Experiments in Economics: should we trust the dismal scientists in white coats?

By

Chris Starmer*

Published in *Journal of Economic Methodology* 1999

Abstract

Is the rapid growth of experimental research in economics evidence of a new scientific spirit at work or merely fresh evidence of a misplaced desire to ape the methods of natural sciences? It is often argued that economic experiments are artificial in some sense which tends to render the results problematic or uninteresting. In the early part of this paper I argue that this *artificiality critique* does not provide a convincing philosophical objection to experimentation in economics. Later sections of the paper argue that methodological discourse in relation to experiments has become somewhat polarized: experimentalists have promoted a position which seeks to defuse objections to experiments; theorists have taken up positions which insulate theory from experimental challenge. I argue that these strategies are overly defensive and tend to stifle rather than promote the goals of economic enquiry.

Introduction

> "unfortunately, we can seldom test particular predictions in the social sciences by experiments explicitly designed to eliminate what are judged to be the most important disturbing influences." Milton Friedman (1953, p.10) "The Methodology of Positive Economics" in Essays in Positive Economics.
>
> "It is rarely, if ever, possible to conduct controlled experiments with the economy. Thus economics must be a non-laboratory science." Richard Lipsey (1979, p39), An Introduction to Positive Economics , Fifth Edition.
>
> "Economists ......cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe" Paul Samuelson and William Nordhaus (1985, p.8), Principles of Economics, Twelfth Edition.
>
> "One of the weaknesses in the claim that the social sciences are sciences at all is their inability to conduct controlled experiments. Physicists can create vacuums, chemists can establish sterile environments, even doctors can conduct blind trials. But economists, sociologists political scientists and those who study management find their subject matter will never stand still." John Kay, Financial Times, 3rd Jan, 1997.

It has been part of the folklore of economics that it is a non-experimental science. Yet, when Charles Plott[1], in his Presidential Address to the Sixtieth Annual Meeting of the Southern Economic Association in 1990, raised the question "Will Economics Become an Experimental Science" he confidently predicted that it would. Was he joking?

Economists have been employing experimental methods for at least 60 years[2]. For a good part of this period, something like the first half, the experimental method played a very peripheral role in empirical economics. Experiments were quite rare, often conducted in a rather informal manner (relative to contemporary standards), and most likely regarded as something of a curiosity by the majority of academic economists. During the last 25 years or so, the picture has changed markedly. The use of experimental methods to address economic questions has grown rapidly and there are clear signs, in at least some quarters, that experimentation has acquired a significant degree of recognition as a legitimate branch of empirical enquiry, relevant to economic discourse. There are numerous indicators of this: articles reporting experimental research are now frequently published in leading international academic journals such as Econometrica[3], the American Economic Review and the Economic Journal; a new specialist journal Experimental Economics began publication in 1998; the Journal of Economic Literature introduced a new and distinct category for the classification of experimental work in the late 1980s; and during the first half of the 1990's a number of textbooks devoted to experimental economics were published[4].

These days it is easy to find economists who speak very enthusiastically about the contribution that experiments can make to economic enquiry. For instance:

> "..experimentation can be clearly seen to have truly begun its now steady and sustained transformation from an occasional curiosity into a regular means for investigating many kinds of economic phenomena" Alvin Roth (1988, p.974), Laboratory Experimentation in Economics.

> " Economics is one of the few sciences that is fortunate to have both the field and the laboratory with which to work........ the laboratory methodology, which has historically been absent, will grow and become an important partner in a joint effort to isolate the principles which govern economic behaviour." Charles Plott (1991, p.918), "Will Economics Become an Experimental Science?"

> "..... as currently practised, economics is ideally suited for experimental investigation" John Hey (1991, p.15), Experiments in Economics.

This enthusiasm for experimentation contrasts sharply with the sceptical tones of the opening quotations. Compare for instance Hey ("economics is ideally suited for experimental investigation") with Lipsey ("economics must be a non-laboratory science"). One might be forgiven for wondering whether these writers are talking about the same discipline! What can account for such a sharp disparity?

One distinction between these writers which may be relevant is that each of the 'enthusiasts' is currently a practising experimentalist while this is true of none of the 'sceptics'. Since none of the sceptics cited above were primarily concerned with discussing the role of experimentation in economics, their passing references to the topic might be interpreted as merely reflecting certain conventional but mistaken and by now outdated perceptions. Given the volume of experimental work produced currently, the view that experiments are rarely possible in economics seems hard to sustain. Perhaps the sceptics had simply failed to realise the potential for experimentation in economics more obvious to those actively engaged in the field?

This interpretation, however, may not do justice to the views of the sceptics. Even though the sceptics comments on experimentation were asides to their main arguments, in some cases, they are taken from works directly concerned with methodological issues surrounding the relation between theory and evidence in economics. It would seem somewhat rash, therefore, to simply dismiss the views of the sceptics as naive without due consideration. If taken simply as off-hand suggestions that one cannot (often) do experiments, the comments do seem out of touch with current events; these days, many hundreds of researchers in economics employ experimental methods. But there is another, more sophisticated, reading of the sceptics' comments. Perhaps the intention is to convey the more subtle point that, for the most part, experimental procedures may not provide very <u>meaningful</u> data relevant to the discussion of economic issues.

The latter is certainly a more plausible reading of Friedman, for instance. At the time of writing his Essays in Positive Economics, Friedman was well aware that economists had been using experimental methods. In fact, Friedman was co-author of a review - Wallis and Friedman (1942) - of one of the first experiments reported in the economics literature: that of Thurstone (1931). Thurstone's experiment attempted to experimentally elicit individual's indifference curves and in their review of Thurstone's paper, Wallis and Friedman wrote:

> "It is questionable whether a subject in so artificial an experimental situation could know what choices he would make in an economic situation; not knowing it is almost inevitable that he would, in entire good faith, systematize his answers in such a way as to produce plausible but spurious results".Wallis and Friedman (op. cit., p.179)

Arguments to the effect that experimental results may be 'spurious' because of the 'artificial' context in which they are generated will, I suspect, be familiar to almost anyone who has presented experimental results to an audience of general economists. Indeed, in a review of "Experimental Methods in Economics" Graham Loomes (1991) notes that one of the questions most frequently posed to experimentalists is:

> "Can you <u>really</u> take the observed behaviour of groups of volunteers spending short periods of time in carefully controlled environments and draw any meaningful conclusions about the way things work in 'the outside world'." (Loomes, 1991, ch.29 emphasis in the original).

It is a good question and it exemplifies a scepticism towards experimental research in economics which persists, in spite of the now widespread use of the method and its new prominence in mainstream dialogues. The purpose of this paper is to examine what basis there might be for such scepticism. In particular I shall be interested in what arguments might support the view that experimental settings are artificial in some non-trivial sense which renders their results of limited relevance to ‘real’ economic phenomenon.

Many of the points I will raise have been discussed elsewhere, though typically as asides to experimental reports or comments on such papers. Part of the purpose of this paper is to bring together possible doubts about, and defences of, experimental methods which one can find scattered through the experimental literature. In many cases, the references one finds to these issues are little more than off-hand comments. I will make an attempt to articulate a variety of potential doubts about the viability of experimental methods and I shall consider what defence an experimentalist might make.

2. The things experimenters do

Most experimental research in economics has focussed on microeconomic issues and two broad strands of research are discernable (i) experiments investigating *individual choice* and (ii) experiments involving *interactive choice* of one kind or another. Individual choice experiments typically involve no experimentally-relevant interaction between subjects (or at least the intention behind these experiments is that there should be no such interaction). The purpose of such experiments is to investigate properties of individuals such as their preferences (or choices) or properties of their decision making processes. A large proportion of the literature on individual choice behaviour has focussed on decision making under risk (see Camerer 1995 for a review). Some experiments in this vein have been directed towards testing specific theories (e.g. do choices under risk conform to expected utility theory). Other experiments have investigated possible biases/anomalies in decision making such as peculiarities in the way that people respond to probability information. Biases observed in the latter type of experiment will, of course, often imply failures of particular theories. There is, therefore, no clear cut line between the bias-hunting and the theory-testing experiment; it is more a question of emphasis.

As an illustration of what might be involved in a very simple individual choice experiment consider this choice problem discussed by Kahneman and Tversky (1979)[5]. An individual is asked to make two (hypothetical) pairwise choices between four alternative prospects labelled A, B, C and D. The choices are presented here in a state/payoff matrix with three states of the world, s1, s2 and s3 whose associated probabilities are 0.33, 0.01 and 0.66. In choice 1, an individual who chooses option A 'gets' £2400 for sure[6] whereas if they choose B, they enter a gamble which gives £2400 with probability 0.66 and £2500 with probability 0.33, otherwise nothing. Options C and D should be interpreted in a similar way.

		s1	s2	s3
	State:	s1	s2	s3
	Probability:	0.33	0.01	0.66
<u>Choice 1</u>				
Choose either	A:	2400	2400	2400
A or B	B:	2500	0	2400
<u>Choice 2</u>				
Choose either	C:	2400	2400	0
C or D	D:	2500	0	0

Kahneman and Tversky presented these problems to undergraduate students and found that the modal responses of subjects was to choose A then D. This pattern of response violates the independence axiom of expected utility theory (notice that choices 1 and 2 are identical apart from the value of the common consequence assigned to state s3) and it is an example of the well-known Allais paradox. Experimental observations like these did much to stimulate the development of alternatives to expected utility theory and a vast number of experiments have investigated choice behaviour in the context of simple gambles like these.

In the case of interactive experiments, (controlled) interaction between subjects usually forms an integral part of the experimental design. A large number of experiments of this latter type have been employed to investigate the properties of 'market' institutions such as different auction rules or pricing mechanisms - see Kagel (1995) for a review - or alternative market structures (see Holt, 1995). Among the questions of interest in these 'market' experiments is whether the market reaches the theoretical equilibrium, and if so, how quickly. Many of the interactive experiments have close connection to issues in game theory such as experiments related to the provision of public goods (see Ledyard, 1995) or bargaining behaviour (see Roth, 1995).

As a illustration of the interacative experiment, consider the problem of public goods. Numerous experiments have investigated individual contributions to public goods using some variant of the following simple design. A total of n experimental subjects are each endowed with a fixed number of tokens. Subjects in such an experiment might then be told that the following rules apply. Each token can be either traded for cash or contributed to a common pool. The total contributions to the common pool will be multiplied by a factor λ and the resulting sum divided equally between the subjects. Each subject must decide how many tokens to contribute to the pool and how many to trade for cash. These rules are intended to create an experimental analogue of a public goods problem and experiments along these lines have investigated a number of questions including the extent to which contributions to the public good (i.e., the common pool) depend upon a variety of treatment variables such as: the multiplication factor; the number of subjects; whether or not the subjects are allowed to communicate; whether or not-binding agreements are allowed; the composition of the subject pool (e.g. are they men or women, are they familiar with economic theory).

These 'example' experiments are intended only to provide a flavour of the sorts of things economists might do in an experimental context. They are far from complete accounts of any experimental design, nor do they reflect state-of-the art methodology. For example, unlike the individual choice 'experiment' described above, most genuine experiments conducted these days would use real, usually financial, incentives to motivate subjects. Presumably, this convention reflects a widely held view that real incentives are necessary to produce results which have meaningful interpretations. Much of the subsequent discussion will be concerned with just what requirements a valid experimental design would have to satisfy for the results to be meaningful.

Before this, however, I wish to consider whether there might be a general case to be made against experimenting in economics per se.

3. Experiments in principle

Two of my representative sceptics, Lipsey and Friedman, are well known for promoting empiricist methodologies in which the confrontation between theory and evidence plays a central role. Both authors suggest numerous analogies between the methods of the natural sciences and those appropriate to economics. Since, experimentation has contributed so much to the development of knowledge in the natural sciences, why can't experiments play a similar in economics?

It is perhaps reasonable to imagine that the problems confronting experimenters in natural and social sciences might be quite different for one basic reason: Natural scientists experiment on the physical world whereas social scientists are concerned with the behaviour of human beings. Might this difference form the basis for doubts about the applicability of experimentation in social science in general and economics in particular and, if so, why? I will argue that although there may indeed be important distinctions between the physical and the social world, such differences ought not to form the basis of any deep philosophical objection to experimenting in economics so long as one holds to the following view: *that theories are intended to predict phenomena and can be tested by comparing their predictions with the actual behaviour of the relevant economic phenomena*. I will argue that there can be no basis for a general objection to experimentation on human subjects in principle given agreement to this *predictivist* stance.

Let us begin, however, by entertaining what might constitute some obvious ground for doubt.

Perhaps one important distinction between humans and inanimate physical objects arises from the fact that, unlike objects in the physical world, humans make conscious choices concerning their behaviour. A bimetallic strip *will* bend when heated. This we may say, is a physical property of bimetallic strips and such objects have no choice in the matter: when they are heated they will behave as all bimetallic strips do. A human being, on the other hand, when placed in a choice situation where more than one course of action is available to them *decides* what to do; they engage in some (more-or-less) self-conscious process of decision making. In this section, I shall discuss two sorts of doubt about the possibility of experimentation which might arise from the recognition that human subjects are conscious choosing agents.

One possible implication of the fact that human action is a product of self conscious choice might be that the behaviour of human subjects is in some sense more *variable*, or less predictable, than the behaviour of physical objects. A second potential doubt could arise from the thought that the behaviour of human subjects may be sensitive to context in such a way that the experimental setting may often induce different behaviour from the natural setting to which it is intended to correspond. I will argue that neither consideration mitigates against experiments in principle though considerations of the second kind might place limits on the range of practically viable experiments.

The first thought can be quickly dispensed with. Suppose that human behaviour were, for whatever reason, more variable than the behaviour of objects in the physical world. In the extreme, one may even question whether human behaviour is governed by any general laws or tendencies and consequently whether human behaviour is predictable at all. Questions about whether human behaviour is ultimately determinate raise deep philosophical questions relating to free will[7] and so on. Fortunately, however, a simple defence of experimental method can be mounted against the "unpredictability argument" without the need to enter such deep waters.

Most economists believe that at least one purpose of building economic theories is to *predict* the actual behaviour of economic entities (e.g. firms, consumers, prices, output and so on). And while there may, of course, be other purposes for theories to serve, some clearly see prediction as the primary purpose of theorising. This is certainly the view of at least two of our opening sceptics, for instance, R. Lipsey & Crystal (1995, p33) argue that "A theory ceases to be useful when it cannot predict better than an alternative theory", similarly Friedman (1953, p23) states that "Viewed as a body of substantive hypotheses, a theory is to be judged by its predictive power for the class of phenomena which it is intended to 'explain'". For those who do take the view that prediction is at least part of the purpose of economic theorising, it will be hard to make a case against experimentation on the grounds that human behaviour is unpredictable, without simultaneously undermining the case for economic theory in its predictive role. Economic phenomena are ultimately reflections of human behaviour and hence, the possibility of economic prediction rests on the presumption (right or wrong) that there are at least some regularities in those aspects of human affairs with which economists are concerned.

More subtly one might argue that while there are indeed patterns in actual behaviour, the underlying explanations for these patterns are 'complex'. Suppose it were the case that a great many different factors influence economic phenomena with complex interactions between them; that might make experimenting difficult and costly. But if the world is complex in this way, economic theories do not typically refer to such complexity. Economists have traditionally built simple models which rely on a small number of variables to predict (or explain) patterns in behaviour. While such theories may not be complete 'explanations' of observed behaviour they could still be useful models so long as they succeed in capturing the *primary* determinants of the phenomena under investigation. But if the test of whether a theory has succeeded in this sense turns on whether its predictions conform with observed behaviour this would seem, on the face of it, to provide a license for experimentation. If the experimenter can create an environment in which the variables entering a theory can be observed, manipulated and controlled, the experimenter may test, directly, whether these variables are related as the theory predicts.

One problem which lurks here is deciding when an experimental setting is one to which the predictions of a particular theory ought to apply. Consider for example a simple public goods experiment. Suppose that one observes a significant degree of contribution in a case where a given theory apparently makes a prediction of complete free-riding. Should one reject the theory? Perhaps not, since it might be possible to defend the theory by objecting that the test did not satisfy all of the conditions necessary for the theoretical implications to hold. For example, perhaps the experiment did not secure appropriate conditions of anonymity, or perhaps the payoffs were not high enough for the subjects to take the problem 'seriously'.

Objections like this raise questions about whether the experimentally generated environment is one to which the predictions of the theory apply. If the theory is only intended to apply, say, in situations where the pay-offs are large enough to ensure subjects pay sufficient attention, experimental evidence from settings where payoffs are 'too small' might be legitimately discounted.

While potentially valid criticisms of a given experimental design, objections like this, which point to a specific limitation of the experimental setting, do not seem to tell against experimentation per se since the experimenter can mount a ready response to each such objection. For example if the hypothesis is that "the free-rider theory failed because the incentives were too small", then run a new experiment with bigger incentives. If it is suspected that communication between subjects enabled them to 'beat' the free-rider problem, design a new experiment which makes communication more difficult. So long as the theory defender identifies some specific aspect of the design which renders it unsatisfactory as a test of the target hypothesis, it seems reasonable to think that a new experiment could be run which could 'correct' the limitation of the earlier test. Hence, criticisms which point to specific reasons as to why an experiment is not a satisfactory test of a hypothesis do not tend to undermine experimenting; they suggest new problems which can be investigated experimentally; they enrich the experimental agenda.

On the other hand, it may be possible to raise objections to experiments which are rather less specific. Some experimentalists might be tempted to respond to such criticisms by arguing that non-specific objections are, by definition, non-credible. There may be some merit in this position. Simply to argue something like "your experiment is not an acceptable test of my theory *but I can't tell you why"* seems unconvincing. This said, there does seem to be some scope for non-specific doubts which it might be rash to dismiss without closer examination.

In experiments with human beings, the very fact that they are conscious of taking part in an experiment may condition their responses. This thought leads to a variety of possible objections to experiments which are both rather general - in the sense that they don't identify very specific limitations of particular experimental designs - but, nevertheless, of genuine concern. A couple of examples will, hopefully, suffice to illustrate the nature of the worry. Subjects might form a view about what the experimenter wants them to do, and act accordingly. Conversely, they may adopt strategies intended to outwit the experimenter. In either case, the validity of the experiment is undermined because one is not observing the intended phenomena, but some artifact of the experimental design. These are illustrations of a general problem of "experimenter effects".

The possibility of experimenter effects does not mean that it is impossible to conduct meaningful experiments with human subjects, rather it has implications for the appropriate design of experiments. If experimenter effects are a potential problem, one can, at least in principle, design experiments to circumvent their influence. For instance, in medical trials of drugs, subjects are randomly allocated to either an "experimental group" (who get the drug to be tested) or a "control group" who receive a placebo (a "dummy" pill which is expected to have no significant physiological influence in its own right). In this way, investigators aim to separate out real effects of the pill under trial from effects which arise simply because the subject is taking part in an experiment and swallowing a pill of some kind. Similar sorts of controls are often implicit in the design of experiments in economics.

For instance, experiments are often designed so that a hypothesis is tested by comparing the behaviour of two randomly selected groups under different experimental conditions. Here one is interested in determining whether behaviour varies in some systematic way between the two groups who have experienced conditions which differ in a way controlled by the experimenter. Any variation detected cannot simply be attributed to "having taken part in an experiment" as both groups are being experimented upon. Any systematic change in behaviour is rather attributable to the change in the experimental conditions under the control of the experimenter.[8]

It is still possible that the variation observed is nevertheless affected *in some way* by the experimental setting: perhaps the impact of the treatment variable is conditioned by aspects of the experimental design which are common to both groups? Consequently, it could be that the observed effect of the treatment variable is different from the effect that would be observed in a 'natural' setting. But if the problem is that some aspect of the experimental setting renders it different, in an important respect from the relevant natural setting, the solution is to adjust that experimental setting so that it then corresponds in the appropriate way.

This solution is possible, *in principle*, even in the case where it is simply the subjects' knowledge that she is taking part in an experiment which creates the disturbing influence. An appropriate control in this case would be found by conducting experiments which were designed so that the participants were completely unaware of having taken part in any experiment. For example, an economist could investigate certain hypotheses about the behaviour of stock market traders by obtaining the agreement of firms to present employees, who had no knowledge of the subterfuge, with fictitious trading opportunities on their computer screens[9].

So long as economists wish to claim that their theories do indeed have some empirical content, that is, they do embody claims of the sort that if certain conditions hold, specified behaviour should follow, then there seems to be, at least in principle, the possibility of creating the stated conditions experimentally to test the implications of the theory. One cannot reject experimentation on the grounds that experimentally constructed scenarios will always be 'unrealistic' in some important respect. In principle, they need not be so. Of course there may be practical, financial and indeed ethical considerations which would render certain sorts of realism impractical. While such considerations might well have implications for the practical scope of experimentation, they do not suggest any deep philosophical objection to experimenting on humans. It would, however, be the start of a worrying doubt about the usefulness of experiments if the practical scope of experimenting were to be drastically constrained by such considerations.

But while considerations of time, cost and in some cases ethics, will place limits on the extent to which experimentalists may replicate all features of some naturally occurring economic environment, such constraints may not be especially binding, or so some well-known experimentalists would assert. Plott (1991, p 906), argues that many early experimentalists mistakenly believed that "the only effective way to create an experiment would be to mirror in every detail, to simulate, so to speak, some ongoing natural process". But a good part of the rationale for experimenting, in Plott's view, is that it allows us to investigate the relationships hypothesised in economic theories while abstracting from other factors which may also be at work in the broader social setting, but which are not part of the theory being investigated. In so far as is possible, the experimenter wishes to close the door on these other factors. Hence:

> "Once models, as opposed to economies, became the focus of research the simplicity of an experiment and perhaps even the absence of features of more complicated economies became an asset. The experiment should be judged by the lessons it teaches about theory and not by its similarity with what nature might happened to have created" (Plott, Ibid).

Plott’s point seems to be that if we are concerned with theory testing, the issue is not how to replicate a real world decision setting *in every detail* but how to create an appropriate abstract setting which isolates the elements of interest; those which feature in the theory under consideration. Under these circumstances, the necessary controls will usually imply abstraction from day to day social settings.

When theory testing is the goal, the idea that experimental scenarios should not need to replicate all aspects of everyday economic settings seems persuasive, once we recognise that the theories being tested do not refer to all aspects of everyday environments. There may still be grounds, however, for questioning whether the scenarios typically created in the laboratory, provide appropriate abstractions, even for the purpose of theory testing. For instance, typically, subjects *do* know that they are taking part in experiments and the experimental setting will, more often than not, place the subject in a situation which is quite different from any day to day experience. Is this a legitimate practice, or is it likely to be prone to the artificiality critique?

By now, I would hope to have convinced the reader that so long as theories are thought of as (at least in part) devices for predicting behaviour no solid ground exists for a generally cast objection to experimentation. But there remains a question, or so it seems, of how we can tell when an experimental design provides an appropriate abstraction for the purpose of testing a particular theory. Within the experimental literature there has been some discussion of when an environment has been suitably controlled to allow a test of an economic hypothesis. This has been widely cited, though rarely has it been closely examined. It is to this that I now turn.

4. Experiments as 'real' microeconomic systems

How can we tell when an experimental design provides an appropriate test of a given hypothesis? Insofar as we are concerned with testing specific theoretical hypotheses at least part of the answer is supplied by the theory itself. In designing an experiment to test a particular theory, the economist will necessarily look to that theory to inform her about what is required to construct a satisfactory test of an hypothesis. The task is to create an experimental setting in which one can manipulate variables relevant to the theory, under controlled conditions, while observing the effects on other variables of interest.

But invariably, economic theories do not come fully specified with all the conditions necessary for conducting a test. Theories must be interpreted before they can be tested. Inevitably, then, one is testing not simply the validity of some model, but the model combined with various *auxiliary* assumptions or hypotheses which allow the theory to make predictions in this situation or that. Further, when an experiment has been run, there will then be questions about the extent to which the results generalise beyond the confines of the laboratory. Let me illustrate the point in relation to a particular class of experiment.

A large number of experiments have been concerned with investigating the behaviour of *microeconomic systems*. Vernon Smith (1982, 1989) argues that a microeconomic system can be thought of as consisting of two components: *environment* and *institution.* In Smith's terminology, the environment refers to: the set of agents participating in the system (including their individual characteristics such as their preferences); the goods existing in the economy; the production technology; and the initial resource endowments. The institution governs the interaction of agents: it can be thought of as a set of rules which define how agents may or may not interact and it specifies the form that any such interaction must take, including the language through which messages may be transmitted. Thus, the institution might govern, for instance, the structure of property rights, the rules of auctions, regulations of trade, and so on. The interplay of the environment and institution generate *behaviour* which refers to all the observable outcomes of the microeconomic system: the actions of agents, and the emergent allocations, for example. It is important to notice that this account of a microeconomic system has a two-way correspondence with theory and reality: it is the way in which economists conceptualise real economic systems. In experiments with microeconomic systems, the experimenter seeks to create and (in most cases) manipulate the institution and environment with a view to investigating the relationships between the behaviour of the system (the choices of the agents and the resultant outcomes) and characteristics of the environment and institution.

A good deal of experimentation can be interpreted in this way. For example, take the public goods experiments described in Section 2. The individuals in the experiment, their preferences, the initial endowment of tokens etc. constitute the environment. The rules of the experiment - how contributions are adjusted and redistributed to participants and so on, for instance - define the institution. What should we make of behaviour in these experimentally generated microeconomies? Is it no more than an artifact of whatever laboratory microeconomic system we have created, or is it behaviour of more general significance?

In a sequence of closely related papers Smith (1976, 1977, 1982) and Wilde (1980) have argued against the view that such laboratory simulations can dismissed as artifactual. Their claim is that, given certain assumptions, the laboratory microeconomies can be regarded as *real* (if small-scale) microeconomic systems; the behaviour of the agents can be thought of as real economic behaviour and, therefore, the observed behaviour is suitable as a test-bed for economic theories which purport to be general theories of economic behaviour.

The assumptions, or 'precepts' as Smith calls them, relate to the reward structure of the experiments. Following Smith (1982, pp.931-935) these can be stated as follows:

1 *“Nonsatition*: given a *costless* choice between two alternatives, identical (i.e., equivalent) except that the first yields more of the reward medium (for example, U.S. currency) than the second, the first will always be chosen (i.e., preferred) over the second, by an *autonomous* individual.” (emphasis in the original)

2 *“Saliency*: Individuals are guaranteed the right to claim a reward which is increasing (decreasing) in the goods (bads) outcomes, x^1, of an experiment”

Wilde (op. cit. p 141), the person responsible for introducing the idea of saliency explains the content of these two precepts as follows:

> “Saliency implies that the amount of the reward medium is linked to the decisions made by the subjects and nonsatiation implies that the amount of the reward medium earned is always important to the subjects.”

Thus nonsatiation amounts to the assumption that individuals will want to acquire the reward medium while saliency is necessary to ensure that the amount of the reward medium they will acquire by participation is determined by their actions in an appropriate way. Smith argues that 1 and 2 are "sufficient conditions for the existence of an experimental microeconomy, that is motivated individuals acting within the framework of an institution". In other words, if these two precepts are satisfied, we are entitled to interpret the behaviour observed in an experimental microeconomy as maximising behaviour which is appropriately tied to the institutional context of the experiment. Two further precepts are introduced to rule out other significant motivational factors as influences which may disturb the reward structure 'induced' by the experimenter.

3 *Dominance*: The reward structure dominates any subjective costs (or values) associated with participation in the activities of an experiment.

4 *Privacy*: Each subject in an experiment is given information only on his/her own payoff alternatives.

These are essentially footnotes to precept 1 covering the qualifications 'costless' and 'autonomous'. Precept 3 ensures that other costs (or benefits), such as transactions costs, do not override the motivation provided by the experimenter's intended reward medium. Precept 4, is intended to guard against motivations arising from interpersonal considerations such as altruism or envy. With the addition of 3 and 4, the agents in an experiment are maximising utility functions which are purely a function of the reward medium designed by the experimenter: the experimenter has then, <u>controlled</u> preferences in the sense that the experimenter has induced <u>known</u> preferences on the subjects. Wilde (1979, p142) argues that given precepts 1-4, "Then, in fact, the laboratory experiment constitutes a small-scale microeconomic environment in which real economic agents make real decisions". On the basis of this, Wilde makes the following claim:

> "if an experiment includes all parameters relevant to a particular theory, and if the theory fails to predict well in the simplified setting of the laboratory, then it cannot be expected to predict well in more complex environments. Again, the only requirements needed to reach this conclusion are that saliency, nonsatiation, dominance and privacy are satisfied by the reward structure. The experiment does not need to be 'realistic' and no presumption need be made about its connection to more complex ('real world') environments." (op. cit., p143).

Similarly, Smith (1982, p.935-6) writes:

> "Insofar as we are only interested in testing hypotheses derived from theories, we are done, that is, Precepts 1-4 are sufficient to provide rigorous controlled tests of our ability as economists to model elementary behaviour. Microeconomic theory abstracts from a rich variety of human activities which are postulated not to be of relevance to human economic behavior. The experimental laboratory, precisely because it uses reward-motivated individuals drawn from the population of economic agents in the socioeconomic system, consists of a far richer and more complex set of circumstances than is parameterized in our theories. Since the abstractions of the laboratory are orders of magnitude smaller than those of economic theory, there can be no question that the laboratory provides ample possibilities for falsifying any theory we might wish to test."

These arguments constitute a head-on challenge to those who argue that the laboratory environment is artificial and that, consequently, the results may be spurious because they fail to capture the complexities of real world economic environments. The Smith/Wilde position, like that of Plott, shifts the focus from comparison between lab and 'real world' to comparison between laboratory and theory. The laboratory microeconomy corresponds more closely with economic theory than does the real economy. Hence, if the theory fails in this context, there are good grounds for questioning the theory and being sceptical about whether its predictions will carry over to any more complex 'natural' economic environment. The burden of proof, or so the argument goes, should then be upon the advocate of the theory to give an account of why the theory has failed and why, if at all, the theory should still be believed as an account of behaviour in the more complex reality. This line of argument has been clearly endorsed in more recent discussions of experimental method[10].

There seems to be some merit in this argument but it is open to objections from at least two directions. In a response to Smith, Cross (1980, p.405) argues "it seems to be extraordinarily optimistic to assume that behavior in an artificially constructed "market" game would provide direct insight into actual market behaviour". Cross (op. cit., p.404) suggests the following as potential problems in experiments (they are not intended as an exhaustive list of objections):

> “(1) experimental situations often project a gamelike atmosphere in which a "subject" may see himself as "matching wits" against the experimenter-designer of the game. Even with relatively large payoffs, a subject may derive personal satisfaction from perceived "victories" which are not necessarily correlated with the performance indices used by the experimenter.
> (2) Experimental subjects are often cast in roles such as "seller," "dealer," or even "monopolist," and the subject may act in accordance with his own (mis)perceptions of these roles rather than in accordance with other incentives which may have been incorporated into that situation.
> (3) "real world" behaviour has usually been learned over many trials or over many years. The relatively short time horizons of experiments cannot hope to capture more than the behavior of the most naive and inexperienced actors who are found in the wider system.
> (4) Among the biological species, human beings are foremost in their capacity to control their own behavior through the implementation of abstract rules. Since these rules can be applied to a variety of different situations, human subjects usually carry many of them into the laboratory. Which of many possible alternative rules is to be applied in the experiment then depends on the background and experience of individual subjects and on their short-run interpretations of the nature of the situation."

The first two of Cross's comments can be interpreted as doubts about whether the Smith/Wilde precepts can be expected to apply in experimental settings. In each case, Cross is suggesting some possible motivation which may contaminate the data: the subjects' preferences are not those assumed by the experimenter. This points to an important qualification which is implicit in the Smith/Wilde argument concerning the *validity of auxiliary hypotheses*. The precepts are auxiliary hypotheses which are tested simultaneously with the hypothesis which motivates the investigator (e.g. does the market theory hold?). If the predictions of the theory appear to be falsified there are always a number of possible interpretations: one possibility is that the theory is false; another that the theory is true but some auxiliary assumption, necessary for the validity of the test, has been violated. It will always be open to the defender of the theory which has apparently failed some test to question, as Cross has done, whether the auxiliary assumptions hold.

To decide which inference to draw, one must make some judgement with respect to how likely it is that the auxiliary assumptions have held. Such judgements are likely to be problematic. For instance, to determine whether dominance has held, one needs to ask whether the rewards in the experiment were sufficiently large to outweigh other possible influences.
Since economic theories typically tell us little or nothing about how big incentives need to be to provide appropriate motivation, experimenters must make assumptions about what kinds of rewards are necessary. Smith and Wilde are not unaware of these lines of attack, but neither do they offer very much in the way of argument to head them off. For instance, in relation to the scale of incentive issue Smith (1982, p934) argues that the conventional approach is to use "payoff levels that are judged to be high for the subject population". But this clearly leaves room for the sceptic to offer the doubt that the rewards may not have been large enough.

Similarly, for salience and nonsatiation to hold, it is important that subjects properly understand the experiment, in particular, how the reward medium is related to their actions. Even if they do understand what has been described to them, there may be a questions concerning whether subjects trust the experimenter, that is, did they believe what they were told, even if they understood it? Can we ever be sure that subjects have understood an experimental design "sufficiently" well? There are strategies which experimentalists can, and often, do adopt for reducing the possibility of misunderstanding for example, opting for simple designs, ensuring that instructions are clear, allowing time for subjects to become familiar with the experimental setting and so on. There may also be possibilities for attempting to assess subjects' understanding, or to explore their trust in the experimenter(s) by including 'test' questions[11] or post-experimental questionnaires. But while such procedures may reduce the scope for scepticism, they are unlikely to eliminate the possibility of doubt. There is no logic of experimentation here which could dislodge the dedicated sceptic. The question must be whether the sceptical position appears a reasonable response to the carefully conducted experiment. This is an issue I shall pursue below.

The second two of Cross's objections suggest a different sort of worry, this time, reflecting upon the *generalisability* of the results from laboratory experiments. Objections 3 and 4 from Cross each propose reasons why behaviour in the laboratory might be different from 'real world' behaviour. The third point suggests that real world behaviour may be a product of learning and adaptation. If laboratories do not embody analogous mechanisms, or allow them sufficient time to operate effectively, we must expect systematic differences between the lab and the naturally occurring behaviour. Objection 4 suggests that behaviour may be sensitive to context since human beings interpret their surroundings to decide which modes, or rules, of behaviour are appropriate to their environment. Again, this suggests the possibility of a wedge between the lab and the world 'out there'. In each case, the wedge provides a possible buffer against apparently refuting evidence from the lab.

Smith argues that attacks along these lines do not subvert the role of experiments in theory testing. His strategy is to argue that theories which pay no reference to such effects can (and indeed should) be tested in environments in which such effects are absent. Smith (1977, p8) puts the case as follows:

> "if the purpose of an experiment is to test a theory, are the elements of alleged unrealism in the experiment parameters of the theory? If not then the criticism must be directed to the theory as much as the experiment".

Smith's argument seems to be that if economists believe that evolution, learning, context, or whatever else are important explanatory factors determining market outcomes, then those factors should be included in economic theories. Once they are, they too can be subject to experimental test. Again, Smith's point has some, force, but it is not immune to counter.

Suppose one were to take the view (i) that the phenomena of interest to economists are real market phenomena and (ii) that the job of the economist is to develop the simplest theory of market behaviour capable of predicting observed behaviour. Given this brief the economist may seek explanations of market phenomena which do not pay reference to all the details of the mechanisms which are responsible for producing that behaviour. There is a case for thinking that this is often what economists do. Take, for example, theories which predict the equilibria of economic systems as the behavioural outcome without necessarily explaining how the system reaches the equilibrium. If pressed to explain how such equilibria are in fact arrived at, an economist could respond that she believes non-equilibrium outcomes are unstable and that there are learning or evolutionary mechanisms which promote a tendency toward equilibrium. This was precisely Friedman's (1953) defence of the theory of competition against the observation that managers of firms were not actively or consciously attempting to optimize in the conventional sense. It is easy to think of other economic theories which predict behaviour without providing an account of how such behaviour is arrived at: for example, the rational expectations hypothesis is rather like this; one of the most basic ideas to economics - that agents are rational utility maximisers - is like this too.

While some might interpret such theories as partial explanations of economic behaviour, there are philosophical arguments which explicitly endorse this mode of theorising. From an *instrumentalist* perspective theories are interpreted purely as devices for predicting behaviour in specific contexts and there is no presumption that a theory should provide any realistic account, or explanation, of the process by which the predicted outcome is generated. An instrumentalist need not be impressed by Smith's defence of theory testing; since, on this view, theory is domain specific it may be argued that the laboratory experiment is not a context to which the theory is intended to apply. While instrumentalism has its own critics, there are other positions in the philosophy of science which endorse domain specific theorising such as the realist-localist stance defended by Cartwright (1996). If theories have restricted domain, an experimental test may tell us little or nothing about the validity of the theory as an account of the phenomena we a truly interested in. We may have learned that it does not apply in the lab, for sure, but the critic may argue that this is precisely why they are not interested in laboratory experiments; they never expected that their theories would apply in the laboratory in the first place.

5. On dry land without a paddle

Smith and Wilde have succeeded in establishing the following proposition: if an economic theory fails to predict behaviour in a context where the four precepts hold, then that theory can be rejected as an *entirely general* theory of economic behaviour. In this respect. Smith and Wilde are on firm ground, or so it seems to me. The practical significance of this proposition, however, is severely limited by two problems. The first problem is that it will *never* be possible to establish that the precepts have been satisfied in a given experimental test. It must always be a matter of judgement and, more pointedly, a matter of judgement about controversial matters such as whether the incentives were big enough, whether the subjects fully understood the design and so on. The second limitation stems from the fact that, in principle, theories can be insulated from refuting experimental evidence by restricting their domain in ways which render the evidence irrelevant. To illustrate the practical significance of these problems, consider this extract from Ken Binmore's book <u>Playing Fair</u> where he discusses the significance of experimental evidence:

> "But how much attention should we pay to experiments that tell us how inexperienced people behave when placed in situations with which they are unfamiliar, and in which the incentives for thinking things through carefully are negligible or absent altogether? ..... In brief, two questions about experiments with human subjects always need to be asked. Does the behavior survive when the incentives are increased? Does it survive after the subjects have had a long time to familiarize themselves with all the wrinkles of the unusual situation in which the experimenter has placed them? If not, then the experimenter has probably done no more than inadvertently to trigger a response in the subjects that is adapted to some real-life situation, but which bears only a superficial resemblance to the problem the subjects are really facing in the laboratory" (Binmore, 1994, pp. 184-5)

Binmore's quote brings both of the problems with the Smith/Wilde position into quite sharp focus. Take for instance, Binmore's comments regarding incentives. I take it that Binmore - perhaps quite rightly - feels that experiments must use adequate incentives before the results can be taken seriously. The problem is to know what constitutes 'adequate'.

This is an issue which can, and indeed has been investigated experimentally. A variety of studies have examined how incentive levels affect behaviour by manipulating the scale of payoff for a given task. Examples include Smith and Walker (1993a) and Hoffman et. al.(1996). Some authors have then sought to use the collected evidence across a range of studies as the basis for judgements about the impact of incentives. For instance, Smith and Walker (1993b) survey over 30 experimental studies which report data relating to the impact of incentives. Their conclusion is that incentives matter, and that higher incentive levels tend to produce behavioural outcomes more consistent with the predictions of economic models. Smith and Walker (ibid) suggest that these data may be accommodated by extending conventional theories to allow for the cost or effort involved in decision making. While this line of research shows that apparently awkward questions of interpretation can stimulate a positive research agenda, it may be naive to think that such research will produce unequivocal answers to the questions which motivate it. The data will not speak for itself, and ultimately, conclusions on these matters will turn on some degree of judgement leaving room for inter-subjective disagreement. A comparison of two recent assessments of experimental evidence relating to choice under uncertainty - one by Glenn Harrison (1994) the other by Colin Camerer (1995) - may illustrate the point.

Camerer argues that the evidence supports the view that violations of Expected Utility theory (EU) are both widespread and robust. Moreover, he suggests the evidence supports the view that increasing incentives makes little difference to conformity with the predictions of EU adding: "Indeed, some studies suggest that the main effect of paying subjects is a reduction in variance of their responses, which *increases* the statistical significance of EU violations" (p. 635). Harrison, on the other hand claims that:

"Many of the experimental anomalies that are claimed to violate EUT do not satisfy the

> Saliency requirement or, if they do, generally fail to satisfy the Dominance requirement for plausible (perceptual or motivational) threshold values”

In Harrison’s view, then, the appropriate conclusion to draw is that most experiments have not been properly designed: they did not satisfy the Smith/Wilde precepts. Thus we have two sharply contrasting commentaries on what the evidence relating to EU shows: one claiming it clearly demonstrates the inadequacy of that theory at the level of description, the other claiming it does not.

The Binmore quote also illustrates the point that there may be cases where some will feel able to discount the relevance of theory refutation, even if there were agreement that the Smith/Wilde precepts had been satisfied. Here I have in mind Binmore’s references to ‘familiarity’ and ‘adaptation’. Some would argue, I take it Binmore is among them, that the implications of some economic theories are only intended to apply to settings where subjects have received *appropriate* feedback on the consequences of their decision and had *sufficient* time to adjust accordingly. Using this line of argument, at a stroke, all of the evidence derived from settings in which individuals make one-off choices, or even repeated choice but without adequate feedback or learning time are dismissed as irrelevant to assessment of the given theory.

Once again, the problem revealed here - that theorists are able to head off disconfirming evidence by suggesting that their theory is not intended to apply in the setting where the test was conducted - would be of little concern if there were general agreement about where a particular theory, or class of theory, is to apply. It is far from clear, however, that any such agreement exists. For instance, for Charles Plott (1991) part of the “logic” of experimentation is that “General theories must apply to simple cases”. But is there anything that binds theorists to interpreting particular economic theories as general theories in the first place? The Binmore example suggests not and it is easy to find other counter-examples.[12]

6. The Rhetoric of Economics and Experimental Economics

The problems I have been discussing all relate to how one should interpret the result of an empirical test. The problems are by no means peculiar to testing in experimental economics - they are common to all forms of empirical investigation. There is a well known and widely accepted argument in the philosophy of science that the results of empirical tests are always open to more than one interpretation. The argument, which is often referred to as the Duhem-Quine[13] thesis goes something like this. In any test of a hypothesis, one is always testing a joint hypothesis which consists of some target hypothesis (the one the investigator really wants to test) combined with a variety of auxiliary hypotheses. A hypothesis could be rejected, even though it is actually true, because some auxiliary assumption has failed. Likewise, we may fail to reject a hypothesis which really is false because of the failure of some auxiliary hypothesis. Since one can *never* be certain that all of the auxiliary hypotheses necessary to bring a hypothesis to test have held, there is an irreducible uncertainty in the interpretation of empirical results and, more specifically, there is no such thing as a crucial experiment.

Each of the problems I have been discussing in this paper can be interpreted as instances of this more general problem. The Smith/Wilde precepts can be thought of as auxiliary assumptions which need to hold for us to have a straightforward interpretation of the results bearing on the main hypothesis at test. Similarly, for a theory to count as a test of a hypothesis even when the precepts do hold, we need a further auxiliary hypothesis: that the theory concerned applies to the experimentally generated context.

It would be wrong to suggest that experimentalists have been unaware of the issues raised by the Duhem-Quine problem and some have explicitly addressed the topic. Smith (1994) suggests that much of the research undertaken by experimentalists is stimulated, if only implicitly, by the sorts of doubt which arise from the Duhem-Quine problem.

> "Experimental economists are intuitively if not formally aware of the problem; this is why they do so many experiments probing the sources of a theory's failure, or success,......(p127)....When a theory works well, they push imaginatively to find deliberately destructive experiments that will uncover its edges of validity.....when a theory works poorly, they re-examine instructions for lack of clarity, increase experience level of subjects, try increased payoffs, and explore sources of "error" in an attempt to find the limits of the falsifying conditions; again, this is for the purpose of better understanding the anatomy of a theory's failure, or the procedures for testing it...." (p.129).

In Smith's (1994, p.127) view "some philosophers have exaggerated the significance of the Duhem-Quine problem". Part of my concern is that, at times, the rhetoric of some experimentalists displays an overly cavalier attitude to the problem. There are some stark examples of this. Take, for example, Roth's (1988) survey of experimental economics. On page p.974 he defines a laboratory experiment as a setting in which "the economic environment is fully under the control of the experimenter". In a later survey he writes that the controlled economic environment allows "observations to be unambiguously interpreted in relationship to the theory" (Roth 1995, p.22). Notice that there is no scope here for the doubt which the Duhem/Quine problem asserts we *always* confront. In fairness to Roth, these may be no more than off-hand asides not intended for close methodological scrutiny. Amongst the explicitly methodological writings of experimentalists, however, lurk more subtle attempts to play down the significance of Duhem-Quine. For instance, in the introduction to the paper where he discusses the conditions under which experiments constitute real economic environments, Smith (1982) gives a spin to the paper which de-emphasises the Duhem/Quine problem when he writes "An important message of the paper which has been set out before ... [Smith then cites references] ..... is that economies *are* real live economic systems" (emphasis added). This interpretation of the paper glosses over two important caveats: (i) the experiment is real (in smith's sense) only *if* the precepts are satisfied; (ii) knowing whether they are satisfied is problematic.

My hunch (though this is hard to establish) is that pioneering experimentalists like Smith and Plott were, in their methodological writing, attempting to persuade an audience of general economists that experimental techniques provided a worthwhile approach for economics. At the time they were writing it seems reasonable to suppose that the prevailing view was the sceptical one: that economics is a non-laboratory science. As such, it would be understandable if those promoting experimental methods saw themselves as fighting a tough corner in which a robust defence of the method was called for. If so, the incentives were there to be up-beat about the benefits from experiments while playing down the difficulties of interpretation. In Section 3, I attempted to make the case that there is no underlying philosophical objection to the application of experimental techniques in economics (or at least none that I can discern). While I stand by this argument, my second aim has been to make the case that there are deep difficulties of interpretation which surround experimental results. My concern is that these difficulties have not received sufficient attention from experimental economists, and that an official rhetoric has emerged - deriving from the Smith/Wilde/Plot discourse - which has tended to bury them. Given the greater (though still somewhat controversial) status accorded to experimentation, I would suggest that the time has come to be more open about the difficulties presented by these issues of interpretation. Let me explain why.

I am struck by a possible parallel in the development of econometrics. Thomas Mayer (1980) explains how the pioneers of econometric methods - many of them writing in the early volumes of Econometrica - promised that econometric techniques would form a bridge between theory and reality; at last economic theories would be confronted by the data in a way which would allow disputes between economists to be finally resolved. For instance, Irving Fisher writing in the Journal of the American Statistical Association (1933, p.3) made this remark about the advent of the new econometric age; he said "I hope it may mark the beginning of the end of 'schools' in the sense of partisan groups or cults". With hindsight, such claims seem extraordinarily bold. Mayer observes that econometrics failed to resolve disputes, different economists were able to assemble econometric results supporting their own point of view and instead of just waving their arms at one another, ended up waving their t-statistics at one another. Good jokes about econometrics are now easy to find: for those interested Leamer (1983) tells many of them but with a heavy heart describing the state of econometrics as "a sad and decidedly unscientific state of affairs...... Hardly anyone takes data analysis seriously, Or perhaps more accurately, hardly anyone takes anyone else's data analysis seriously" (p. 37).

What happened to the econometric dream? Relatively few people have written on this topic, but it is interesting that among those who have there seems to be a unanimous view that there is nothing wrong with econometric techniques per se. The techniques themselves are quite sophisticated and statistically well-founded[14]. The diagnosis of, among others, Mayer, Leamer, Grahl (1977) is to point to a lack of methodological sophistication amongst practitioners and economists more widely. Each suggest that economists have been guilty of operating within an overly simplistic methodological framework; one which plays down philosophical difficulties of interpretation at the interface between theory and evidence, and generates a misplaced sense of objectivity[15]. This engendered over-optimistic expectations in relation to what the methods of econometrics could achieve. Consequently, when econometrics failed to deliver the goods, economists became disillusioned.

I sense that a similar danger exists for experimentalists today. These days, the results of experimental investigation are reaching a wide audience of economists. The rhetoric of experimental economics suggests the potential for an objective confrontation of theory with fact under controlled laboratory settings. But it also seems clear that after something approaching fifty years of experimentation, many basic questions which motivated experimental studies remain highly contentious. Take, for example, research into expected utility theory; one of the earliest targets for experimental testing in economics. If we were to ask a wide range of experimentalists a simple question like "do the results of experimentation support expected utility theory" we should expect a diversity of answers[16]. That being the case, will experimental techniques become the focus of the next generation of economic satirists?

Some outside observers have expressed doubts as to whether much progress in economics has been achieved as a result of experimentation. For example, Frank Stafford (1980, p.410) in a comment on two papers by Smith (1980) and Kagel and Battalio (1980) argues that while experimental results have been used to reinforce certain beliefs that economists have traditionally held they "have not yet resulted in new beliefs or in new ways of conceptualizing economic problems". In my view, such a position would be hard to defend today. Many new streams of theory have been stimulated, and are continuing to evolve, in response to the challenges set by experimental evidence: examples include new theories of choice under uncertainty (see Sugden 1998) and advances in game theory (see Camerer 1997). Indeed a case can be made for thinking that experimental economics has generated a number of progressive research programmes (see Smith et. al. 1991, for example). Furthermore, experimentalists can claim credit for developing some new and distinct approaches to tackling economic problems. One such approach would be the use of experiments as test beds for institutional design (see Banks et al 1989; McCabe et al. 1989,1991; Backerman et al. 1997). This genre of experiment provides a social science analogue of the wind-tunnel: in this case the “models” being tested are typically new institutional forms, such as proposed regulatory mechanisms. The purpose is to investigate the behaviour of these proposed institutional designs as a prelude to potential application in the field.

Notwithstanding these advances, there are some areas of research where it is harder to see what the, sometimes quite extensive, volume of research has accomplished. Consider, for instance, research in relations to the preference reversal phenomenon. This phenomenon (first reported by Lichtenstein and Slovic (1971) and Lindman (1971)) has been observed in experiments of the following basic form. A subject is asked to carry out two distinct tasks (usually separated by some other intervening tasks). The first task requires the subject to choose between two prospects: one prospect (often called the "$-bet") offers a small chance of winning a "good" prize; the other (the "P-bet") offers a larger chance of winning a smaller prize. The second task requires the subject to assign monetary values to each of the prospects. In terms of standard economics theory, the two tasks amount to different ways of eliciting the preference ordering of the P- and $-bets, hence, both tasks should produce the same ranking of options for a given individual. Repeated studies have found, however, that there is a significant tendency for individuals to choose the P-bet (task 1) while placing a higher value on the $-bet (task 2). Preference reversal looks like quite a serious issue: it not only violates EU, but seems to pose quite a challenge for economic theory in general as Grether and Plot (1979. p623) point out "It suggests that no optimisation principles of any sort lie behind even the simplest of human choices".

After a good deal of research on this phenomenon, what do we know? For instance: is preference reversal a real phenomenon? Surely we can agree that it does occur in specific experimental contexts. But, is it robust to increases in payoffs; does it occur in economically meaningful contexts; how should it be explained? Each of these questions remains highly controversial, or so it seems to me. Moreover, despite the very large volume of effort devoted to empirical investigation of this phenomenon there seems to have been very little effort devoted to developing or applying economic accounts of the phenomenon. Under these circumstances, an outside observer could perhaps be forgiven for reaching the conclusion that the sustained empirical effort involved here has had little impact upon economic thinking more broadly. However, to view this as evidence that experimental methodology is not worthwhile and should be curtailed before more research funds are expended on some blind alley would be to reach exactly the wrong conclusion.

Experiments *attempt* to confront theory with reality. In doing so puzzles, like the preference reversal phenomenon will arise. Understanding the nature of these puzzles is unlikely to be a straightforward business at least in part because of some of the problems discussed above. The question must be whether as a profession we are seriously interested in understanding such puzzles. If we are, then I would suggest that the way forward is to be less defensive in our attitudes and to be more upfront about the limits of our knowledge and methods. Let me provide a couple of suggestions about what such 'up-frontness' might involve.

Amongst theorists one can sometimes observe a tendency to play down the fact that it is rather unclear as to what contexts economic theories are intended to apply. For instance, Lipsey and Crystal (1955, p.31) argue that the statement of a theory will include assumptions which outline "the conditions under which the theory is to apply". In general, this is just not true[17]. As theorists, then, let us give up this illusion and engage in an open-minded debate about these issues. As experimentalists, for the most part, we simply don't know whether the results of laboratory experiments apply more generally in everyday contexts of economic significance. The official theory-testing rhetoric of experimentalists attempts to insulate us from this fact. As experimentalists, then, lets admit this limit to our present knowledge and pursue a broader exploration of the extent to which laboratory findings generalise outside the laboratory.

There are certain, potentially very important, obstacles to this 'upfront' strategy, many of them institutional. Research requires funding; funding is very often controlled by bodies who's interests are focussed on research with practical applications. Consequently, the research that is funded will often be that which can demonstrate 'relevance'. The strategy I propose urges us to acknowledge that, in certain important respects, economics is both theoretically and empirically disconnected from precisely those things which fund-holders may perceive relevant. While it would be naive to ignore this constraint, I would suggest that it is precisely the research which seeks to address these, currently, weak links that leads the way to increasingly practical and relevant economic knowledge.

A second, crucially important, institutional factor surrounds the behaviour of professional journals (or rather their editors and selected referees). The research agenda I am suggesting would imply researchers engaging in work which might be regarded as somewhat unglamourous. Think again about the preference reversal phenomenon. When this topic first hit the economics Journals it was regarded as exciting, challenging material and the early debate featured on the pages of top-flight international journals. Would the work of tenacious researchers pursuing a detailed and fine-grained analysis of its causes and significance catch the attention of an editor of a highly respected journal today? While it seems only reasonable to think that such work would be best suited to specialist journals, there must be some doubt about whether the incentives which operate within contemporary academia are suited to sustaining a continuing research programme along these lines.

These are genuine difficulties no doubt. But then the question must be are we, as economists, really interested in understanding the way that things work in the economic world "out there"? If so, then I would argue that we should support the contributions of the economists in white coats, even if their chore may seem dismal at times.

References

Backerman, S., M. Denton, S. Rassenti, V. Smith (1997) "Market Power in a Deregulated Electrical Industry: An Experimental Study". Mimeo, University of Arizona.

Banks, J., J. Ledyard and D. Porter (1989) "Allocating Uncertain and Unresponsive Resources: An Experimental Approach", Rand Journal of Economics, 20, 1-25.

Bernoulli, D. (1954), "Exposition of a New Theory on the Measurement of Risk", Econometrica 22, 23-26.

Binmore, K. (1994), Playing Fair, MIT Press.

Blaug, M. (1992), The Methodology of Economics, Cambridge University Press.

Bohm, P. (1994), "Behaviour under Uncertainty without Preference Reversal: A Field Experiment", Empirical Economics, 185-200.

Butler, D. and J. Hey (1987) "Experimental Economics: An Introduction", Empirica, 14, 157-86.

Camerer, C. (1995), "Individual Decision Making", in Kagel and Roth (1995, eds).

_________ (1998), "Progress in Behavioral Game Theory", Journal of Economic Perspectives, 11, 167-88.

Cartwright, N. (1996), "Fundamentalism vs. the Patchwork of Laws", in D. Papineau (ed.) The Philosophy of Science, Oxford University Press.

Cross, J. "Some Comments on the Papers by Kagel abd Battalio and by Smith", in J. Kmenta and J. Ramsey, eds., Evaluation of Econometric Models, New York University Press, 1980.

Caldwell, B. (1982), Beyond Positivism, George, Allen and Unwin.

Darnell, A. and J. Lynne Evans (1990), The Limits of Econometrics, Edward Elgar.

Davis, D. and C. Holt (1993), Experimental Economics, Princeton University Press.

Fisher, I. (1933), "statistics in the Service of Economics", Journal of the American Statistical Association, 1-13.

Friedman, M. (1953), Essays in Positive Economics Chicago University Press.

Friedman, D. and S. Sunder (1994), Experimental Methods, Cambridge University Press.

Grahl, J. (1977), "Econometric methods in Macroeconomics: A Critical Assessment", The British Review of Economic Issues, 11-37.

Grether, D. and Plot, C. (1979), 'Economic Theory of Choice and the Preference Reversal

Phenomenon. American Economic Review, 69, 623-38.

Harless, D. W. and C. Camerer (1994), 'The Predictive Utility of Generalised Expected Utility Theories', Econometrica, 62, 1251-89.

Harrison, G. (1994), "Expected utility and The Experimentalists", Empirical Economics, 19, 223-53.

Hey, J. (1991), Experiments in Economics, Basil Blackwell.

______ and C. Orme (1994) "Investigating Generalizations of Expected Utility Theory Using Experimental Data", Econometrica, 62, 1291-1326.

Hoffman, E., K. McCabe and V. Smith (1996) "On Expectations and Monetary Stakes in Ultimatum Games", International Journal of Game Theory, 289-301.

Hollis, M. (1994), The Philosophy of Social Science, Cambridge University Press.

Holt, C. (1995), "Industrial Organization: A Survey of Laboratory Research", Ch. 5 in J. Kagel and A. Roth (eds, 1995).

Kagel, J. (1995), "Auctions: A Survey of Experimental Research" in Kagel and Roth (1995, eds).

Kagel, J. and A. Roth (1995, eds), The Handbook of Experimental Economics, Princeton University Press.

Kagel, J and R. Battalio (1980) "Token Economy and Animal Models for the Experimental Analysis of Economic Behavior" in Kmenta, J. and J. Ramsey (eds, 1980).

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

Kmenta, J. and J. Ramsey (eds, 1980), Evaluation of Econometric Models New York: Academic Press.

Leamer, E. (1983), "Let's Take the Con out of Econometrics", American Economic Review, 31-42.

Ledyard, J. (1995), "Public Goods: A survey of Experimental Research", Ch. 2 in J. Kagel and A. Roth (eds, 1995).

Lichtenstein, S. and Slovic, P. (1971), 'Reversals of Preference Between Bids and Choices in Gambling Decisions. Journal of Experimental Psychology, 89, 46-55.

Lindman, H. (1971), 'Inconsistent Preferences Among Gambles', Journal of Experimental Psychology, 89, 390-97.

Lipsey, R. (1979), An Introduction to Positive Economics, fifth edition, Weidenfeld and

Nicholson.

Lipsey, R. & K. Crystal (1995), Positive Economics, Oxford University Press.

Loomes, G. (1991), Experimental Methods in Economics, in Companion to Contemporary Economic Thought (eds) Greenaway, D., M. Bleaney and I. Stewart, Routledge.

Mayer, T. (1980), "Economics as Hard Science: Realistic goal or Wishful Thinking", Economic Enquiry, 165-78.

McCabe, K., S. Rassenti and V. Smith (1989), "Designing 'Smart' computer-assisted markets: an experimental auction for gas networks", Journal of Political Economy, 5, 259-83.

______________________________ (1991) "Smart Computer-Assisted Markets", Science, vol 254, 534-54.

Palfrey, T., and R. Porter (1991), "guidelines for submission ofManuscripts on Experimental Economics", Econometrica, 59, 1197-98

Plot, C. (1991), "Will Economics become and Experimental Science", Southern Economic Journal, 57, 901-19.

Roth, A. (1987) (eds), Laboratory Experimentation in Economics: Six Points of View, Cambridge University Press,

_________ (1988), "Laboratory Experimentation in Economics: A Methodological Overview", Economic Journal, 974-1031.

__________ (1995),"Introduction to Experimental Economics" in Kagel and Roth (1995, eds).

Samuelson, P. and W. Nordhaus (1985) Principles of Economics, 12th edition, McGraw-Hill.

Smith, V. (1976), "Experimental Economics: Induced Value Theory", American Economic Review Proceedings, 274-9.

_________ (1980), "Relevance of Laboratory Experiments to Testing Resource Allocation Theory", in Kmenta, J. and J. Ramsey (eds, 1980)

_________ (1982), "Microeconomic Systems as an Experimental Science", American Economic Review, 72, 923-55.

________ (1989), "Theory, Experiment and Economics", Journal of Economic Perspectives, 3, 151-69.

_________ (1994), "Economics in the Laboratory", Journal of Economic Perspectives, 8, 113-31.

________ , K. McCabe and S Rassenti (1991) "Lakatos and Experimental Economics" in Niel de Marchi and Mark Blaug (eds) Appraising Economic Theories Edward Elgar.

________ , and J. Walker (1993a), "Monetary Rewards and Decision Cost In Experimental Economics", Economic Inquiry, XXXI, April, 245-61.

_____________________ (1993b), "Rewards, Experience and Decision Costs in First price Auctions", Economic Inquiry, XXXI, April, 237-245.

Stafford, F (1980), "Some Comments on the Papers by Kagel and Battalio and by Smith" in Kmenta, J. and J. Ramsey (eds, 1980).

Sugden, R. (1998), 'Alternatives to Expected Utility Theory', Handbook of Utility Theory, P. J. Hammond and C. Seidl (eds.), Kluwer. Forthcoming.

Thurstone, L. (1931), "The Indifference Function", Journal of Social Psychology, 139-67.

Wallis, W. and M. Friedman (1942), "The Empirical Derivation of Indifference Functions". In Studies in Mathematical Economics and Econometrics in Memory of Henry Schults, O. Lange, F. McIntyre and T. Yntema (eds.), University of Chicago Press.

Wilde, L. (1980), "On the Use of Laboratory Experiments in Economics", in Joseph Pitt, ed., The Philosophy of Economics, Dordrecht: Reidel, 1980.

Notes

1. Plott, C. (1991), "Will Economics become and Experimental Science", Southern Economic Journal, 57, 901-19.

2. Some would say that experiments in economics date back much further than this. For instance, Roth (1988) cites Bernoulli (1738) as an early example of 'informal' experimentation in economics. For reviews which discuss the history of experimental economics see Davis and Holt (1993; Friedman and Sunder (1994) and Kagel & Roth (1995).

3. Econometrica also published Palfrey's and Porter's (1991) "guidelines for submission ofManuscripts on Experimental Economics".

4. Hey (1991); Davis and Holt (1993); Friedman and Sunder (1994) and Roth and Kagel (1995).

5. The problem discussed here is a variant of the famous Allais paradox invented by Maurice Allais (1953).

6. In the original example the units of the payoffs are Israeli pounds.

7. For a discussion of some of the issues here see Hollis (1994, Ch.1)

8. This is subject, of course, to the usual provisos concerning sampling error and subject to the proviso that some other variable is not at work and so on.

9. I should make it clear that I am not endorsing such practices, and elsewhere, I have argued against the use of deceptive practices in economics experiments (see McDaniel and Starmer, 1998).

10. For example, Glenn Harrison (1994, p.223), speaks of the precepts as "widely accepted sufficient conditions for a valid controlled experiment". Similarly, Butler and Hey (1987, p.159) argue that "When these conditions are met, a microeconomic environment, where real agents make meaningful decisions, exists. It follows that by constructing the required institutions, competing hypotheses may now be tested."

11. One strategy used in some experiments has been to include dominated options in the choice set presented to subjects. It may seem reasonable to suppose that a subject who believed what they were told and understood the situation would never select a dominated option. The level of dominance violation might therefore constitute some proxy measure of the degree of misunderstanding (or mistrust) amongst the sample.

12. See, for example, Peter Bohm's (1994) discussion of the preference reversal literature which ends as follows: "Therefore, for preference reversal to be acknowledged as a blow to expected-utility theory and as a general problem that decision theory has to come to grips with, real-world issues must be identified where preference reversal is actually shown to exist" (p.196)

13. For a more extensive discussion of this point see Blaug (1992, p.18).

14. There is a nice parallel here between the Smith/Wilde precepts and certain of the results in econometric theory. For instance, if the standard Gauss-Markov assumptions hold, then it follows as a matter of logic that Ordinary Least Squares estimators will be unbiased and minimum variance. The difficult issue is whether these assumptions hold in particular applications. As with the experimental precepts, we can never be sure.

15. Take for example Leamer (1983, p36) who argues that "Economists have inherited from the physical sciences the myth that scientific inference is objective, and free from personal prejudice. This is utter nonsense.... the false idol of objectivity has done great damage to economic science" or Grahl (1977, p12) "With hindsight, it is clear that, as the research method proper to economic science, econometrics never had any substantive methodological underpinnings and that all its claims to logical rigour rest on a positivism so shallow that it would now probably be rejected as an illegitimate off-spring by the positivists them-selves."

16. Two papers which appeared in the same issue of Econometrica (November 1994) may illustrate the point. Camerer and Harless (1994) argue that analysis of a rich body of experimental data supports the use of non-expected utility models while only a few pages later Hey and Orme (1994) seem to conclude just the opposite.

17. One might object that I have drawn here on a simple text-book account of theory rather than a more careful and sophisticated methodological statement. But that begs the question of where students of economics learn the methodological conventions that they are supposed to adopt?