is important, subjects would not group all the decision and treat them as an interconnected portfolio of decisions (see also Read et al., 1999).

Starmer and Sugden (1991) ran a first experiment to show that the RLI would have no part in the phenomenon of preference reversals, because effective in revealing true preferences. In fact, in their experiment, no significant difference was detected between real single choices and random lottery designs, although people's responses have been demonstrated to not be always consistent with the reduction axiom.

Furthermore, in Harrison (1994) the mechanism has been criticized on the basis that it may not guarantee that all the costs involved in the decision process are overcome by the relative benefits (see Section 5.1 in which the dominance precept is introduced), since the expected monetary payoff per task is generally very low. Wilcox (1993) actually provides a first evidence that the incentive levels are important only when the decisional environment is complex enough, whereas when the effort is minimal, there is no reason for subject to economize on that effort not only at normal incentive levels, but even when incentives are absent.

Other experiments have followed up, in which Beattie and Loomes (1997) and Cubitt et al. (1998) test its validity: the former compares single choice and random lottery treatments, the latter designed also to detect *cross-task contamination effects* in treatments employing this system. In general, they again found no significant differences among the treatments in which subjects faced just a single choice task and the treatment in which the RLI system was used[10], and therefore Cubitt et al. (1998) conclude that there may be no evidence for cross-task contamination effects.

Finally, a more recent paper by Hey and Lee (2005) is a further evidence that subjects actually consider each question in isolation, rather than consider the experiment as a whole, also because this would require a "singular degree of sophistication on the part of subjects" (p. 235).

Anyway, these are still preliminary and not general results, since all the experiments reported up to now involved the same kind of task: to choose among prospects. With regard to this, to reach to more robust results, further empirical investigation may be necessary.

Finally, it should be always considered that every extra amount of instruction and incentive mechanism complexity has its cost in terms of subjects' patience and attention.

[10] The only case in which Beattie and Loomes (1997) found a significant difference among the treatments is when people had to choose among compound lotteries, instead than among simple prospects.

## 4. Subject Pool and Sample Size

One of the much debated issue in making experiments in social sciences is the general use of college students as experimental subjects. The reasons for which the researchers usually observe students' behaviour to test their hypotheses are numerous.

First of all, the reason is matter-of-fact: students are easy to recruit. In fact, usually researchers find their experimental subjects by means of notices around university campuses, or, more simply, inviting students to volunteer during lectures. More recently, making use of the new technologies, some online recruitment systems have also been developed (e.g., Greiner, 2004). In this way, it may be also better achieved an indirect control by randomisation in recruiting subjects.

Moreover, students are convenient subject pool since generally quick on understanding their task in the experiment.

Third, and probably more importantly, the particular low opportunity costs make students the ideal subjects to make experiments: generally, people participating in an experiment are paid in manner to matching, on average, their hourly wage rates, besides that according to the decisions they (and other participants) make, and consequently it is clear as much convenient is to use students as subject pool compared to professionals.

Another reason is of empirical nature: the very first experiments that compared students' behaviour with other subject pools found virtually no significant differences. Interestingly, Lichtenstein and Slovic (1973) replicated the preference reversals phenomenon with regular players at Las Vegas casino, whereas Burns (1985), comparing students' and businessmen's behaviour in an auction, found that professionals showed even lower learning. Also Dyer et al. (1989) addressed the same question comparing students and executives, finding no substantial differences.

More recently, in reviewing some results in labour market experiments, Falk and Fehr (2003) can conclude: "subject pool differences may be a real issue. However, the studies also show that the different subject pools do not behave in fundamentally different ways. [..] Thus, although there are some quantitative subject pool effects, the qualitative pattern of behaviour were rather similar across the different subjects pools".

Probably, a latest paper that, at least in part, better tries to face the issue, is due to Guillén and Veszteg (2006). In this paper, the authors look for demographic effects that may be the source of an important subject pool bias if not properly accounted for. At this aim, they consider a particularly large data set from different economic experiments, where no less than 70 per cent of subjects belongs to student subject pool. They find that age and gender may have some effects, whereas experience and education do not. Notwithstanding this, demographics seems to explain less than 4% of variability observed in monetary payoffs. Consequently, subject pool effects may be not so effective in explaining differences in payoffs, and, thus, in subjects' behaviour. Also Carbone (2005) finds no link between strategic behaviour and demographics in a life-cycle consumption experimental task. However, Harrison et al. (2005) found different behaviour between (adult) students and the other part of the adult population in Denmark.

The issue dealing with subject pool bias is particularly important especially in the light of *external validity*, that is, if we want to be confident in transferring experimental results to real life. We will come back to the question in Section 7.

Another related issue is the fact that subjects, independently of the subject pools used, actually volunteer for the experiment (for a very first examination of the issue, Kagel et al., 1979. For a more recent investigation, Jones and Seaman, 2003). This could create the so called *self-selection bias*, that, in turn, may distort the experimental data. For example, in the public goods game in which we want to test for subjects' willingness to cooperate, the results we obtained could overestimate the degree of cooperation itself, since the subjects that have decided to participate may have stronger social preferences than the ones do not. Actually, this is even a more serious problem, since it involves an ethical dimension. Even though we can rather easily test for the robustness of our results obtained from a student subject pool comparing them with a different subject pool, nevertheless it is rather difficult, instead, to force people to take part in an experiment, because of ethical or political constraints.

### *4.1. People as the main feature in experiments*

The idea that someone can have formed of experiments in social sciences, and actually it is true, is that the difficulties in implementing experiments in this field arise just for the fact that experimental economics deals with 'people'. With regard to this, on the mark seems to be Sugden's definition for experimental economics: "theory with people in it" (2000).

Dealing with real people in the laboratory without doubt has been and is one of the most tricky characteristics of experimental economics, and very likely this has had its part in slowing down the awareness that it could be really feasible to realize experiments in economics at all.

#### *4.1.1. Experiments versus simulations*

This peculiarity should be used also to overcome a common misinterpretation, given that sometimes experiments are mixed up with simulations, especially by 'outsiders'.

Actually, the feature that more can help in distinguishing experiments from simulations lies just in the employment of real people in the former, and of 'computerized' agents in the latter ("computer simulations are useful for creating and exploring theoretical models, while experiments are useful for observing behaviour", Roth, 1988, p, 1000).

For simulations, or better, in Agent based Computational Economics (ACE) approach (Tesfatsion, 1997, 2001), researchers usually develop software in which they programme as the simulated agents are supposed to behave in a given scenario according to the particular hypothesis they want to test, and then observe the attainment or not of a equilibrium state (Arthur, 2006. In experimental economics, for a robustness test of the double auction with zero intelligence traders, Gode and Sunder, 1993). More recently, in order to test some learning theories, and also to make simulations more realistic, more and more sophisticated software have been developed that allow simulated agents to change their strategies according to changes in environment and/or interactions with others (Holland 1995, 1997)[11].

On the contrary, in experiments, researchers put real people, with their own background, emotional states, and so on, in a laboratory, where they face real decisions, and generally earn real money. More and more often, the focus of investigation is not on equilibrium properties, but on behaviour itself, and this has very often produced so astonishing results that no programmer could have ever conceived of (for a review of results in individual decision making, see Camerer, 1995).

### *4.1.2. Framing and labelling: the 'psychological interferences' behind experiments*

It is easy to realize that, since human beings are the object under investigation, the variables to be taken into account are theoretically in so large number that it would be virtually impracticable to control for each of them (even the colour the laboratory is painted may have an effect). This is due to the fact that every single change in the experimental environment, even if very little, may cause large changes in behaviour.

Especially for this reason, it has been established to pose the experimental setting to subjects as more neutral as possible, in order to prevent that some 'psychological interferences' spreading in experimental settings. The possible psychological processes and the related shrewdness to be undertaken can be countless to list, so in this work I would content myself with spending some words about framing[12] and labelling.

Experimenters usually prepare their experiments taking care of providing no cue about what may be the aim of the work and of conveying no misleading or ambiguous or loaded meaning to any aspect of experimental environment and institution, but rather of using abstract or context-free terminology, in order to not contaminate behaviour or to not lose control (Smith, 1976: "it may be preferable *not* to embellish the instructions with well-intentioned attempts at "realism"; emphasis in the original). For instance, it has become classical as example the prisoner dilemma experiment: the recommended practice is to not label the two available strategies as 'defect' or 'cooperate', but rather as 'strategy A' and 'strategy B', or 'strategy 1' and 'strategy 2'.

Probably, one of the first to report the phenomenon was Schelling (1957). In his experiments, he noticed a higher coordination rate when experimental subjects had the opportunity to recognize some "prominent" outcomes among the available ones compared with the cases in which no prominent outcomes existed.

[11] "Agent-based computational economics is the computational study of economies modelled as evolving systems of autonomous interactive agents. ACE is thus a specialization to economics of the basic complex adaptive systems paradigm" (Holland).

[12] In the psychological literature, we observe *framing effects* when different ways of describing the same choice problem change the choices that people make, even though the underlying information and choice options remain the same (Tversky and Kahneman, 1981).

Nevertheless there are also authors that rather strongly oppose the practice of an abstract-free experimental set-up. They argue that in this way we misrepresent the environment in which people usually take decisions in real life and consequently we concur in confounding them, distorting the decisions themselves. If people have no idea about the environment in which they are acting, we cannot know if they are imposing a their own context on the experiment. If they actually do this, we may have completely lost the control. Substantially, this is the opinion supported by Harrison and List (2004). They think that field referents can facilitate subjects in surmounting confusion about the experimental task, avoiding lack of understanding, and in drawing on some specific heuristics from the field to solve the decisional problem in the lab. At the same time, they note that not adequate choices of field referents could provoke uncontrolled psychological motivations.

Except for the cases in which framing effects themselves are under investigation, it may be preferable that experimenters avoid to provide any cues to experimental subjects. Harrison and List argue that "the choice between an abstract script and one with filed referents must be guided by the research question".

One reason for avoiding framing is simple: not transferring any cues in experimental settings preserves the replicability of experiments, even across different experimenters. This can be useful also to have experimental data that can be related to some 'conventional' and clear reference point. To make an example, just consider what happens with blood tests. The convention is undergoing them on an empty stomach. This is not just because this is by far the best way to make them, but it is simply because clinicians set the benchmark blood values according to this reference point. At the same time, since economic theories to be tested assume the *description invariance*, we may agree on a particular convenient reference point in order to test it. Personally, I see no more general reference point than this. At the same time, I agree that it should be convenient a broader testing involving environments framed in different ways. We refer to Section 7.4 for a further discussion about the more general issue of *context effects*.

#### *4.1.3. Anonymity and the demand-induced effect*

In order to avoid as much as possible some complex social phenomena entering in the experimental environment, that even in a laboratory setting could prove to be difficult to control for, another measure usually undertaken is ensuring anonymity.

Generally, anonymity extends to many features of the experiment. For instance, except for the cases in which the effects of communication on behaviour are explicitly under investigation (with regard to this, there is a rather wide literature that it is difficult to completely account for[13]), generally the participants are strictly not allowed to communicate among them (for this reason, very often in the lab each participants' seat is provided with some kinds of partitions), and they do not know the subjects they interact with during the experiment. Moreover, if experimental

[13] For example, in the experimental literature it is now known as stylized fact the finding that the communication among participants raise the degree of cooperation in a public goods game (e.g., Isaac and Walker, 1991).

subjects are monetary rewarded (see Section 5), the common practice consists of paying them at the end of the experiment strictly in private.

In order to reduce social distance, the implementation of computerized treatment can prove to be very useful. At this aim, specific software have been developed (an experimental software widely used is z-Tree, Fischbacher, 1999).

It has been investigated also the effect of double anonymity, that is, as the behaviour changes when the participants do not interact even with the experimenter (Hoffman et al., 1996). With regard to this, the *demand-induced effect* has been detected. According to this, subjects may try to understand researcher's own goal and then endeavour to go in that direction or to behave in the opposite way, depending on subject's attitude towards the experimenter. Clearly, if nothing is done to try to prevent this, experimental data may be seriously contaminated.

Levitt and List (2007), for example, have provided empirical evidence in a social preferences experiment indicating that factors like the nature and extent of scrutiny by others, the context in which a decision is to be taken, and the ways to select participants and tasks are important and that differences between laboratory and real experience diverging in one or more of these factors impact on results obtained in the lab and their comparison with behaviour observed in naturally occurring settings. In other words, the observation is itself a problem in experiments.

Again, the fact that in all these cases the investigator itself (the experimenter) is part of the issue investigated raises very sensitive problems that can be prevent the achieving of objective results. Moreover, in some experiments experimenter may be considered by participant not only as a neutral observer, but also as an extra player, whose presence can alter participants' behaviour (for an example, the reader can refer to Harrison and Johnson, 2005).

### *4.2. Sample size and statistical inference*

Related to the issues involving student subject pool, as discussed above, there are some statistical problems. Indeed, since very often experimental subjects are only students, it may make problematic any kind of inference, since the data could be not representative enough. Moreover, usually no more than one/two hundreds subjects are involved in an experimental study, but often even less.

As regards the first point, critics may generally be right when the research question directly involves decisions that are clearly different depending, for example, on age. Just to cite an example, we can refer to retirement plan choices. Obviously, it is rather unquestionable to believe that decisions taken by 'real people' that have spent more than half of their life at work are different from decisions taken in a laboratory by students. Very likely, they have never worked up to that moment, and still do not have a clear idea about what job they intend to embark on in their future. However, it is still a matter that depends on own work's goals. On the contrary, if we want to tackle an explorative work about time preferences in general, we can affirm that, beyond age and education, factors that undoubtedly are important in shaping subjects' decisions, nevertheless there are further idiosyncratic factors that influence subjects' choices and that may be reliable caught by a student subject pool, being representative enough of the entire population.

In conclusion, I believe that, whenever the socio-demographics characteristics are proven to be no so effective in differentiating behaviour, the problem is not so serious as generally considered : "the "problem with students" is the lack of variability in their socio-demographic characteristics, not necessarily the unrepresentativeness of their behavioural responses *conditional on their socio-demographic characteristics*" (Harrison and List, 2004; emphasis in the original).
As for the second point, generally the experimenters reply that the problem with the small numbers is not effectively a real problem, since easily surmountable (usually they use small size sample or because of financial constraints or of practical implementation). Actually, it is sufficient to replicate the same experimental design with fresh subjects to enlarge the sample size[14]. At this point, it is important to note how much imperative is having experimental designs that have incontrovertible methodological standards in order to improve replicability. If so, it is possible for researchers to have their experimental laboratories even kilometres and kilometres far, and still be sure to replicate exactly the same experiment and to have no doubt in drawing statistically reliable conclusions[15].
Finally, we can just mention another potential statistical issue. The fact is that classical statistical procedures assume that each unit of observation is an independent draw. Clearly, this is hardly true in experimental economics. Hence, much will depend on the experimental design and on the experience experimenter has in handling these problem. As a further solution, usually it may be recommendable to implement also non-parametric statistical tests (Conover, 1980; Brunner et al., 2002).
Anyway, concerns about data analysis are broader and would deserve a deeper investigation. This would turn to be a fruitful field of research in the future. Someone has also already coined a specific new term: "experimetrics" (e.g., Bardsley and Moffat, 2005).

## 5. Incentives

One of the features that more concurs in distinguish experimental practices among economics and psychology and that fuels much debate is surely the use of monetary incentives as rewards for participants at the experiments[16] (Roth, 1995, p. 86; Hertwig and Ortmann, 2001. For the idea that psychologists have regarding this, as classical references we can cite Kahneman and Tversky, 1979, p. 264-265, or Thaler,

[14] However, often experimenters complain of too low opportunities to have published follow-up works (Rubinstein, 2001).

[15] For example, data collected in different experiments could be aggregated, provided that some econometric devices are taken into account, as the inclusion of a dummy variable for each different experiment. If replicability has been achieved, these variables should be not significantly different from zero.

[16] Curiously, the first author advocating the use of monetary incentives was a psychologist, Siegel (Siegel and Fouraker, 1960). In this work, they were the first in noting the effects of the presence and the size of monetary payoffs on behaviour, meanwhile Smith (1962) reported just one treatment in which real monetary payoffs were awarded.

1987, p. 120). We should remember that this was also in the Wallis-Friedman critique (1942).

Several and different are the reasons for which experimentalists in economics choose to pay participants at their experiments. We will try to present the two most relevant ones.

As recognized by Hertwig and Ortmann (2001), it is likely that the primary rationale for financial incentives is that for economists it is quite natural to recreate in the laboratory the same framework as in standard economic theory they want to test, generally theories built on maximization assumptions. In this sense, given that economic agents are seen as maximizing their own utility, or profit, theoretically, there is no reason in believing that participants' actions are not driven by the same incentives in the laboratory.

Another common alleged reason for financial incentives is the belief that rewarding participants has an important role in reducing variance in experimental data. The rationale behind this is that the financial incentives would induce more effort (or, in the cases of particularly time-spending or tedious experiments, they would help in counteracting boredom) and would help in maintaining concentration. Consequently, they would produce more statistically reliable and more informative data. This belief has been supported with a survey study carried out by Smith and Walker (1993), in which they did find that "in virtually all cases rewards reduce the variance of the data around the predicted outcome".

It is worth citing another source of critique, that is, the fact that generally incentives in laboratories are far lower that the ones in the real world. While considering the issue important, I do not develop it further here[17].

### *5.1. Smith: The induced value theory*

Most of the relevance and the "necessity" of monetary incentives (for example, Camerer and Hogarth (1999) found that virtually no experimental work without payments has been published in a top-rank economic journal between 1970 and 1997) could be imputable also to the influence Vernon Smith's argument has had among the experimentalists and that he stigmatized in the "induced value theory". In particular, in a couple of very influential papers (1976, 1982), he advanced the now so called *precepts* of experimental economics, that is, in author's opinion, are the sufficient conditions for experimental validity.

His argument is that the way to obtain control in experimental domain goes through inducing prescribed monetary value on actions. In order to achieve this, some requirements are to be met in reward medium: a) *non-satiation*: this is no more than a reformulation of the monotonicity assumption, whereby more money is preferred to less; b) *saliency*: according to this precept, the incentives in order to be motivationally relevant have to be in a clear and direct relationship with the performance achieved in the experiment[18] (see also Wilde, 1980); c) *dominance*: in

[17] As far as the level of rewards is concerned, I can cite in the experiment by Cameron (1999). Since very often the results about the ultimatum game (Güth et al., 1982) were criticized on the ground of low stakes, Cameron replicated previous experiments in Indonesia in which the pie to be split was up to three times the average monthly expenditure of a participant.

order to not lose control, in presence of subjective not monetary costs related to the decision process, it is important that the reward structure is sufficiently high in order to more than compensate for them (as we have already seen, this is one of the reason for which as subject pool students are usually used, for their particularly low opportunity costs); d) *privacy*: in order to rule out any interpersonal comparison that could distort the desired induced valuation (with regard to this, see the literature developed about fairness (Fehr and Gächter, 2004), it is advisable to keep the information about other participants' payoffs incomplete, in the sense that any subject is given only information about his/her outcome. This principle may underlie the established practice to pay subjects privately.

If these requirements are met, we have been able to put in existence an experimental microeconomy, that is, a real live economic system, "where real people earn real money for making real decision". Anyway, Smith adds a further precept, the parallelism, which we deal with below (Section 7), referring to the external validity of the experiments. As we will see, this is probably the hardest attack to experimentation in economics.

### *5.2. Harrison: The 'marginal payoff critique'*

In a famous and frequently cited paper, Harrison (1989) criticized not the use of monetary rewards in themselves, but rather the way in which the payoffs are computed from the actions observed during the experiment.

The critique is that very often, whereas the deviation between the observed action and the optimal predicted action seems to be so large such that someone can be inclined to reject the theory being tested, nevertheless the difference in the foregone expected payoffs is rather insignificant. According to this, especially when there are not negligible decision costs, the subjects could not make much effort to reach the optimal choice, given that the gain in expected payoff is not so much as to offset the extra effort. Consequently, on the margin, there should not be so much difference in choosing actions that deviate from the optimum, even if these actions lead to completely different theoretical conclusions: the cost of misbehaviour (measured in foregone expected payoff) appears to be no so serious. On the contrary, from a scientific point of view, the cost associated to the rejection of a theory could be rather large. Harrison concludes arguing that in these kinds of experiment the precept of dominance would be often not satisfied.

As expected, this paper raised debates and comments, collected altogether in an issue in *American Economic Review* (vol. 82, number 5; 1992).

In their comment, Friedman (1992) and Kagel and Roth (1992) agree in thinking that Harrison overstated his case, rejecting his assertion whereby it would be more natural to test hypothesis in payoff space rather than in action space, but at the same

---

[18] However, very often, besides the payoff earned during the experiment, a flat *show-up fee* is scheduled. Various the reasons for this. For example, since the risk of giving up is always present, often a number of participants larger than necessary is recruited. So, supernumerary people that will have no opportunities of participating will be compensated in any case. Sometimes, when experimental subjects can incur in some losses (but generally campus regulations do not allow students pay out of their own pocket), the show-up fee is scheduled just to cover the losses.

time appreciate his attempt to reconsider the cost of deviations as a useful robustness check, as well, more generally, his effort to bring to attention these kinds of issues.
Instead, Cox et al. (1992; we should note that a series of their studies were the original attack in Harrison, 1989) defend their own work, asserting that they have always been interested in motivational questions and showing that the measure suggested by Harrison is not an informative aid in their auction experiments[19].
On the other hand, in their contribution, Merlo and Schotter (1992) emphasize that the strength of Harrison's critique is, in fact, rather restricted. They show how his criticism holds only for a particular subset of experiments (in which the task is neither too easy, nor too difficult, and in the cases in which subjects have good opportunities to learn) and depends also on the particular participants' attitude towards the experiment itself. Moreover, using experimental data, they demonstrate that the shape of theoretical payoff function (steep, as Harrison suggests it should be, or flat) in fact seems to have a little impact on a large fraction of subjects (only 35% of subjects consistent with Harrison's critique).
Finally, in Harrison (1992), the author restates the critique in wider terms. He cleverly presents as an example the BDM mechanism (for further details, see Section 3.1.2). He supposes a subject reports as a selling price a value of 5 cents less than her true value, showing as this false report results in a foregone expected payoff of 0.015 cents only. However, as it is simply to note, a rational player that prefers more to less (and Harrison himself reminds the nonsatiation as precept) should still prefer the true telling to false reporting, until decision costs are completely negligible, as it is likely in so straightforward tasks.
In the rest of the paper, Harrison strikes back each of the previous comments. From his reply a series of interesting methodological reflections can be derived.
This has been a clear example of how the dialogue among experimentalists is a so important and necessary path to be undertaken for the developing of experimental research programme, so that these kinds of symposia cannot be more than appreciated.

### *5.3. To pay or not to pay?*

We can surely affirm that, if there are valid theoretical reasons to motivate experimental subjects with monetary rewards, to what extent they really matter remains essentially an empirical issue.
Probably, the first work that tried to make the point was Smith and Walker (1993). As we will already noted, their survey (31 experiments) substantially leads to the finding that increased payoffs have a two-fold effect: in increasing the consistence of experimental data with rationality, and in reducing the variance in the data. They conclude asserting that whenever the data fail to reach the rational models prediction, this may be attributed to the low opportunity costs of deviating from theoretical predictions.

[19] In fact, the methodological digression in Cox et al. (1992) is more wide-ranging and involves accurate details about the design and the analysis of auction experiments. Even if this is an interesting issue, nevertheless this is clearly beyond the scope of this paper.

More structured the conclusion Camerer and Hogarth (1999) reached in their wider survey (74 experiments). They also observe that higher payoffs lead to reducing variance, but additionally they show that the effects of higher incentives are rather complex. Actually, the fact that monetary rewards are at stake and their actual amount seem to affect mean results for some kinds of tasks (in particular, in judgement and decision tasks) in which at higher payoffs correspond a better performance, but for the most part rewards are showed to have almost no effect on other kinds of tasks, for a series of reasons. Moreover, there are cases in which monetary rewards seem even to hurt, for example when higher payoffs lead to exacerbating the use of a heuristic and consequently to observing more instances for a bias. Another finding is that higher payoffs may reduce *self-presentation effects*. This term usually is used to intend that behaviour chosen in order to make a good impression from a social desiderability perspective, and it can be seen as a sub-case of demand-induced effect. As an example, we can cite the dictator game (Forsythe et al., 1994). In this game some subjects, while recognizing as rational the strategy to give nothing to the other player, nevertheless they would choose to give her a share of their own endowment in order to present themselves as generous.

Of some relevance also the survey carried by Hertwig and Ortmann (2001), that reaches substantially the same results.

More recently, there have been two studies that have used experimental methods to investigate the methods experimenters themselves use, what I have defined *meta-experiments*. In Holt and Laury (2002), in a simple lottery-choice experiment, is found that financial incentives practically matter, since higher the magnitude in real payoffs, higher the degree of risk aversion, but the same thing does not occur under hypothetical payoff treatment. Conversely, Tenorio and Cason (2002), considering the Wheel Game performed in a TV programme and its replication in the laboratory, found that the deviations from the theoretical predictions are largely independent of the stake levels of the game.

Finally, Read (2005) asserts that monetary incentives do not assure the achievement of the three effects that are usually ascribed to: cognitive exertion (Wilcox, 1993; Harrison, 1994), motivational focus, and emotional triggers, and even that there may be no basis at all for requiring the use of monetary incentives, but that these effects can be achieved in other ways. For instance, Camerer and Hogarth (1999) advance experience as substitute for financial incentives.

So far, we can affirm that there is a nearly general agreement on the fact that using financial incentive produces more reliable data given the reduction in variance empirically observed. Notwithstanding this, there is no agreement on the fact that the attainment of less variable data is in any case sufficient to require anyway the use of financial incentives to motivate subjects. It should be noted, however, that at this point a new methodological issue would arise: sometimes, if no financial incentives are scheduled, it could be not easy to distinguish survey data from proper experimental data, if we use monetary rewards as divide between them (Friedman, 1988).

In conclusion, being fundamentally an empirical issue, it is clear that further works, in the same spirit of Holt and Laury's (2002), are more than useful to try to settle the issue and in order to adopt a common standard in future experiments. For instance, it is now rather established that financial incentives matter particularly  for decision tasks, but not in other kinds of tasks. Probably, in the future we will have a common protocol in which rewards are necessarily required for these tasks, but only optional for other ones. Eventual possibility if and only if further empirical evidence will be collected.

## 6. Training and Learning

### *6.1. Instructions*

Once subjects have entered and taken their seat in the laboratory, the first thing generally done is instructing them about the task they will be asked to perform during the experiment[20].

With regard to this, experimenters have developed rather widespread rules across experimenters and experimental laboratory centres.

Firstly, instructions are required to be kept as much simple as possible and to be framed in neutral words (see Section 4.1.2). For this reason, it is a useful practice to circulate them among colleagues before the experiment in order to collect comments and suggestions to improve and simplify them. Indeed, researchers designing the experiment may fall into the so called *knowledge's curse* (Camerer et al., 1989). In this particular context, it means that often, who is very involved in a research project, can take for granted each single detail that, although insignificant at all appearance, could be very significant from participants' perspective. Some researchers told they sometimes experienced problems in explaining how probabilities actually work even to students with statistical background.

Secondly, instructions are usually intended to give subjects only the relevant information to perform the experimental task. It is common practice to avoid to provide them experiment's goal in order to not produce undesiderable induced-demand effects (see Section 4.1.3): "The instructions make clear the opportunities available to the subjects, but the motivation is supplied by the people" (Plott, 1982, p.1490). Moreover, this practice proves to be essential in order to not contaminate behaviour with extrinsic motivation, so that the only (or, at least, the most prominent) motivation for participants remains the monetary reward. In this perspective, this way of conceiving the instructions is another manner to attain control in the laboratory.

Thirdly, usually experimenters read the instructions aloud: in this way it is assured, at least theoretically (we cannot avoid that some subjects do not pay attention at all even during instructions reading), that every participant is given the same amount of information before the experiment starts, establishing a scenario of *common knowledge*[21] among them. Indeed, if experimental subjects are told to read the

[20] In experimental economics is also quite common the fact that authors add instructions to their paper or make them promptly available on request. This practice also concurs in reducing ambiguity and in enhancing replicability.

[21] I realize the term common knowledge is used in a very specific way in game theory (Gibbons, 1992). In this context, I have used this expression just to mean that each subject, at the least, receives publicly exactly the same amount of information, that could be considered as a part of her personal

instructions on their own, we cannot say with certainty if participants actually read them. Indeed, if a no negligible fraction actually do not read the instructions, we may obtain misleading data. That instructions are read aloud do not exclude that subjects are always provided with some scripts: in this manner, they have always the opportunity to read again instructions if not clear enough in some parts or at the moment in which are actually relevant during the experiment. This practice also distinguishes experimental designs in economics and psychology (Hertwig and Ortmann, 2001).

Clearly, sometimes the experimental design forces experimenters to not read aloud the instructions. For instance, this can happen in experimental games, where there are two kinds of players with completely different roles and strategies to be played, or in the cases in which subjects are provided with different private information. Consider the case of oral double auction: as explained in Smith (1962) and Plott (1982), the payoff functions are induced by individual redemption values and cost schedules, for buyers and sellers, respectively, and the key-variable design is that everyone knows nothing about others' schedules.

The standard practice should be the public reading of written instructions, unless opposite reasonable motivations are at stake.

### *6.1.1. Understanding the instructions: training*

Usually, experimenters do not settle for reading instructions aloud and providing subjects with scripts. After reading, they also give subjects the opportunity to ask as many questions as they want and are often encouraged to do this. The choice of answering privately or publicly depends on the nature of the questions itself. Actually, this may be the optimal opportunity to overcome some deficiencies in instructions not detected up to that moment.

Moreover, generally a more direct test to control if subjects have entirely understood the instructions is conducted. Several the methods used at this aim, and generally one rules out another one. Up to now, no structural methodological debate has come out about this, but there is some disagreement among different experimental centres.

One method consists of running a written test, known as *control questions* in the experimental literature, just after the reading of instructions. Generally, participants are required to fill in a paper or an electronic form with some questions about the available strategies for playing the game or about some examples regarding how to compute the payoff. The answers are checked before the experiment starts so that the experimenter has the opportunity to clarify further the design. The main shortcoming of this method is that it could create an *anchoring effect*, i.e., the subjects may consider the examples in the questionnaire having a particular meaning so that they could be tempted to use them as reference points for the entire experiment, or, as regards the strategies presented as illustration, decide to play according to them or against them.

---

"endowment" in order to be enabled to run the experiment. I am not the only one to use this term with this meaning. For some instances, Smith (1994, p.120), and Friedman and Sunder (1994, p. 212).

Another method often used consists of running some *trial periods*. In this case, subjects play exactly the same game that they will play in the proper experiment for some periods, with the only difference that in these very first periods they will receive no rewards for the decisions taken and they are informed about this in advance. This is the reason that leads Friedman and Sunder (1994) to name these as "dry-run periods". Also this method has advantages and disadvantages. The principal advantage is that subjects have the opportunity to familiarize themselves with the game (as can be noted, this is not possible with the control questions method). This could be especially important for computerized treatments, in which some may present problems in interfacing with computers. On the other hand, the usual criticism against this method is that, given the data referring to the trial periods are not considered in further analysis, and given the most part of learning usually takes place just in these very first periods, as a result we would have experimental results show more consistency and rationality than if not.

Finally, also for obviating these deficiencies, while maintaining the advantages at the same time, a different method that we could term "dummy practice treatment" has been developed. This is substantially equivalent to running trial periods, only that the parameter values in these periods are distinct enough from the values used in the proper experiment, in order to avoid any possibility for anchoring effects and to develop some training about the task in general.

As above underlined, up to now no work has been done to detect if these pre-experiment training methods have any systematic and/or different effects on following experimental results. This will be another question in which the use of meta-experiments may prove to be very useful.

### *6.2. Stationary replications and experience*

As it will be already clear reading the preceding section, usually subjects performing an experiment in economics are required to complete the same task more than one time, for several "periods", with no changes occurred in the meantime. In some parts, this feature is referred as the "Groundhog Day[22] replication" (Camerer, 1997). For opposite appraisals of this procedure, see Loewenstein (1999), that substantially criticizes it, and Binmore (1999), who, instead, is in support of it.

The rationale to do this is that during the first task performance people could be confused, in case of computerized treatments, they could not have properly understood the roles or how the experimental software properly works, such that data referring to very first periods may be misleading, until enough experience has not been acquired. Indeed, beyond the mere repetition, usually subjects are provided with *performance feedback* of their previous decisions.

The reason for repeated trials is also that people have usually several opportunities for learning in the real world, and these opportunities are recreated, to some extent, in laboratories, with replications of the task.

[22] "Groundhog Day" is an American movie directed by Harold Hamis (1993), in which the main character finds himself living the same day over and over again.

Anyway, and maybe more importantly, it is the general interest of economists in reaching equilibrium situations that suggests repeated trials.
As an empirical investigation whether replications of the task actually improve the consistency of the decisions taken, we can refer to Hey (2001). In this study, the author is primarily interested in investigating the nature of noise that is generally observed in the experimental data. His results are supportive of individual heterogeneity: indeed, there are subjects whose number of inconsistencies decline over the five repetitions of the same task, for which we can affirm that actually the noise is temporary, and there are subjects whose inconsistency rate remains substantially constant during the experiment, so that repetition does not seem to improve consistency.
Furthermore, given the repetitions, researchers have also the opportunity to observe how fast learning processes in human beings occur for tasks at different complexity levels. With regard to this, we should mention that a wide literature is spreading about learning theories (Roth and Erev, 1995; Camerer and Ho, 1998; Erev and Roth, 1998, just to cite a few) and that the observation of the behaviour in experimental laboratories is undoubtedly a favourable place where these tests can be conducted in a controlled environment.
Another debated issue and a common source of criticism is the role and the employment of experienced or unexperienced people in experiments, issue that to some extent is intertwined with subject pool bias (Section 4). However, further and specific methodological questions may arise. For example, it could be necessary to control directly for experience, introducing or a dummy variable (0 = no experience; 1 = experience) or a discrete or continuous variable (e.g., number of participations in previous experiments or number of hours spent in previous experiments, respectively) in econometric analyses. As an alternative solution, we may considering the employment of experienced vs. unexperienced subjects as treatment variables on its own (Friedman, 1988).
To date, however, it seems that no general agreement has been reached, but these decisions are generally left at researchers' discretion.
Before leaving this issue and turning to a probably much debated one, I will give just a short account of a theory that aroused interest in recent years: the Discovered Preference Hypothesis (DPH) developed by Plott (1996), and that, to some extent, is related to experience with a task. DPH theory is based on observation and tries to explain how "with practice and experiences, under conditions of substantial incentives, and with the accumulating information that it is obtained from the process of choice, the attitudes stabilize in the sense of a consistent decision rule, reflecting the preferences that were discovered through the process" (p. 228). According to this theory, preferences are stable and 'innate', but they need to be *discovered*. With regard to this, it is clearly contrasting with the Constructed Preference Hypothesis (CPH), developed among others by Slovic (1995), whereby people *construct* different preferences relative to different environments they face each time. These two theories have offered new directions of how interpreting experimental data, and will not fail to suggest new lines of reserch in the future, as they have already done (for an example, see Hoeffler and Ariely, 1999).

### *6.3. Deception*

In experimental economics community, deception is almost generally considered a taboo (Hey, 1991; Ledyard, 1995, p. 134: in his passage, honesty in conducting experiments is famously likened to a public good). Probably, the clearest explanation of the rationale behind this stance can be found in Davis and Holt (1993, p. 23-24) and this point of view is shared nearly by the majority of economics experimentalists:

> "the researcher should be careful to avoid deceiving participants. Most economists are very concerned about developing and maintaining a *reputation* among the student population for *honesty* in order to ensure that subject actions are motivated by the induced monetary rewards rather than by psychological reactions to suspected manipulation. Subjects may suspect deception if it is present. Moreover, even if subjects fail to detect deception within a session, it may jeopardize future experiments if the subjects ever find out that they were deceived and report this information to their friends" (emphasis added).

The main concern about the deception is clearly the potential loss of control in which the experimenter could incur if only participants have even the least doubt about whatever procedure employed by experimenter.

The key word all the debate spins around is the relevance of building and maintaining up a *reputation*, since economics experimenters fear negative spillover effects to spread if only some of them implement deceptive practises.

Another illuminating illustration against the deception can be found in Hey (1998). Particularly, in this paper, it is also clarified the difference "between not telling subjects things and telling them the wrong things. *The latter is deception, the former is not*" (italics in the original, p. 397). This is an important methodological point that should be always kept in mind. Actually, subjects hardly ever are told the experiment's goal, but this is never conceived as a deceptive device.

Probably, the only one that strikes a discordant note is Bonetti (1998a, 1998b). In his papers, he states that there are no supportive evidence against the use of deception and, quite the opposite, in some circumstances the deception could lead to potential benefits.

To some extent, an intermediate position between these two extremes is expressed in McDaniel and Starmer (1998). They consider the use of deception as potential treatment variable in some very limited cases, but are very cautious in promoting a general use of deceptive practises.

Ortmann and Hertwig (2001), after reviewing relevant research in psychology, conclude that "prohibition of deception is a sensible convention that economists should not abandon", since also in psychology experiments deception is proven to affect subjects' behaviour.

My personal judgement is that there are no reasonable grounds on which one can sustain the use of deception, even if there would be no empirical evidence, as Bonetti asserts, that proves as harmful the deception could be. It makes no sense to tell subjects something and then implement another one. From a methodological point of view, whether a researcher acts in this way, then, how will she interpret the experimental data she obtains? Referring to what the subjects are told at the beginning of the experiment, or referring to what they could have thought or discovered during the experiment? Indeed, as an experimenter, it is essential to have an idea how participants will interpret the game, otherwise she could lose control completely. In addition, to what extent can the experimenter learn something about the opinion that the subjects could have formed during the course of a deceptive experiment? At the end, running very badly designed experiments, what is she testing?

Someone supports some deceptive practises, asserting that in this way would be possible to implement experimental design otherwise hardly feasible. The example commonly cited is the experiment where in a public goods game subjects were told to play with 100 subjects, but actually there were only five (Kim and Walker, 1984). These problems can be overcome also thanks to particularly talented experimenters and their inventiveness, and also thanks to new technologies. For example, nowadays the same 100-people public goods game could be replicated in several laboratories connected in real time by Internet.

## 7. The "triangular relationship": Experiments, theory and world

In this section I will try to sketch some relevant issues and then to reach to a (provisional) conclusion. At this aim, some papers published in a symposium in a recent issue of *Journal of Economic Methodology* (issue 12, number 2; 2005) may be useful and may shed some light on the matter.

As regards the related question associated with the peculiarity of human beings being the object of investigation, see the Section 4.1 above. However, here I can add just a couple of remarks. First, in some parts the attack against economics experiments is grounded on the basis that human behaviour is unpredictable or the relating explanations too complex to investigate. However, since it should be clear that this argument is not only against experimentation in particular, but more generally against the predictive role of economic theories, I will not go into it further. Second, and probably a most serious issue, there may be the doubt that since human beings are conscious of participating in an experiment, this fact could be sufficient in its own to affect their behaviour (in literature, it is known as "Hawthorne effect"; Landsberger, 1958). It is not difficult to note that this effect is a much wider than the "demand-induced effect". However, this is a problem related to experimentation in every social science, and, furthermore, follow-up research has shown almost no evidence for this effect (e.g., Jones, 1992).

### *7.1. The "artificiality critique"*

As we already noted (Section 2), the first critique, in chronological order, against experiments in economics was the one advanced by Wallis and Friedman (1942), referring to the alleged artificiality of experimental situation.

At present, more than sixty years after, this attack to experimental economics is still relevant. Indeed, if there is a roughly wide consensus among economics experimentalists about the practices to be implemented, so that the *internal validity* is generally preserved, more critical the situation regarding the *external* (or ecological) *validity*. That is, provided that experimenter follows the "common practices" broadly shared by profession, namely, the ones that consider primarily important the employment of adequately motivated subjects, put in a controlled environment, where they are not deceived, but rather are given the opportunity to understand properly the task they are required to perform and given the opportunity to acquaint themselves with it, generally the experiment so designed is considered internally valid, in the sense that the causal relations derived from it can be considered as fundamentally true and replicable. More problematic, instead, to assess to what extent we can "export" the laboratory results to the real world. This is a not marginal methodological issue. Indeed, if results achieved in the laboratory are found to be "artificial", not being representative of people's behaviour in the real world, then these results would be of limited interest, or even meaningless. Even the best practices implemented in the laboratory would lose any significance. With regard to this, Loewenstein (1999, F33) considers the lack of external validity "the Achilles heel of all laboratory experimentation".

Schram (2005) provides a deep evaluation of the artificiality critique, reviewing all the relevant literature and pointing out the tension, or, even, the trade-off between external and internal validity. Also, he strongly relates the question of external validity with the specific goal an experiment is aimed to. In particular, when experimental research is aimed at discovering behavioural regularities ("searching for facts" category in Roth's taxonomy (1995)), the urgency for external validity is more pressing. Given an increasing number of theories has been developed on the basis of experimental results, the author warns experimenters and theorists of the possible complete lack of connection with outside-the-lab world that this procedure could cause.

Bardsley (2005), on the other hand, is more drastic, asserting that in some instances the situations explored in the lab are completely different from those supposed to be investigated, since some 'relational' phenomena occur only in the real world. According to the 'artificiality-of-alteration critique', only some individual decision problems and games would be possible to study in the laboratory, and he also asserts how the experimental reports should conform to this assertion.

Hogarth (2005) lingers over the representativeness of experimental designs. Given that these are framed in abstract environments, the results so obtained would be generalized only to abstract environments, and not to the real outside-of-the-lab world. He also identifies the way of making the experiment representative: considering any experiment as a sample, two the suggested dimensions to take into account: participants (but see Section 4, and Cooper et al., 1999), and, probably as the most demanding requirement, situations.

### *7.2. Are experiments the suitable setting where to test theories?*

Being one of the primary purposes in running experiments the theory testing ("speaking to theorists" category in Roth's taxonomy (1995)), it is clear that also this question becomes of preliminary importance: to assess to what extent the experimental test of theories are appropriate. The problem is clearly stated, for example, in Starmer (1999a).

Sugden (2005a) recognizes the "triangular relationship" between experiment, theory and world, showing that experimenters have often used in their own defence against the artificiality critique the fact that their experiments resemble as much as possible the theory to be tested (e.g., Plott, 1982, 1991), so that the alleged "unrealism" is to be sought ultimately in the theory,  returning the critique to sender. In turn, this response has raised further points, issues developed by Guala (2005a), Cubitt (2005) and Hausman (2005), respectively. In particular, Guala (2005a) lingers over the problem of completeness, that is, the importance of an exact specification of the domain for economic theories, while Cubitt (2005) points out the distinction between the fundamental domain of an economic theory, and the domains of its applications and of its empirical testing, and the legitimacy to test economic theories in the lab. Finally, Hausman (2005) is concerned with the specific problem whether game theory could be actually tested in the laboratory, concluding that the experimenter test only "some particular empirical interpretation of game theory" (p. 212).

A closely connected problem is the fact that theoretical assessment may heavily affect the decisions regarding the empirical researches to be undertaken and the way of interpreting the experimental data (Starmer, 2005. On this point, the reader can also refer to Roth, 1988). Further interesting remarks can be found in Sugden (2005b), about the different roles of experiments in behavioural economics, and the interconnection about some specific behavioural phenomena found in experimental setting and theory; in Mäki (2005), that deals with the close similarities between models and experiments, and even their interchangeable role in some cases, being the difference only in the means of control; and in Morgan (2005), who argues that both theory and experiments have the same mediator role, but experiments may have a greater epistemic power, both since experiments share the same ontology with the real economic world, and since they give the opportunity to discover new phenomena, provided that participants have a certain degree of freedom during the experiment.

### *7.3. The Duhem-Quine thesis*

In passing, I think is worth citing this common problem to any experimental science (Duhem, 1953; Quine, 1953), no less, therefore, to experimental economics, and so often cited in the literature (Smith, 1994). According to this, a theoretical claim can be empirically tested not in isolation, but only if conjoined with some 'ancillary' assumptions: "economic theories do not come fully specified with all the conditions necessary for conducting a test. Theories must be interpreted before they can tested [..] one is testing not simply the validity of some model, but the model combined with various assumptions" (Starmer, 1999a). Often, Smith's precepts are considered as auxiliary  hypotheses in experimental economics. According to Duhem-Quine thesis, results of every empirical test are always susceptible of more than one interpretation: for example, in the case a theory fails, there is the doubt that this may be due to the fact that auxiliary hypothesis have not held, and not the proper theory being tested.

### *7.4. The parallelism precept as principle of induction*

As we have seen, some of the most prominent economists have tried to settle the artificiality critique elegantly, remarking the connection between the theory and the laboratory, rather than between the laboratory and the real world:

> "the more that accepted theory can be invoked, the less the experimental process needs to "mirror" the natural analog" (Plott, 1982, p. 1521).

> "The experiments should be judged by the lessons it teaches about theory and not by its similarity with what nature might happen to have created" (Plott, 1991, p. 906).

> "what is most important about any particular experiment is that it be relevant to its purpose. If its purpose is to test a theory, then it is legitimate to ask whether the elements of alleged 'unrealism' in the experiment are parameters in the theory. If they are not parameters in the theory, then the criticism of 'unrealism' applies equally to the theory and the experiment" (Smith, 1982, p.937).

In order to reply to artificiality critique, the fifth precept advocated by Smith (1982), *parallelism*, may come to experiments' rescue. Smith asserts that:

> "since economic theory has been inspired by field environments, we would like to know [..] if such (experimental) results are transferable to field environments.
> A sufficient condition or this transferability of results can be summarized as a final precept [..].
> *Parallelism*: Proposition about the behaviour of individuals and the performance that have been tested in laboratory microeconomies apply also to nonlaboratory microeconomies where similar *ceteris paribus* conditions hold" (emphasis in the original, p. 936).

However, granting this position, particular attention should be paid to determine exactly what are the *ceteris paribus* conditions for each single situation, and this determination is far from being simple in most of cases.
Roughly the same position is shared by Wilde (1980), Loomes (1999a) and Starmer (1999b), whereas in Starmer (1999a), he also argues that most of artificiality criticism is based on implementation of specific procedures, but no arguments would be against the experimentation on humans in more general philosophical terms.
It should be noted that this precept is especially helpful when experiments are aimed at being test-bed for economic theories, as cited authors themselves admit. Meanwhile, in all the cases in which experiments are aimed at different goals, the parallelism precept may not be sufficient or even not applicable, and so more attention to the concern about the external validity should be paid.
Some recent works may prove to be enlightening. For example, Dohmen et al. (2005) test behavioural relevance of a survey by means of a field experiment. In like manner, we should conduct some kinds of representative survey to validate experimental results. Benz and Meier (2006), on the other hand, test if lab experiment results are correlated with field experiment results.
As an alternative solution, there should be the requirement of a systemic investigation of the 'context effects', that is, the study of the effects of different descriptions of the same experimental situation to participants, one description made in 'abstract' terms, and the other made in 'realistic' terms. In an experimental market setting, for example, in one treatment subjects could be told that they are buyers or sellers, that they are required to trade mugs, pencil, or whatever else, and so on ('realistic' treatment), whereas in the other treatment they could be told that during the experiment they act or in the role A or in the role B, that they have to trade

simply a good, and so on ('abstract' treatment)[23] (substantially the same idea is shared by Harrison and List, 2004).
As regards the last two issues, a work by Bernasconi et al. (2006) is particularly relevant in these two main respects. On the one hand, concerning comparison of real with lab data, it should be noted that, in a forecast task, their experimental subjects are exposed to real stimuli, rather than artificially constructed time series. On the other hand, concerning the test of context effects, they implement a control treatment where subjects are not informed about the actual origin of the times series given as stimulus.
Finally, it is clear that the "artificiality critique" remains one of the most important open issue in experimental economics, and probably in a very recent future an ever increasing amount of work will be lavished on it, even though a laboratory experiment should be judged by its impact on understanding, rather than by its fidelity either to reality or to formal model. At the same time, however, it should be clear that this criticism equally applies to some theories. Even though some experiments, to some extent, can appear unrealistic (because they provide no cue at all to participants, contrary to what they usually experience in the real world, so that they could feel confused) nevertheless it is surely no more realistic, for example, a theoretical model entirely grounded on the representative agent approach (e.g., Kirman, 1992). Since "the abstractions of the laboratories are orders of magnitude smaller than those of economic theory" (Smith, 1982, p.936), the "artificiality critique" should lead to question also some methodological issues on the theoretical side of the matter.

## 8. Conclusions

Being the experimental economics a "subdiscipline [..] still relatively young" (Loomes, 1999b), it is understandable that experimenters have delayed a little developing wide methodological discussions (Schram, 2005, p. 225). However, at the same time, it is undeniable to affirm that some incontrovertible firm points have been reached among the profession, first, maybe tacitly and implicitly, but afterwards, an ever increasing shared "common practice" whose I have tried to account for in this work.
Probably, someone will brand the methodology I have tried to illustrate to be the "lower-case-m methodology" (McCloskey, 1985, p. 25, as cited by Hands, 2001), as opposed to the proper Methodology, the one dealing with the philosophical issue of scientific knowledge (at the moment, in the field of experimental economics, maybe one of the first attempts in this direction is Guala, 2005b). Nevertheless I have judged noteworthy to provide a survey of the methodological works appeared, among other things, in mainstream economics journals, along with my modest reasoned comments, for two main reasons.

[23] In addition, often in financially motivated experiments, in the course of the experiment itself the payoff is not computed in national currency, but rather in "tokens" and then cashed at a known exchange rate at the end of the experiment. Therefore, also the use of national currency instead of some more abstract kind of currency should be tested to be relevant.

First, the level of the methodological debate is far from being held cheap. Indeed, maybe because trying to counter against the numerous attacks, experimental economists have promptly begun to affirm the scientific dignity of their research programme with a high intellectual effort devoted to it.
Second, notwithstanding this praiseworthy effort, at the current state of the literature, a systematic survey of methodological contributions was still missing, and this has been the stimulus behind this work, with which I hope to have filled the gap, partly at least.
At this point, I need hardly summarize the main points discussed above. However, I find a couple of remarks noteworthy to express.
First, a clear pattern can be derived from the methodologies put in practice. As a desiderable property, we should try to design experiments having in mind as much as possible the experimental aim and the research question. With regard to this, the researcher's sensitivity, able to weight in a wise way all the ingredients, is of paramount importance. In order to obtain this goal, experimenter should always takes in consideration a series of tradeoffs are at stake: between realism and control, between control and external validity, between what is considered theoretically ideal and what is actually feasible.
Second, but on this experimenters have already gone rather a long way, we should preserve as much as possible reliability and replicability, and at this aim common practice more widely shared are essential. At this point, the habit to make available rough data and all the material used in the experiment is more than commendable[24].
Third, the fact that the situation is, on the whole, not so bad, should not let us rest on laurels. In fact, important open issues remain, especially regarding the external validity of experiments. At this aim, I find rather doubtless that a more frequent use of what I have defined *metaexperiments* will prove to be very fruitful in the future.
In conclusion, experimental economics could have started the inductive turn off in economics.

[24] We should mention a Web forum especially devoted to discussion in experimental methodology promoted by ESA members as a more than welcomed initiative.

## References


Abbink, K., Selten, R., and Sadriech A., 1999, Money does not Induce Risk Neutral Behaviour, but Binary Lotteries do even Worse, *Theory and Decision*, 46, 213-252.

Arthur, W. B.,2006, Out-of-equilibrium economics and agent-based modelling, in K. Judd and L. Tesfatsion, eds. *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*, North-Holland Press.

Bardsley, N., 2005, Experimental Economics and the Artificiality of Alteration, *Journal of Economic Methodology*, 12(2), 239-251.

Bardsley, N., and Moffat, P., 2005, The Experimetrics of Public Goods: Inferring Motivations from Contributions. Discussion Paper, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.

Beattie, J., and Loomes, G., 1997, The Impact of Incentives Upon Risky Choice Experiments, *Journal of Risk and Uncertainty*, 14, 149-162.

Becker, G.M., DeGroot M.H., and J. Marschak, 1964, Measuring Utility by a Single-Response Sequential Method, *Behavioral Science*, 9, 226-232.

Benz, M., and Meier, S., 2006, Do People Behave in Experiments as in the Field? Evidence from Donations. Working Paper, Institute for Empirical Research in Economics.

Bernasconi, M., Kirchkamp, O., and Paruolo, P., 2006, Do Fiscal Variables Affect Fiscal Expectations? Experiments with Real Worls and Lab Data. CRIEFF Discussion Paper, Centre for Research into Industry, Enterprise, Finance and Firm.

Binmore, K., 1999, Why Experiment in Economics?, *Economic Journal*, 109, F16-F24.

Bohm, P., Lindén, J., and Sonnegård, J., 1997, Eliciting Reservation Prices: Becker-DeGroot-Marschack Mechanism vs. Market, *Economic Journal*, 107, 1079-1089.

Bonetti, S., 1998a, Experimental Economics and Deception, *Journal of Economic Psychology*, 19, 377-395.

Bonetti., S., 1998b, Reply to Hey and Starmer and McDaniel, *Journal of Economic Psychology*, 19, 411-414.

Brunner, E., Domhof, S., and Langer, F., 2002, *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*, Wiley, NY.

Burns, P., 1985, Experience and Decision Making: A Comparison of Students and Businessmen in a Simulated Progressive Auction, in V. Smith ed. *Research in Experimental Economics*, Vol. 3, Greenwich: JAI Press.

Camerer, C.F., 1995, Individual Decision Making, in A. Roth and J. Kagel, eds. *Handbook of Experimental Economics*, Princeton: Princeton University Press, 587-703.

Camerer, C.F., 1997, Rules for Experimenting in Psychology and Economics, and Why They Differ, in W. Albers, W. Güth, P. Hammertein, B. Moldovanu and E. von Damme, eds. *Understanding Strategic Interaction. Essays in Honor of Reinhard Selten*, Springer, 313-327.

Camerer, C.F., and Ho, T.H., 1998, Experienced Weighted Attraction Learning in Normal-Form Games, *Econometrica*, 67(4), 827-874.

Camerer, C.F., and Hogarth, R.M., 1999, The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework, *Journal of Risk and Uncertainty*, 19, 7-42.

Camerer, C.F., Loewenstein, G., and Weber, M., 1989, The Curse of Knowledge in Economic Settings: An Experimental Analysis, *Journal of Political Economy*, 97, 1232-1254.

Cameron, L., 1999, Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia, *Economic Inquiry*, 37, 47-59.

Carbone, E., 2005, Demographics and Behaviour, *Experimental Economics*, 8(3), 217-232.

Chamberlin, E.H., 1948, *The Theory of Monopolistic Competition*, Cambridge: Harvard University Press.

Conover, W.J., 1980, *Practical Nonparametric Statistics*, Wiley, NY.

Cooper, D.J., Kagel, J.H., Lo, W., Gu, Q.L., 1999, Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers, *American Economic Review*, 89(4), 781-804.

Cox, L.C., Smith, V.L., and Walker, J.K., 1992, Theory and Misbehavior of First Price Auctions: Comment, *American Economic Review*, 82(5), 1392-1412.

Cubitt, R., 2005, Experiments and the Domain of Economic Theory, *Journal of Economic Methodology*, 12(2), 197-210.

Cubitt, R., Starmer, C., and Sugden, R., 1998, On the Validity of the Random Lottery Incentive System, *Experimental Economics*, 1(2), 115-138.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G.G., 2005, Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. IZA Discussion Paper.

Davis, D.D., and Holt C.A., 1993, *Experimental Economics*, Princeton: Princeton University Press.

Duhem, P., 1953, Physical Theory and Experiment, in H. Feigl and M. Brodbeck eds. *Reading in the Philosophy of Science*, Appleton Century Crofts, 235-252.

Dyer, D., Kagel, J.H., and Levin, D., 1989, A Comparison of Naïve and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis, *Economic Journal*, 99, 108-115.

Erev, I., and Roth, A.E., 1998, Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria, *American Economic Review*, 88(4), 848-881.

Falk, A., and Fehr, E., 2003, Why Labour Market Experiments?, *Labour Economics*, 10, 399-406.

Fehr, E., and Gächter, S., 2004, Fairness and Retaliation: The Economics of Reciprocity, in Camerer, CF., Loewenstein, G., and Rabin M., eds. *Advances in Behavioral Economics*. Princeton: Princeton University Press.

Fischbacher, U., 1999, z-Tree – Zurich Toolbox for Readymade Economic Experiment – Experimenter's Manual. Working Paper, Institute for Empirical Research in Economics, University of Zurich.

Forsythe, R., Horowitz, J.L., Savin, N.E., and Sefton, M., 1994, Fairness in Simple Bargaining Experiments, *Games and Economic Behavior*, 6, 347-369.

Friedman, D., 1988, Experimental Methods: Points of Consensus and Points of Contention. Working Paper, University of California at Santa Cruz.

Friedman, D., 1992, Theory and Misbehavior of First Price Auctions: Comment, *American Economic Review*, 82(5), 1374-1378.

Friedman, D., and Cassar, A., 2004, *Economics Lab: An Introduction to Experimental Economics*, Routledge, London & NY.

Friedman, D., and Sunder S., 1994, *Experimental Methods*, Cambridge: Cambridge University Press.

Friedman, M., 1953, *Essays in Positive Economics*. Chicago: University of Chicago Press.

Gibbons, R., 1992, *Game Theory for Applied Economists*, Princeton University Press.

Gode, D.K., and Sunder, S., 1993, Allocative Efficiency of Markets with Zero Intelligence Traders, *Journal of Political Economy*, 101, 119-137.

Greiner, B., 2004, The Online Recruitment System ORSEE - A Guide for the Organization of Experiments in Economics. Discussion Papers on Strategic Interaction, Max Planck Institute of Economics.

Grether D., and Plott C.R., 1979, Economic Theory of Choice and the Preference Reversal Phenomenon, *American Economic Review*, 69(4), 623-638.

Guala, F., 2005a, Economics in the Lab: Completeness vs. Testability, *Journal of Economic Methodology*, 12(2), 185-196.

Guala, F., 2005b, *The Methodology of Experimental Economics*, Cambridge University Press.

Guala, F., A Short History of Experimental Economics. Unpublished paper, Department of Sociology and Philosophy, University of Exeter.

Guillén, P., and Veszteg, R.F., 2006, Subject Pool Bias in Economic Experiments. Working Paper, Departamento de Teorìa e Historia Econòmica, Universidad de Granada.

Güth, W., Schmittberger, R., and Schwarz,, B., 1982, An Experimental Analysis of Ultimatum Bargaining, *Journal of Economic Behavior and Organization*, 3, 367-388.

Hands, D. Wade, 2001, *Reflection without Rules. Economic Methodology and Contemporary Science Theory*, Cambridge: Cambridge University Press.

Harrison, G.W., 1989, Theory and Misbehavior of First Price Auctions, *American Economic Review*, 79(4), 749-762.

Harrison, G.W., 1992, Theory and Misbehavior of First Price Auctions: Reply, *American Economic Review*, 82(5), 1426-1443.

Harrison, G.W., 1994, Expected Utility Theory and the Experimentalists, *Empirical Economics*, 19, 223-253.

Harrison, G.W., and Johnson, L., 2005, Identifying Altruism in the Laboratory. UCF Economics Working Paper No. 04-04.

Harrison, G. W., Lau, M.I., and Rutström, E.E., 2005, Dynamic Consistency in Denmark: A Longitudinal Field Experiment. UCF Economics Working Paper No. 05-02.

Harrison, G.W., and List, J.A., 2004, Field Experiments, *Journal of Economic Literature*, 42, 1009-1055.

Hausman, D.M., 2005, 'Testing' Game Theory, *Journal of Economic Methodology*, 12(2), 211-224.

Hertwig, R., and Ortmann, A., 2001, Experimental Practises in Economics: A Methodological Challenge for Psychologists?, *Behavioral and Brain Science*, 24, 383-403.

Hey, J.D., 1991, *Experiments in Economics*, Oxford: Blackwell.

Hey, J.D., 1998, Experimental Economics and Deception: A Comment, *Journal of Economic Psychology*, 19, 397-401.

Hey, J.D., 2001, Does Repetition Improve Consistency?, *Experimental Economics*, 4(1), 5-54.

Hey, J.D., and Lee J., 2005, Do Subjects Separate (or Are They Sophisticated)?, *Experimental Economics*, 8(3) 233-265.

Hoeffler, S., and Ariely, D., 1999, Constructing Stable Preferences: A Look into Dimensions of Experience and Their Impact on Preference, *Journal of Consumer Psychology*, 8(2), 113-139.

Hoffman, E., McCabe, K., and Smith, V.L., 1996, Social Distance and Other-Regarding Behavior in Dictator Games, *American Economic Review*, 86(3), 653-660.

Hogarth, R.M., 2005, The Challenge of Representative Design in Psychology and Economics, *Journal of Economic Methodology*, 12(2), 253-264.

Holland, J.H., 1997, *Emergence, from chaos to order*, Addison-Wesley.

Holland, J.H., 1995, *Adaptation in Natural and Artificial Systems*, Univ. of Michigan Press.

Holt, C.A., 1986, Preference Reversal and the Independence Axiom, *American Economic Review*, 76(3), 508-515.

Holt, C.A., and Laury, S.K., 2002, Risk Aversion and Incentive Effects, *American Economic Review*, 92(5), 1644-1655.

Isaac, M.R., and Walker, J.M., 1991, Costly Communication: An Experiment in a Nested Public Goods Problem, in T.R. Palfrey, ed. *Laboratory Research in Political Economy*, Ann Arbor: University of Michigan Press.

Jones, M.K., and Seaman, P.T., 2003, Sampling Bias in Economics Experiment: An Analysis of the Effect of Attrition among Subjects who Sign Up for Experiments. Discussion Paper, Economic Studies, University of Dundee.

Jones, S.R.G., 1992, Was There a Hawthorne Effect?, *American Journal of Sociology*, 98(3), 451-468.

Kagel, J.H., Battalio, R.C., and Walker, J.M., 1979, Volunteer Artifacts in Experiments in Economics: Specification of the Problem and Some Initial Data from a Small-Scale Field Experiment, in V. Smith ed. *Research in Experimental Economics*, Vol. 1, Greenwich: JAI Press.

Kagel, J.H., and Roth A.E., 1992, Theory and Misbehavior of First Price Auctions: Comment, *American Economic Review*, 82(5), 1379-1391.

Kagel, J.H., and Roth A.E., 1995, *Handbook of Experimental Economics*, Princeton: Princeton University Press.

Kahneman, D., and Tversky, A., 1979, Prospect Theory: An Analysis of Decision Under Risk, *Econometrica*, 47(2), 263-291.

Kim, O., Walker, J.M., 1984, The Free-Rider Problem: Experimental Evidence, *Public Choice*, 43, 3-24.

Kirman, A.P., 1992, Whom or What Does the Representative Individual Represent?, *Journal of Economic Perspectives*, 6(2), 117-136.

Knetsch, J.L., 1989, The Endowment Effect and Evidence of Nonreversible Indifference Curves, *American Economic Review*, 79(5), 1277-1284.

Landsberger, H.A., 1958, *Hawthorne Revisited*, Ithaca, NY: Cornell University.

Ledyard, J., 1995, Public Goods: A Survey of Experimental Research, in A. Roth and J. Kagel, eds. *Handbook of Experimental Economics*, Princeton: Princeton University Press, 111-194.

Levitt, S.D., and List, J.A., 2007, What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?, *Journal of Economic Perspectives*, 21(2), 153-174.

Lichtenstein, S., and Slovic, P., 1971, Reversal of Preferences between Bids and Choices in Gambling Decisions, *Journal of Experimental Psychology*, 89, 46-55.

Lichtenstein, S., and Slovic, P., 1973, Response-Induced Reversals of Preferences in Gambling: An Extended Replication in Las Vegas, *Journal of Experimental Psychology*, 101, 16-20.

Lipsey, R.G., 1979, *An Introduction to Positive Economics*, 5th ed, London: Weidenfeld and Nicholson.

Loewenstein, G., 1999, Experimental Economics from the Vantage-Point of Behavioural Economics, *Economic Journal*, 109, F25-F34.

Loomes, G., 1999a, Some Lessons from Past Experiments and Some Challenges for the Future, *Economic Journal*, 109, F35-F45.

Loomes, G., 1999b, Experimental Economics: Introduction, *Economic Journal*, 109, F1-F4.

Mäki, U., 2005, Models are Experiments, Experiments are Model, *Journal of Economic Methodology*, 12(2), 303-316.

McCloskey, D.N., 1985, *The Rhetoric of Economics*, Madison: University of Wisconsin Press.

McDaniel, T., and Starmer, C., 1998, Experimental Economics and Deception, *Journal of Economic Psychology*, 19, 403-409.

Merlo, A., and Schotter, A., 1992, Theory and Misbehavior of First Price Auctions: Comment, *American Economic Review*, 82(5), 1413-1425.

Mill, J.S., 1874, On the Definition of Political Economy and the Method of Investigation Proper to It, in *Essays on Some Unsettled Questions of Political Economy*, 2nd ed. London: Longmans, Green, Reader, and Dyer [1st edition 1844].

Morgan, M.S., 2005, Experiment versus Models: New Phenomena, Inference and Surprise, *Journal of Economic Methodology*, 12(2), 317-329.

Ortmann, A., and Hertwig R., 2002, The Cost of Deception: Evidence from Psychology, *Experimental Economics*, 5(2), 111-131.

Ortmann, A., and Tichy, L., 1999, Understanding Gender Effects in the Laboratory: Evidence from Prisoner's Dilemma Games, *Journal of Economic Behavior and Organization*, 39, 327-339.

Palfrey, T., and Porter R., 1991, Guidelines for Submission of Manuscripts on Experimental Economics, *Econometrica*, 59(4), 1197-1198.

Plott, C.R., 1982, Industrial Organization Theory and Experimental Economics, *Journal of Economic Literature*, 20, 1485-1527.

Plott, C.R., 1991, Will Economics Become an Experimental Science?, *Southern Economic Journal*, 57, 901-919.

Plott, C.R., 1996, Rational Individual Behaviour in Markets and Social Choice Processes: the Discovered Preference Hypothesis, in K. Arrow, M. Perlman, and C. Schmidt eds. *Rational Foundations of Economic Behavior*, London: MacMillan, 225-250.

Quine, W.V., 1953, *From a Logical Point of View*, Harvard University Press.

Read, D., 2005, Monetary Incentives, What Are They Good For?, *Journal of Economic Methodology*, 12(2), 265-276.

Read, D., Loewenstein, G., and Rabin, M., 1999, Choice Bracketing, *Journal of Risk and Uncertainty*, 19 (1-3), 171-197.

Roth, A.E., 1988, Laboratory Experimentation in Economics: A Methodological Overview, *Economic Journal,* 98, 974-1031.

Roth, A.E., 1995, Introduction to Experimental Economics, in A. Roth and J. Kagel, eds. *Handbook of Experimental Economics*, Princeton: Princeton University Press, 3-109.

Roth, A.E., and Erev, I., 1995, Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term, *Games and Economic Behavior*, Special Issue: Nobel Symposium, 8(1), 164-212.

Roth, A.E., and Malouf, M.W.K., 1979, Game Theoretic Models and the Role of Information in Bargaining, *Psychological Review*, 574-594.

Rousseas, S.W., and Hart A.G., 1951, Experimental Verification of a Composite Indifference Map, *Journal of Political Economy*, 59, 288-318.

Rubinstein, A., 2001, A Theorist's View of Experiments, *European Economic Review*, 45, 615-628.

Samuelson, P.A., and Nordhaus W.D., 1985, *Principles of Economics*, 12th ed.

Schelling, T.C., 1957, Bargaining, Communication, and Limited War, *Journal of Conflict Resolution*, 1, 19-36.

Schram, A., 2005, Artificiality: The Tension between Internal and External Validity in Economic Experiments, *Journal of Economic Methodology*, 12(2), 225-237.

Siegel, S., and Fouraker, L.E., 1960, *Bargaining and Group Decision Making*, New York: McGraw-Hill.

Slovic, P., 1995, The Construction of Preference, *American Psychologist*, 50(5), 364-371.

Slovic, P., and Lichtenstein, S., 1968, Relative Importance of Probabilities and Payoffs in Risk Taking, *Journal of Experimental Psychology Monograph Supplement*, 3(2), 1-18.

Smith, A., 1776, Inquiry *into the Nature and Causes of the Wealth of the Nations*, Edinburgh.

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

Smith, V.L., 1976, Experimental Economics: Induced Value Theory, *American Economic Review*, 66(2), 274-279.

Smith, V.L., 1982, Microeconomic System as an Experimental Science, *American Economic Review*, 72(5), 923-955.

Smith, V.L., 1994, Economics in the Laboratory, *Journal of Economic Perspective*, 8(1), 113-131.

Smith, V.L., and Walker J.M., 1993, Monetary Rewards and Decision Costs in Experimental Economics, *Economic Inquiry*, 31, 245-261.

Starmer, C., 1999a, Experiments in Economics… Should We Trust the Dismal Scientists in White Coats?, *Journal of Economic Methodology*, 6(1), 1-30.

Starmer, C., 1999b, Experimental Economics: Hard Science or Wasteful Tinkering?, *Economic Journal*, 109, F5-F15.

Starmer, C., 2005, Normative Notions in Descriptive Dialogues, *Journal of Economic Methodology*, 12(2), 277-290.

Starmer, C., and Sugden, R., 1991, Does the Random Lottery Incentive System Elicit True Preferences? An Experimental Investigation, *American Economic Review*, 81(4), 971-978.

Sugden, R., 2000, Credible Words: The Status of Theoretical Models in Economics, *Journal of Economic Methodology*, 7(1), 1-31.

Sugden, R., 2005a, Experiment, Theory, World: A Symposium on the Role of Experiments in Economics, *Journal of Economic Methodology*, 12(2), 177-184.

Sugden, R., 2005b, Experiments as Exhibits and Experiments as tests, *Journal of Economic Methodology*, 12(2), 177-184.

Tesfatsion, L., 1997, How economists can get alife, in W. B. Arthur, S. Durlauf and D. Lane, eds. *The economy as an evolving complex system, II*, Addison-Wesley.

Tesfatsion, L., 2001, Introduction to the special issue on agent-based computational economics, *Journal of Economic Dynamics and Control*, 25, 281–293.

Tenorio, R., and Cason, T.N., 2002, To Spin or not to Spin? Natural and Laboratory Experiments from "The Price is Right", *Economic Journal*, 112, 476, 170-195.

Thaler, R., 1987, The Psychology of Choice and the Assumptions of Economics, in E. Roth ed. *Laboratory Experimentation in Economics. Six Points of Views*, Cambridge University Press.

Thurstone, L.L., 1931, The Indifference Function, *Journal of Social Psychology*, 2, 139-167.

Tversky, A., and Kahneman, D., 1981, The Framing of Decisions and the Psychology of Choice, *Science*, Vol. 211. no. 4481, 453 – 458.

Vickrey, W., 1961, Counterspeculation, Auctions, and Competitive Sealed Tenders, *Journal of Finance*, 16(1), 8-37.

Wallis, W.A., and M. Friedman, 1942, The Empirical Derivation of Indifference Functions, in O. Lange, F. McIntyre, and T.O. Yntema eds. *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz*, Chicago, University of Chicago Press, 175-189.

Wilcox, N., 1993, Lottery Choice, Incentives, Complexity and Decision Time, *Economic Journal*, 103, 1397-1417.

Wilde, L., 1980, On the Use of Laboratory Experiments in Economics, in J. Pitt ed. *The Philosophy of Economics*, Dordrecht: Reidel.