



Dundee Discussion Papers in Economics

On the Autonomy of Experiments in Economics

Martin K. Jones

Department of
Economic Studies,
University of Dundee,
Dundee.
DD1 4HN

Working Paper
No. 194
October 2006
ISSN:1473-236X

On the Autonomy of Experiments in Economics

Martin K. Jones*

Abstract: Most methodological discussion in experimental economics has been pursued by justifying the use of experiments as theory testing vehicles. More recently it has also been argued that the external validity of experiments requires the use of non- experimental field evidence. Therefore, it is argued, experiments are intermediaries between theories and field evidence. In this paper it is proposed that this picture of experiments is mistaken in the general case and that experiments can be justifiably undertaken as autonomous vehicles of discovery independently of theory-testing or field evidence.

JEL Classification: B41

Keywords: Autonomy, Experiments, External Validity, Theory Testing

* Department of Economic Studies, University of Dundee, Dundee. DD1 4HN.
M.K.Jones@dundee.ac.uk

Experimentation is now an accepted method of research within economics and, as a result there have been an increasing number of papers which have examined it from a methodological point of view. Typical examples of this phenomenon can be found in the Economic Journal symposium (1999) and the symposium in the Journal of Economic Methodology (2005). In the past few years methodological research has gone from a fairly rigid defence of experimentation while experiments were a novel technique in economics (See Smith 1982), to a more questioning attitude which attempts to show the strengths and weaknesses of the experimental technique.

The main thesis of this paper is that experiments are autonomous research activities which contribute in their own right to economic knowledge. This means that they can be wholly independent of theory testing and do not necessarily need to be externally validated by field studies (as argued by Guala 2003). This proposition is not novel- Sugden (2005) for example argues that experiments which investigate phenomena (known as “exhibits”) are as valid as theory- testing experiments.

However, while it is accepted that such experiments are possible there has been little real analysis of why these “exhibits” are valid or in what way they are methodologically useful. It is often the case that acknowledgement of their existence is followed by them being ignored, while the methodological analysis concentrates on aspects of theory testing (examples of this are Starmer 1999, Cubitt 2005 and Guala 2005). It is the aim of this paper to demonstrate that a certain group of experiments cannot be externally validated using field studies, while still being externally valid and do not rely on a theory testing methodology, while remaining very useful. In doing this I will draw on literature from the new philosophy of experimentation (c.f. Radder 2003, Hacking 1984, Franklin 1986). In the conclusion I will then argue that this analysis can be applied to experiments in general.

CAUSAL CONTROL EXPERIMENTS

In order to proceed with the analysis it will be necessary to define the class of experiments which will be used in the subsequent analysis. This class of experiment will intentionally be an extreme type which will help to clarify the argument in subsequent sections. Then, at the end of the paper, the arguments made will be

transferred to experiments in general. First of all we will divide experiments in economics into two types. The first type of experiment is that which includes all tests of theories. An example of this would be various tests of expected utility and the further tests of the successor descriptive theories (see Starmer 2000 for an overview). This is the conventional role of a theory- testing relationship which is analysed in Guala (2005). The second type of experiment is that which produces an “exhibit” (Sugden 2005). An exhibit is a phenomenon produced by an experiment which is constructed to produce it rather than to test a theory. An example of this is the long run of experiments which were used to prove and isolate the preference reversal phenomenon (see Guala 2000 for an overview). As Starmer (1999) points out, whether an experiment is an exhibit or whether it is theory testing is quite often a matter of circumstance. A test of a theory may turn up an anomalous phenomenon which is then isolated as an exhibit. On the other hand theories may be formulated and tested which are based on exhibits (Sugden 2005).

Following Guala (2005) and Siakantaris (2000) we will say that the primary aim of experimentation is to uncover *causal relationships* in the external world. In other words, we aim for explanation by finding out what causes are responsible for an effect. In the case of an economics experiment, the causes are usually found in the beliefs and desires of the agents taking part in the experiment or in the institutions set up within the experiment. In the former case, since in experiments we only currently have access to people’s behaviour, we *infer* the causal beliefs and desires from their behaviour.

Following Harre (2003) we will say that these exhibit producing experiments can be further divided into two types. The first type is the “domesticated system” type of experiment. In this case the experimenter tries to model the outside world in the experiment, using her understanding of how the system being studied works and the relevant causes. This “domestication” works by mimicking the system as it occurs in nature, while controlling inessential factors, and then running it using human subjects. The aim is to create a simplified version of the external world in the laboratory.

An example of this would be experimental markets where different market set- ups have been found to have drastic effects on behaviour (See Kagel & Roth ch.5 1995 for an overview). While the *institutions* created in such an experiment are all backed by economic theory, there have been experiments which have been set up to investigate effects rather than to test this theory. Examples of this include those

experiments set up to examine effects created by specific institutions existing in the external world such as posted price or uniform price auctions. Insofar as these markets replicate (as far as economists know) the workings of the market, while excluding the irrelevant elements which usually exist in real- world markets, then they can be called domesticated. Since, quite often, the effect being sought is not predicted by any theory, it should be described as an “exhibit”.

The second type of exhibit producing experiment is what will be described here as a Causal Control Experiment¹ (or CCE for short). CCEs do not aim to mimic economic systems but, instead, to find the cause(s) for a particular effect within the system. This is done by intervening in the workings of particular mechanisms within that system by using controls. These controls hold possible confounding factors constant while only allowing the perceived cause to be uncontrolled. The control on the perceived cause can then be varied to find out whether there is a corresponding effect as a result of the variation.

It can be seen that CCEs need not be realistic in the sense of representing any part of the economy in the same way that domesticated experiments do. The aim is purely to find causal relationships rather than to represent even a simplified model of the world. Given that an effect, such as preference reversal, may have multiple causes; it follows that isolating each individual cause will result in experiments where most causes of a phenomenon are effectively held constant. This would be highly *unrealistic* because in the external world all of the causes would be operating at once. However, in spite of this lack of realism these experiments are useful because they give an explanation for why effects happen.

An example of this type of experiment can be found in some of the research done into choice under uncertainty. Some of the phenomena found by experimentalists such as event- splitting effects, violations of monotonicity and intransitivities (see Starmer 2000) are experimental effects which have no real link to any theory and stand on their own as exhibits within experimental research. It is true that some of these experiments were done in the framework of a given theory and can be interpreted as *contradicting* various theories. However, they were often replicated in other experiments whose sole purpose was to follow up on the initial discoveries. Furthermore the experiments involve very tight, unrealistic controls on the types of

¹ Harre (2003) describes this as a “Bohrian” experiment but this label has little meaning in an economics context.

decision which can be made by subjects. They cannot be seen as representative of real decisions made in the external world. As such they can be seen as CCEs.

AUTONOMY FROM THEORY- TESTING

Having defined CCE's as a separate class of experiment it can be seen that they are also defined to be independent of theory. However, is this supposed independence genuine and, if so, is it valid? The latter is obviously believed to be so by Sugden (2005) but apart from his paper, none other has seriously analysed the idea of non theory- testing experiments. Starmer's (1999) defence of experimentation is presented purely in terms of theory testing. To take an example, when answering worries about how to tell whether Smith's (1982) precepts of experimentation are satisfied, his answer is essentially based around the idea of theory testing. He states that the theory being tested is always in a joint test with the auxiliary assumptions (such as Smith's precepts) and so one is faced with the Duhem- Quine problem² of how to deal with an anomalous result. Guala's (2005) comprehensive study of experimental methodology approaches the relationship between theory and experiment from essentially the same angle. Many of his chapters focus on traditional philosophy of science problems such as Bayesian updating and the Hypothetico-Deductive model, all of which assume the primacy of theory testing.

Ironically, therefore, there seems to be very little interest in the philosophy of *experiments* as opposed to the philosophy of theory testing as applied to economic experiments. This is quite unlike a movement which has developed to create just such a set of ideas within the philosophy of natural science (see Hacking 1983, Radder 2003, Franklin 1986³). This philosophy developed in reaction to a perceived excessive focus of philosophy of science on theorising and the testing of theories. It also moved away from the purely logical approach often previously used to focus on how experiments were actually carried out. This philosophy concentrates on the *instruments* used in the experimental laboratory and also how these instruments both enhance and constrain the knowledge which can be obtained from experiments. It

² This is the idea that any individual test of a theory is always to some extent a joint test of the theory and the network of theories in the background which support the original theory's presuppositions.

³ It is true that the authors mentioned have been referenced by Guala (2000) and Morgan (2005) however, they have not drawn out the implications for this point of view.

should be emphasised that these “instruments” do not have to be solid pieces of apparatus such as dice or computers but could be any device used for eliciting or controlling behaviour.

Instrumentation within economic experiments varies from experiment to experiment and within individual experiments. Examples include the computers that are used to carry out many interactive experiments as well as the software that is run on them. These are enabling pieces of apparatus that allow the construction of the “institutions” in the experiment and so elicit behaviour from the subjects. An example of these is displays of game matrices on the screens of computerised experiments. There are also the *rules* of the experiment that allow for control of behaviour. An example of the “controlling” function of experimental apparatus includes the common custom in game theory experiments of keeping subjects isolated and anonymous. The final function of experimental apparatus is to measure outputs from the experiment. Examples of this are the binary lottery mechanism as well as the Becker- de Groot- Marschak elicitation device (Becker et al. 1963).

The *point* of instruments within an experiment is to elicit phenomena (Hacking 1983). One uses instruments to control causes and to create an effect. This effect is not necessarily one which would ever exist in the outside world so there is the question of whether one can tell if a phenomenon is real or is simply an artefact of the experiment (Woodward 2003). Part of the answer to this is to eliminate experimental artefacts before the experiment in a process of calibration (see Franklin 1986, also Guala 2005). However, as Hacking states, the reality of the phenomena is ultimately assured when it can be *manipulated*.

Manipulation is a very important process in establishing the reality of a phenomenon. The example Hacking gives is that of a microscope where one can tell, for example, if one is injecting a fluid into a particular part of a cell because one can see the microneedle being pushed into the cell through the microscope. The act of pushing the needle through the cell wall correlates with one seeing the injection of the cell. This establishes that the absorption of fluid by the cell is a real thing and not an artefact because one’s manipulation corresponds with one’s observation in the experiment. At a more fundamental level, if one thinks that an instrument such as a particular (light) microscope is suspected of producing artefacts then one can apply an electron microscope to the same phenomenon. If the two (quite distinct) types of

microscope show the same effect then it is likely that the phenomenon is not an artefact and increases the likelihood of it being real.

A similar idea of manipulation can be applied within economic experiments. However, in contrast to the experiments described by Hacking (1983), economic phenomena tend to be much more difficult to manipulate than a simple microscopic observation session. Each manipulation may take one experiment to do, or at least one control group per experiment. For this reason an economic phenomenon may take longer to establish as being real. An example of this is in ultimatum game experiments where a variety of manipulations (Camerer 2003) have been tried to establish the reality of the phenomenon of equitable division between proposers and responders.

At the more fundamental level of testing instruments, there are examples of such testing although this type of activity is comparatively rare. The most complete example of this type of instrument- testing has taken place with the testing of the preference- reversal phenomenon (see Guala 2000 for a review). Part of the problem with some of the original experiments in preference reversal (e.g. by Grether and Plott 1979)⁴ was that they solely used the Becker- de Groot- Marschak (BDM) mechanism (an elicitation instrument) to measure the prices subjects were willing to accept for a lottery. However, the BDM mechanism depends on the subject being an expected utility maximiser. When the subjects were found to suffer from preference reversals then this assumption was violated. Karni and Safra (1987) picked up this problem when they demonstrated that the preference reversal phenomenon could occur if one assumed that the subject did not obey the independence axiom and one used the BDM mechanism to elicit preferences. This meant that the preference reversal was not a fundamental problem but simply an artefact of a far less fundamental problem (violation of independence) and an artefact of an instrument (the BDM mechanism). Tversky et al. (1990) demonstrated that this effect did not occur and that preference reversal existed independently of the BDM instrument.

If one is trying to find the causes of a particular phenomenon then using CCEs is, by definition, the best way to discover them as one can control the experiment without having to replicate a mechanism in the external world. However, it must be asked,

⁴ Although, ironically not by some of the very first experiments carried out by Lichtenstein and Slovic (1973)

what is the relationship between theory- testing and this type of exhibit? Just how free of theory-testing is it possible for an experiment to be?

It could be said that if, for example, one was testing for the effect of gender on behaviour in the ultimatum game (as in Eckel and Grossman 2001) then one could say that one was testing a hypothesis. This could be interpreted as a test of a hypothesis along the lines of “In all Ultimatum games, men offer less money than women”. In fact, in Eckel and Grossman’s paper this is roughly at the level of generality that the hypothesis was actually stated. It is this type of hypothesis that Guala (2005) calls a “low-level” hypothesis that can be fitted into a theory- testing framework.

As it stands there does not seem to be any need to distinguish between CCEs and other theory testing experiments. However it is plausible that many experiments don’t even start with this type of hypothesis but instead start as a “hunch” about how people will behave in an experimental environment. In that case a statement of a “hunch” would be something like: “In this particular ultimatum game environment, with subjects divided into males and females, men will offer less money than women”. This is not to say that all CCEs need to have statements this specific (and Eckel and Grossman certainly never did) but to point out that such statements are possible, if extreme.

Is this a hypothesis? I would suggest that it is one only in the most trivial sense that it is a statement predicting what will happen in that particular event. The statement effectively reflects the structure of the experiment and has no generality. This is not a theory in the sense of a generalised set of explanatory statements but rather a hunch about a possible phenomenon (See Radder 2003). However, notice that the experiment used to test it is *identical* to the one needed to test the more general hypothesis. There is no difference between the experiments *per se*, merely in the statements being tested. The test of such a statement is effectively identical with the experiment itself while the experiment consists of the whole domain of the supposed hypothesis.

However such a minimalist statement could be highly appropriate, for example, in the case where one has constructed a CCE that produces a phenomenon which has no existence in the outside world. This would occur because the phenomenon in a CCE may not exist independently of the instruments used to elicit it. If the causal factors controlled in the experiment are not, in fact, controlled then the phenomena elicited in the experiment would not occur. The phenomenon would be

jointly produced by the behaviour of subjects in the experiment together with the instruments used (Woodward 2003, Hacking 1983).

The problem is not that one cannot form some kind of description of the research being carried out in such an experiment. This can obviously be done. Instead the problem is that such a statement would be trivial and have minimal explanatory power. Because of this many CCEs cannot be said to “test” any hypothesis as this would empty the term “test” of any meaning. However this does not mean that such experiments are useless, as they enable us to explore causal linkages in the world, even though the phenomena produced may not exist outside the laboratory.

This is not to say that there are no theoretical entities involved in an experiment. On the contrary, experiments have large amounts of theoretical knowledge embedded in the instruments used in the experiment. To take an example, the Becker- De Groot- Marschak elicitation device presupposes that the subjects are expected utility maximisers. Another example is the isolation of individuals in interactive experiments, which derives from theoretical concerns about them acting as autonomous agents. Furthermore, there is nothing in what has been said above to suggest that experimental results are not theoretically interpreted. Obviously they are, as Sugden (2005) has pointed out. What is denied here is that experiments are necessarily guided by, or are testing, some non- trivial hypothesis.

However, if one assumes that the instruments and behaviour jointly produce an effect in a particular experiment then how can we derive any general hypothesis from it? The answer is through the process of manipulation described above. A series of experiments must be carried out to isolate the phenomenon by manipulation using different treatments and by using different instruments to measure and elicit behaviour. At the end of this process the phenomenon can be seen to be *independent* of the instruments used to elicit it. When one gets to this stage then the phenomenon can be accepted as being real and used to generate more general hypotheses.

The emphasis in this section has been on the reality of the phenomenon being examined in an experiment. This has an important implication for the suggested autonomy of CCEs from theory in that it is claimed that the causal dependencies highlighted by an experiment are independent of the theory used to interpret its results (Franklin 1999). The economic theory behind (say) expected utility may change but the behaviour elicited by the instruments in an experiment remains the same. In other

words the *meaning* of the behaviour may be determined by current theory⁵ but the material facts behind it are not determined by this meaning. (Woodward 2003). As such experimental results may survive theoretical changes and remain important even under new theoretical paradigms. This suggests that CCEs are, given the caveats above, autonomous of theory.

AUTONOMY AND EXTERNAL VALIDITY

External validity has recently become a serious topic of debate amongst experimentalists (see Schram 2005, Starmer 1999, Bardsley 2005, Siakantaris 2000 and Guala 2002, 2005). Much of this debate follows on from similar concerns which have been circulating in psychology for many years (see Lowenstein 1999). The problem of external validity⁶ is quite simply stated. Why should an experiment taking place in a laboratory, often involving an abstract problem be relevant to the external world? How can we guarantee that there is “parallelism” between what happens in the laboratory and what happens in the external world?

Guala (2005) has provided the most in-depth answer to the question of external validity. In his view experimentation is an intermediate stage between theory and the outside world. Theories are formulated by theorists and are then tried out in experiments to find out which are applicable (in the sense of Binmore (1999)) when confronted with real people in a laboratory session. These phenomena can then be assessed for their applicability in the outside world. For him, the role of experiments is to create a “library of phenomena” which can then be tried out in the outside world for a good fit to the facts. Guala’s main tool for assessing applicability externally is the “field study” whereby experimentally verified phenomena are examined in uncontrolled natural settings⁷. Experiments therefore are “intermediaries” in the sense that they act as possible arguments against the applicability of theories in certain circumstances and create interesting phenomena which may exist in the external world.

⁵ Indeed it seems impossible to argue otherwise (see Radder 2003).

⁶ I assume here that the so-called problem of artificiality is simply a version of the problem of external validity.

⁷ It is unclear whether “Field Experiments” (see List 2003) are included in this definition. One would suspect not as these experiments still involve some controls.

Guala's ideas, while being useful for certain experiments, such as the theory testing experiments or the "domesticated" type of experiment are not useful for CCEs. The latter experiments, by definition, are not representative of external phenomena since their main aim is to *disrupt* "natural" systems in order to achieve control. It is unlikely therefore that a field study could ever be found which would confirm the results of a CCE.

In fact, CCEs may involve many causes and effects which do not have the same pattern as causes and effects in the external world. The aim of a CCE may be to identify multiple causes of an effect (by holding all but one of the causes constant) or even to find out why an effect doesn't happen in the external world. The latter case could happen if, for example, two causes cancel each other out, leading to no change in the external world. An experiment in this case could isolate each cause and so create effects that do not happen in the external world (Woodward 2003). Such experiments would not only have no analogue in the external world but could be actively misleading if an attempt was made to find one.

If this is the case then how can the external validity of a CCE be established? It is insufficient to say that CCEs should be ruled out as legitimate experiments as it is obvious that they *are* useful for discovering causal mechanisms. To eliminate this type of experiment from the economist's armoury would be to cripple one's investigative ability and would also prevent large areas of causal linkage from ever being discovered. It is the idea that CCEs focus on controlling causes that allows us to establish external validity. If a CCE is not externally valid then it must be because the designer of the experiment has misconstrued causal linkages in the external world.

Suppose that a sceptic claims that an experimental phenomenon is not externally valid. If the phenomenon has been replicated one would then have to ask the specific reason for why it is not valid. The only reasonable response (ignoring blanket denials of experimental validity) is that there is another cause that affects the phenomenon in the external world but not in the experiment or there is an additional cause within the experiment that does not exist in the external world. In both cases the response would be that, in principle, an experiment could be undertaken to control for that cause. If this experiment was undertaken and the cause shown not to be a problem then that specific objection would be overcome. Testing for external validity, therefore, is a matter of undertaking experiments to eliminate potential confounding

factors. Starmer (1999) gives some examples of possible ways in which such claims can be dealt with in the experimental framework⁸.

This notion of external validity as an experimental problem rather than as a theory- testing problem gives us resources to tackle some of the problems set out in the literature. One is that highlighted by Bardsley (2005). Bardsley points out that the artificiality critique of experiments does hold when there are social relations in society that cannot be replicated within the laboratory. An example of this would be an experiment on tax evasion, where the subjects cannot realistically be forced into the role of taxpayers to the government if this is not actually happening within the laboratory. In essence they do not perceive themselves as being in the position of taxpayers and do not see the experimenter as being the government. One cannot therefore make a prediction about how they will act in an experiment in which they do not have the relevant expectations.

This is a general problem that ties in with Starmer's (1999) worries about the "practicality" of some experiments. However it does fit in quite neatly with the focus on instrumentation in this paper. The problem is not artificiality as such- as Smith (1982) says people are answering real problems with real money. The problem is that the instruments available in current experiments are not technologically sophisticated enough to induce the relevant beliefs in experimental subjects. If such instruments did exist then such experiments would be possible (if not necessarily ethical). As the growth of neuroeconomics (see Camerer et al. 2004 for a review) shows, the growth of technology can push experimental economics in directions never previously thought possible. However, it has to be accepted that this does create a tight constraint on the type of experiment that can be undertaken.

One very general line of attack on experiments that uses the notion of external validity is that of Siakantaris (2000). Siakantaris based his critique of experiments on two supposed problems with external validity⁹. The first problem is how to assess whether a given experiment is externally valid. This can be done by a field study or by another experiment. Since Guala (2002) has answered the case of the field study (and they cannot be done for CCEs anyway), I will focus here on the additional experiment. According to Siakantaris, if an additional experiment is carried out and it

⁸ The argument given here is very close to Starmer's although it is explicitly set out in terms of causal relations rather than in terms of theory testing.

⁹ Guala (2002) has put forward a critique of Siakantaris. However the aim here is to answer Siakantaris' points from an instrument- based point of view.

is found that the first experiment is externally valid then there is the question of whether the second one is also externally valid. As the second experiment will require a third and so on, this leads to an infinite regress. If the first experiment turns out *not* to be externally valid there may then, again, be doubt as to the external validity of the second experiment. If the latter is invalid then the first experiment may actually be valid since the validation experiment was worthless. This leads to a paradoxical situation of alternating answers. Even if the second experiment is valid then this needs to be proved which leads, again, to an infinite regress of further experiments. External validity therefore cannot be proved for any given experiment by experimental means.

From the point of view of this paper, the main problem with this argument is that it defines the external validity of a phenomenon as a binary property where an experiment is either externally valid or not. The argument given here is that the external validity of an experiment cannot be definitively tested but one *can* isolate various causes which may act as confounding factors in the external world. These causes can be tested in independent experiments. There would never be one experiment to test for external validity as such. Instead there would be a series of experiments that would establish the reality of the phenomenon and then would establish whether it replicated causes existing in the external world. If one accepts that a phenomenon is real then the only remaining question is whether the pattern of causes in the experiment corresponds to the pattern of causes in the external world.

Siakantaris' second criticism is of the ability of experiments to properly isolate a part of the external world. This is because the "holistic" nature of the external world is so full of interrelationality that phenomena cannot be isolated in the laboratory. This critique is partly claiming that the social world is so complicated that it cannot be replicated and is partly about the inability of certain social relationships to be replicated in the laboratory. The latter has been dealt with above when commenting on Bardsley's paper. In the former case it can simply be said that Siakantaris' characterisation of the social world is unacceptable. While it is true that some areas of the social world may be too complex for experimental research this cannot be the case in all areas since this would prevent any kind of theorising about causes. If one can pick out certain causes as being more important than others then these causes, allowing for social relationships, can be tested in the laboratory.

CONCLUSIONS

This paper has outlined a set of reasons for why a certain class of experiments in economics- known as Causal Control Experiments- can be considered to be autonomous both of economic theory and of the necessity of Field Studies to establish External Validity. Contrary to Guala (2005) therefore, CCEs are *not* intermediaries between theories and the external world. Furthermore CCE's are not simply marginal cases but are an important class of experiment in that they investigate the causes of phenomena while not directly trying to model the external world.

However, should these conclusions be restricted to CCE's? I would argue not. Most of the arguments made here are quite general and are not restricted to CCEs. The validity of running experiments in the form of CCEs is based around investigating causal links and the reality of phenomena in experiments. However, exactly the same argument applies even if one is testing a theory. Phenomena will still be real and one will still be investigating causes even if these causes are specified by a particular theory. Similarly for external validity: one may be able to use a field study to establish external validity but this does not mean that one cannot establish it by running further experiments.

It follows that the argument given here applies to all types of experiment and not just the special case of CCEs. While the arguments given for external validity or for theory- testing by Guala (2002) and Starmer (1999) are valid¹⁰, they should be seen as *additional* to those given above. Furthermore these theory based arguments are not valid for all experiments but only to special classes of experiment. Experiments can be justified on their own without recourse to theory- testing or external validation.

The Instrument- based point of view also allows one to understand many of the habits which experimenters have developed within economics (and other experimental disciplines). One point that has been emphasised within the philosophy of experiment literature is the distinction between the theory that supports the instruments within the experiment and the phenomenon that is being tested. The two should not interact in a manner that will cause the results of the experiment to be biased (Hacking 1983). Ideally the theory supporting the instruments and the theory

¹⁰ Strictly Guala (2005) has moved away from a theory- testing view towards Binmore's (1999) idea of testing for the applicability of theories. However, the arguments given here apply in both cases.

being tested should be completely separate. This has become plain with the use of the BDM device when testing for preference reversal.

Another implication of an instrument- based point of view is that repetition of experiments becomes an *essential* part of the experimental process. The main reason for this repetition is to establish the reality of the phenomenon via manipulation. By examining the phenomenon using different treatments one can establish that it is not just an artefact but is a genuine phenomenon. In addition one can establish that it is a phenomenon that is independent of the particular instrument used¹¹ to elicit it. Not only does this view endorse extensive repetition to establish a phenomenon but it also requires the experiment to be calibrated (Franklin 1999). Calibration, in the form of pilot experiments, feedback questionnaires and tests before the experiment allow potential artefacts to be eliminated if, for example the subjects fail to understand the experiment or if the instructions are misleading.

The emphasis of the argument in this paper is on the fact that experiments are not simply devices for picking up observations of the world or for testing theories. They have their own complicated structure based around the fact that they use instruments to access phenomena. This allows them to be used to investigate phenomena autonomously of theory and of any potential field evidence. It allows the development of scientific, economic research programmes independent of theory and field studies which focus around experimental phenomena. This is a useful attribute and it also places emphasis on certain aspects of experimentation- such as repetition or calibration that are largely ignored by conventional methodology. This is not to say that experimentalists and methodologists have not endorsed these aspects of experimentation. Indeed they are an integral part of the methods used by experimentalists. However they have not been given an adequate justification and this is precisely what is done by the instrument- based viewpoint.

¹¹ Indeed it is the absence of such consistency between instruments that causes preference reversal. The requirement of “procedure invariance” is precisely the requirement that different measuring instruments come up with the same preference ordering.

BIBLIOGRAPHY

- Bardsley N. (2005) "Experimental Economics and the Artificiality of Alteration" *Journal of Economic Methodology* 12 p. 239-251
- Becker G., DeGroot M. & Marschak J. (1964) "Measuring Utility by a Single Response Sequential Method" *Behavioral Science* 9 226-232
- Binmore K. (1999) "Why Experiment in Economics?" *Economic Journal* 109(453); F16-F24
- Camerer C. (2003) *Behavioral Game Theory* Princeton University Press Princeton NJ
- Cubitt R. (2005) "Experiments and the domain of economic theory" *Journal of Economic Methodology* 12 p. 197- 210
- Camerer C. Loewenstein G. & Prelec D. (2004) "Neuroeconomics: Why economics needs brains" *Scandinavian Journal of Economics* 106 555-579
- Eckel C. & Grossman P. (2001) "Chivalry and Solidarity in Ultimatum Games" *Economic Inquiry* 39; 171-88
- Franklin A. (1986) *The Neglect of Experiment* Cambridge; Cambridge University Press
- Franklin A. (1999) *Can that be Right? Essays on Experiment, Evidence and Science* Boston Studies in the Philosophy of Science vol. 199 Kluwer academic Publishers Boston
- Grether D.M. & Plott C.R. (1979) "Economic Theory of Choice and the Preference Reversal Phenomenon" *American Economic Review* 85: 260-66
- Guala F. (2005) *The Methodology of Experimental Economics* Cambridge Cambridge University Press
- Guala F. (2002) "On the Scope of Experiments in Economics: Comments on Siakantaris" *Cambridge Journal of Economics* 26: 261-67
- Guala F. (2000) "Artefacts in Experimental Economics: Preference Reversals and the Becker- De Groot- Marschak Mechanism" *Economics and Philosophy* 16: 47- 75
- Hacking I (1983) *Representing and Intervening* Cambridge; Cambridge University Press
- Kagel J. & Roth A (1997) *The Handbook of Experimental Economics* Princeton University Press Princeton NJ
- Karni E. & Safra Z. (1987) "Preference Reversal" and the Observability of Preferences by Experimental Methods" *Econometrica* 55: 675-85

- Lichtenstein S. & Slovic P. (1971) "Reversals of preference between bids and choices in gambling decisions" *Journal of Experimental Psychology* 89: 46-55
- List J. (2003) "Does market experience eliminate market anomalies" *Quarterly Journal of Economics* 118: 41- 73
- Loewenstein G. (1999) "Experimental Economics from the Vantage Point of Behavioral Economics" *Economic Journal* 109(453): F25-F34
- Morgan M.S. (2005) "Experiments versus Models: New Phenomena, Inference and Surprise" *Journal of Economic Methodology* 12 p. 317-329
- Radder H. (2003) *The Philosophy of Scientific Experimentation* Pittsburgh; Pittsburgh University Press
- Papers:
- Harre R. "The materiality of instruments in a metaphysics for experiments"
- Heidelberger M. "Theory- ladenness and Scientific Instruments in Experimentation"
- Radder H. "Technology and Theory in Experimental Science"
- Woodward J. "Experimentation, Causal Inference and Instrumental Realism"
- Schram A. (2005) "Artificiality: The Tension between Internal and External Validity in Economic Experiments" *Journal of Economic Methodology* 12 p. 225- 237
- Siakantaris N. (2000) "Experimental Economics under the microscope" *Cambridge Journal of Economics* 24; 267-81
- Smith V. (1982) "Microeconomic Systems as an experimental science" *American Economic Review* 72, 923-55
- Starmer (1999) "Experiments in Economics: should we trust the dismal scientists in white coats?" *Journal of Economic Methodology* 6 p.1-30
- Starmer (2000) "Developments in non- expected utility theory: The hunt for a descriptive theory of choice under risk" *Journal of Economic Literature* XXXVIII; 332-82
- Sugden R. (2005) "Experiments as Exhibits and Experiments as Tests" *Journal of Economic Methodology* 12 291-302
- Tversky A., Slovic P. & Kahneman D. (1990) "The Causes of Preference Reversal" *American Economic Review* 80 204-17