Dundee Discussion Papers in Economics

Against Applicability: A critique of Guala's Methodology of Experimental Economics

Martin K. Jones

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

Working Paper
No. 205
December 2007
ISSN:1473-236X

Against Applicability: A critique of Guala's Methodology of Experimental Economics

Martin K. Jones*

Abstract: The methodology of experimental economics has developed rapidly over the last ten years with many exciting debates within the field. One of the main contributors to this debate has been Guala who has written several articles and a well-received book on the subject. This paper argues that, while much of what he argues is correct, his views on external validity are not justified and the conclusions which he draws from these views could fatally undermine the experimental economics enterprise. In rejecting the justification of these views, the paper reaffirms the importance of the experiments in economics.

JEL Classification: B41

Keywords: Experiments, External Validity, Applicability

* Economic Studies, School of Social Sciences, University of Dundee, Dundee. DD1 4HN. Tel: (01382) 345164. e-mail: m.k.jones@dundee.ac.uk fax: (01382) 344691

Introduction

Recently the methodology of experimental economics has become a flourishing sub – discipline within experimental economics, attracting both practicing experimentalists and philosophers to the debate. This debate has resulted in an in-depth examination of the assumptions and methods used by experimentalists. As experimental economics becomes more accepted as a method within economics then this examination becomes more important as we need to understand the strengths and weaknesses of these methods.

One philosopher who has taken a keen interest in experimental economics and has been central to most debates on the subject is Francisco Guala. In a series of papers (Guala 1998, 1999, 2003, 2005b) and a book (Guala 2005a) he has put forward a view of experimental economics which, if accepted, would have serious repercussions on the way in which experiments are done and the way in which experiments are evaluated. This view places experiments as an intermediary between theoretical modelling and the external world where experiments do not "test" theories but instead act as filters, examining the applicability of models which can then be compared with evidence in the external world. Under this interpretation, experiments are nearer to models than to theory testing vehicles and so should be seen as methods by which economists can manipulate ideas rather than as empirical tests.

This view of experimentation is radical in that it "blunts" the edge of many of the claims made by experimentalists (c.f. Smith 1982, Plott 1991, Starmer 1999). If experiments cannot be used for testing then their results are of a secondary nature- they need to be further compared with external world evidence before they can be said to be empirically interesting. Experiments can be avoided altogether, even in areas where it is a practical possibility, if one has a sufficiently rich model. Under this view experimentation is simply a method of accumulating phenomena which may or may not be useful in the external world.

The aim of this paper is to critique this point of view and to demonstrate that it relies on too- sharp distinctions and a false picture of how experimentalists go about their work. However, the paper will also go deeper in analysing why applicability is a poor method and why the current beliefs of experimenters in what they are doing is far more useful than is characterised by Guala. In this critique we will not try to criticise all of his claims as this will take up far too much space (and indeed, many of his claims are reasonable[1]). Instead we will look at particular claims which seem to be central to his characterisation of experimental economics.

Preliminaries- Cartwright's work

The first problem when discussing the issues is to distinguish Guala's claims from other claims which have been made in the literature. In particular we wish to look at those philosophers who have put forward methodologies which have similar points to those put forward by Guala. This allows us to distinguish Guala's novel contributions from that of other philosophers whose work is similar in some respects but varies widely in others.

In particular we will look at another philosopher of science who has published some work on experimental economics, namely, Nancy Cartwright. Superficially, her work "The Dappled World" (Cartwright 1999) seems similar to that of Guala's but is different at many points. It will be shown that many of the critiques offered by Cartwright of the usual model of scientific methodology, while similar to some of those offered by Guala, differ in certain respects crucial to this paper.

Cartwright argues strongly against the idea that the aim of experimentation is to test "laws of nature" or even universal theories. In her view this is a metaphysical mistake as these so called "universal" theories are almost never universal. There are always exceptions to the rule and so theories only hold *ceteris paribus*. The only time laws can be said to hold are during experiments, which tend to have many factors deliberately controlled and so are heavily "shielded" from the external world.

Cartwright puts forward an alternative metaphysical picture- that of *capacities*. An entity in a theory has the capacity to do something if there is a tendency for an action to occur. So, for example, a magnet has the capacity to be attracted to a piece of iron. This is not universal because something could get in the way- a piece of copper could interpose itself for example. However the tendency to attract iron still exists. Capacities allow us to generalise across circumstances including those in the external world and those in experiments.

[1] Indeed it is worth saying that I have little quarrel with the first six chapters of Guala's 2005 book and I think that much of what is said there is fair and reasonable.

Regularities in the world are not the result of universal laws but are the result of “nomological machines” which link together in a system a variety of capacities and allow them to interact with each other. An experiment is set up as a special type of nomological machine which suppresses certain capacities and allows others to operate without hindrance. Generalisations emerge from the bottom up as a result of a particular network of capacities being replicated in different contexts. This means that a regularity in an experiment can be isolated across a series of experiments and gradually generalised as it is shown to operate with differing systems.

In economics models therefore are *designs* of appropriate types of nomological machine. However, because of the large number of factors involved in economics, regularities tend to be few and far between. In a similar ways to physics, some of these regularities can be highlighted in experiments but, in general, economists, as with physicists, tend to work with models and experiments tend to be constructed to test these models. Some mechanisms, such as demand, tend only to work within a wider system such as a market, so isolating this as a regularity, as opposed to within a model, in an experiment is not useful. Physics and economics are therefore similar because they tend to use models rather than general theories.

The idea that theory testing is a poor reflection of how science actually operates is reflected in work by Hacking (1983), Kincaid (1996 Ch. 3) and Morrison & Morgan (1999). In all of these cases it is demonstrated that discussion about theories and theory testing actually bears little resemblance to what scientists do. Instead scientists build models and then test these models against the external world. Models however also act as mediating instruments in that they act to represent how the world works and to allow the modeller to understand it. Models are often constructed *using* theories but they are not purely theory derived, instead deriving parts from disparate sources, including theories, but also empirical sources and even arbitrary modelling decisions. Significantly, models can also be used to represent the external world as well as representing theories.(Morrison & Morgan 1999). This means that they provide a functioning “mini- world” including the main known causes and effects.

Guala’s ideas

Guala’s work is primarily associated with experiments in economics rather than other sciences so one of the first distinctions he makes is designed to explain why these experiments are different from those in physics. He uses a claim made by Hacking about laboratory sciences that they are “those whose claims to truth answer primarily to work done in the laboratory” (1992). Guala then defines a “Non-laboratory science” as one whose claims to truth do not and cannot answer primarily to work done in the laboratory because the aim is to explain and control non-laboratory phenomena (Guala 1998, 2005a p.209).

Guala therefore sees a fundamental difference between laboratory sciences such as physics and non-laboratory sciences such as economics. In physics (as well as chemistry and biology), one is dealing with idealised circumstances where the entities controlled are often unnaturally pure and are often heavily shielded from the external world. Of course this “shielding” exists in economics experiments but in physics it doesn’t matter as most theories are concerned with the pure entities and the unnatural environments anyway. The theories in physics answer primarily to the evidence gathered in experiments and when one does experiments one is dealing with one’s target phenomena.

However, in economics this is not true. Economic theory, according to Guala (1999), is dependent on the institutional background or context in which the theory is supposed to be applicable. Economic theory is specifically constructed for this environment rather than to be tested in experiments. Because of this the argument that one can test a theory in an *economics* experiment becomes dubious, even though it is legitimate in a physics experiment. One cannot hope to have the same context in the laboratory as one has in the external world.

It may be argued that it may still be possible to test an economic theory in the laboratory if one has sufficient causal factors represented in one’s theory for one to be able to test them. In other words, if one was able to completely specify the *ceteris paribus* clauses for the model being tested then such a test would be legitimate. This Guala (2005b, 2005a p. 150-156) refers to as the necessity for “completeness”. However completeness is a very difficult thing to achieve as it requires that in all experimental conditions there should be no unaccounted confounding factors. If such a factor is found then the theory is no longer complete.

This problem grows when one realises that completeness does not just apply to economic factors. In fact, many ceteris paribus clauses/ confounding factors in economic experiments will originate outside the domain of economics. Even worse, economic relations tend to supervene on other more fundamental relationships (for example in psychology). Moreover, the confounding factors which emerge from outside the domain of economics will tend to originate in these lower, more fundamental levels. It follows that completeness requires reduction to those more fundamental laws and an accounting of the possible confounding factors found at this level. However, these fundamental laws in turn supervene on more fundamental laws (say in physiology) and work only with respect to confounding factors and ceteris paribus clauses originating from this more fundamental level. This leads to a chain of reduction all the way down to physics which, according to Guala, is genuinely universal.

Since reductionism is, at least practically, impossible and because the specification of all possible areas of applicability of a theory is not feasible, the requirement of completeness is far too onerous a condition for an economic theory. In fact, according to Guala, this impossibility is acknowledged in economic theories by an implicit ceteris paribus clause that effectively accounts for such confounding factors but does not explicitly state what they are.

For this reason, there is no hard definition of the domain of applicability for a theory and so it is impossible to specify the conditions for an experiment to replicate that of the external world. Instead the link between an experiment and the external world is ensured empirically by the use of an analogy. A comparison has to be made between the elements of the experiment and the elements of a field study to ensure that the experiment is representative of the external world. An experiment therefore cannot be said to be “externally valid” unless it corresponds with some data from the external world.

One implication of this view is that one cannot test external validity by bringing external factors into the laboratory as suggested by Starmer (1999) and Jones (2006). Any test carried out in the laboratory is a test of the robustness of a phenomenon under different conditions. (Guala 1999;2005a p. 228-229). A test of robustness is a generic test of a phenomenon with different conditions. However it cannot be construed as a test of external validity because it does not, and cannot specify in enough detail the conditions for a given target system. To establish external validity one needs to have a further hypothesis of a link between the experiment and the target system which can be confirmed or disconfirmed by an analogy.

The question then arises as to what is the status of experiments in economics. It may seem from Guala's analysis, stripped of its empirical role of testing theories, that the experiment's role has changed. Guala endorses this change because, in his view, experiments are best seen as being similar to models in that they are used for demonstrating the effects of artificial systems. The main difference is that experiments have more concrete elements, such as real people as subjects. In this sense experiments are closer to the external world than are models. Models have a formal similarity with the external world, while experiments' similarity is material as well. Of course, this does not mean that experiments are in any way more externally valid as both models and experiments abstract from the outside world.

Specific Objections

It can be seen that Guala's ideas do have some resemblance to those of Cartwright, particularly in his distrust for the role of theories and also because of his emphasis on *ceteris paribus* conditions and the fact that experiments are heavily shielded from the external world. However, while there are similarities it can be seen that there are substantial differences. Cartwright does not use the distinction between laboratory and non- laboratory sciences. As far as she is concerned there is no distinction between the two and the problems involved in one apply to the other as well.

Also, while Cartwright emphasises the necessity for *ceteris paribus* conditions on all theories and the impossibility of finding a "pure" universal theory, she does not take this to mean that one cannot do a test of the external world using the experimental method. Instead she sees this as evidence for a different metaphysical picture containing capacities and nomological machines rather than universal theories. Experiments in Cartwright's view are simply controlled (or "shielded") versions of the real world in which capacities are allowed to operate in their pure form. The shielding however does not mean that experiments are fundamentally *different* from what is happening in the external world.

Given that Guala's work is not necessarily supported by Cartwright's ideas, do they stand up on their own merits? I would argue not, as Guala relies on a series of distinctions which are debatable. I will argue against his ideas on two levels. First of all, I will give specific reasons for why his general arguments do not stand up to scrutiny and then I will give more general arguments for why his methodology leads to a dead end for economics in general and economic experimentation in particular.

The first problem with Guala's characterisation of economic experiments comes from his conceptual split between laboratory and non-laboratory sciences[2]. It is a crucial part of his conceptual schema because it allows him to split apart economic experiments from natural science experiments and to attribute problems to the former which do not exist in the latter. However it does not seem that this split is as sharp as Guala claims or has the implications he claims.

To see this we can look at a modern "fundamental science" such as cosmology. Cosmologists have devoted much time towards the study of the origins of the universe. In studying these events, the cosmologists have built models, they have used astronomical observations and they have used the results of particle accelerator experiments. This has resulted in a powerful mix of observation and theory to produce the highly sophisticated models in use today. It is not obvious that the cosmologists have used experiments as the sole target of their theory because they are using their theories to describe the universe as well while they obviously think that the results of particle physics experiments can be used in the external world without analogies to field studies. Experiments are being used, together with observation, as equally valid sources of evidence about the universe (see for example Hawley and Holcomb 1998).

A similar story could be told about evolutionary biology, where molecular biology, palaeontology, experimental evolution and ecological genetics combine to provide a widely varying differing variety of sources of evidence for various aspects of evolution. (see for example Stearns and Hoekstra 2000). The evidence produced in evolutionary experiments is not the sole target of theories but it is seen as being immediately relevant to the external world without the need for field evidence.

It seems that the "laboratory science" in Guala's terms, while it may exist, is a rare thing indeed. This poses an awkward problem as it means that Guala's critiques of laboratory work effectively apply to the "fundamental sciences" which he was

[2] It should be emphasised that this split is of Guala's own creation- Hacking (1992), whom he quotes for the definition of laboratory sciences, does not draw any philosophical implications from it.

trying to exclude. Therefore (using Guala's methodology), if an experiment is done in a particle accelerator, the results can only be accepted as externally valid if similar effects are seen outside the accelerator, by observation in the field. This would tie down the scope of scientific research to an unacceptable level as many effects (as Hacking 1983 pointed out) cannot be seen in the external world[3].

Another problem comes with the issue of *ceteris paribus* conditions. We have already seen how it is possible to accept Nancy Cartwright's critique of conventional scientific theories and still maintain that experiments tell us something meaningful about the external world. This is because experiments are simply nomological machines which are constructed in the right way to produce effects without the *ceteris paribus* conditions. They do, however, contain the same capacities as the external world and we can derive knowledge about these capacities from experiments. Guala however maintains that this is not the case and that the need to specify all of the conditions for applicability in effect means that the experiment is cut off from the world. In effect, any nomological machine constructed will always lack enough capacities to be a useful proxy for the external world.

However, this insistence on the necessity for specifying all possible conditions of applicability in order to test a theory (or just a Cartwright- style regularity) is strange. Guala himself (2005a) comments that this is not done by experimental economists and draws the conclusion that it is not done because it is not possible. However an alternative reading is that experimental economists try to control *some* possible confounding factors but not all. Instead of trying to control all possibilities experimentalists, like all scientists, try to control those factors which are *relevant* (Franklin 1986). As Franklin points out, while logically there are a vast number of possible factors there are usually, in practice, only a few which are strong enough for the experimenter to need to take notice.

Indeed, arguably, if we are to take Guala's criterion seriously then the requirement to take account of all possible factors in a situation should be applied to *all* empirical research and not just experiments. Any kind of empirical, statistical relationship should explain as many factors as possible. An example of this can be seen in econometric studies of the labour market, where a whole series of factors may be put into a regression equation as variables as well as the ones which are predicted

[3] Particularly in the case of particle physics since particles found in a particle accelerator can almost never be observed in a "free" state (Hawley and Holcomb 1998).

by a theory. Of course, these are not explicitly stated by the theory being tested and they are put in as a result of empirical knowledge of probable confounding factors from previous studies. However, these econometric studies do not include all possible confounding factors- a selection is always made from the infinite number of possibilities.

A similar attitude is emerging in economic experiments where the use of "experimetrics" (Camerer 2003) is becoming far more widespread. In addition the use of experiments with different treatments is used to test for different possible confounding factors. However, in all these cases there are no attempts to claim that *all* imaginable factors have been included while there is an attempt to include as many likely factors as possible. By contrast, if the effects on the main dependent variable are minimal then they will not be thought useful for the analysis[4].

It is also debatable as to whether causes of confounding factors/ ceteris paribus clauses outside the domain of economics have to be found in lower level sciences. According to Guala (2005a p. 153-156) this results in a chain of justification which goes all the way down to physics. However, this is not found in Fodor (1987), Guala's source, who merely points out that explanations for some confounding factors have to be found in other sciences. While going down a level is the "most familiar" strategy, Fodor also suggests that it could be a science at the same level e.g. sociology. This is unexceptionable and does not lead to any necessity for reductionism if one is attempting to explain *ceteris paribus* clauses/ confounding factors. A desire for completeness or even a desire to include all *relevant* factors, does not rely on reductionism[5].

Also dubious is the claim that economic models are only meaningful in the context of a given institutional background and that, since experiments do not have the correct institutional context, there is always going to be a problem in their external validity. This claim presumes that the simplification involved in creating experiments

[4] One possible reason for the success of this method may be that the economic external world is in fact "modular" (Simon 1969) in that, while there are many "minor" causal linkages of low relevance between entities there are in fact very few major causal linkages. In artificial systems such as economies this occurs because of bounded rationality- humans inability to comprehend large, complex networks. This means that a researcher is relatively safe if she concentrates on the major causes of a phenomenon.

[5] Curiously, in a note (Guala 2005a p. 154) he states "See also Kincaid (1996 Ch. 3) for a reductio ad absurdum of this sort". While Kincaid does discuss the necessity (or otherwise) of reductionism in the social sciences he does not link it up with completeness and *ceteris paribus* conditions.

inevitably means that they are not the same as the target system[6] and, as a result, need an analogy to transmit the findings across to the external world.

However, if experimentalists have included in their experiment those variables that are believed to be the most relevant to the particular situation then there is no reason why an experiment should not be seen as an empirical test in itself. If one follows Cartwright's methodology, then an experiment is simply set up to test a regularity resulting from a nomological machine. If an experiment finds that regularity then it can be said to be directly testing the capacities of entities in the external world. An experiment is simply a "purified" nomological machine but the "shielding" does not make the experimental environment fundamentally different from that of the real world. (See also Hacking 1983)

The fact that there are a large number of factors which differ between an experiment and the external world should not, of itself, prevent an experiment being a vehicle for testing a phenomenon. *Any* empirical test of a phenomenon or a theory will involve data where there are differing factors between contexts. Where Guala recommends a "Field Study" for example (Guala 2005a p.219-221) when he talks about applicability of the Prisoner's Dilemma in experiments to the external world, one can query whether applications to oligopoly or pollution are legitimate given the widely varying circumstances in these two cases. Given that they are assumed to be applications of the Prisoner's Dilemma then it could be argued that field studies on oligopoly or pollution could be used as tests of the Prisoner's Dilemma model.

However, why should an experimental Prisoner's Dilemma be different? Why can an experimental context not be used for such a test? In the former case it may be argued that oligopoly and pollution are modelled as prisoner's dilemmas because the main aspects of the situation resemble a prisoner's dilemma. However, exactly the same response can be made for an experimental version. Many details may vary but as long as the main relevant features are included, other factors need not be considered.

Guala makes a strict division between testing in the laboratory for robustness and testing for the generality of a phenomenon by using analogy to the external world. This follows from his beliefs that external validity can only be assured if there is an analogy to a specific, concrete target system and that experimenters cannot reproduce

[6] This would be inconsistent with Morgan and Morrison's (1999) claim that models can represent the outside world and can (in many cases) be moved closer to the outside world if necessary.

such a system in the laboratory because of the lack of “completeness” in any such test. However, in principle, it would seem that testing for confounding factors would be a prime method of making an experiment externally valid.

This can be seen by realising that the aim of experimentation is to test for causal relations. One is therefore searching for the causes of a behavioural pattern or some other effect. If an experiment does not correspond to the external world then this must be because its (relevant) causal relationships do not correspond. In principle therefore one could carry out a further experiment to ascertain whether this causal relationship holds or one could carry out more “experimetrics” on the current experiment. Since, as we have seen, the supply of relevant factors is usually not infinite, this is not an impossible task It follows therefore that possible confounding factors should be a motive for further tests rather than signs of an insuperable barrier.

A final criticism relates to Guala’s ideas on how external validity can be established, given that one cannot do it experimentally. Guala’s solution is that of analogy between the experimental system and the system in the external world. When a correspondence is established then so is external validity. This is admitted to be fallible but does not answer a question asked earlier: how general is this analogy? If one is testing a prisoner’s dilemma within one particular context in the external world then an analogy may be made between the experiment and this context. However, this is of little use. One needs to then form an analogy between this context and other contexts in the external world which have similar systems in order to establish a general relationship. Each of them will be different from each other and can only be made approximately the same by abstracting away from the context. Given that experimentation involves a similar process of abstraction, this leads to the same problem of why the experiment is the only context which is not seen as a valid testing- ground of the theory.

Even if we ignore this problem, then the proposed analogical relationship is still problematic. It begs the question of how we can tell whether an analogical test has been successful. Guala claims that this occurs when the relevant mechanisms can be mapped onto each other. However, this seems to contradict his claim that completeness is necessary for a test in the laboratory. Surely the “relevant mechanisms” would include the background context as well? Why are the relevant mechanisms sufficient for an analogical comparison with the external world but not for a full- blown test of the model?

General Objections

Guala argues (p. 157 in Guala 2005a, p. 192 in Guala 2005b) that his methodology is not the same as mere instrumentalism (such as that of Friedman 1953), partly because he imposes an inductive constraint whereby any redrawing of the boundaries of the domain of applicability of a theory must have testable implications. However this still leaves the notion of applicability open to several general worries. The primary claim is that, since the applicability conditions cannot be fully specified, one cannot be claiming to test a theory or model at all but can only demonstrate the circumstances under which a model can hold (Guala 2005b).

One problem with this picture is that it seems to have a confused view of the role of *regularities* in scientific experiments. The notion of applicability suggests that an experiment's role is to test for applicability in different circumstances. However, suppose that we do such a test and find that a given model's predictions do not hold i.e. it is not applicable. How far do such a test's results hold? How different do circumstances have to be for the results not to hold? The point is that to extend this negative result beyond the immediate experimental test involves positing another regularity. In this case the regularity is the negative result attached to the model's prediction in the circumstances of the test.

To use an example used by Guala: Chu & Chu (1990) demonstrated that arbitrage in an experiment can result in the preference reversal phenomenon being considerably reduced. This is obviously a counter- example to the generality of the preference reversal phenomenon. However can we apply it to all situations where there is arbitrage? Can we apply it even to identical experiments with arbitrage? In order to do so we would have to posit a regularity: "Preference reversal is reduced in the presence of arbitrage".

However, this regularity in turn would require testing for applicability and the resulting regularities would require testing for applicability and so on *ad infinitum*. Under the applicability methodology one has to accept that a regularity will never be universal and so one becomes trapped in a vicious circle. It may be the case that these new regularities are testable but this does not release one from the vicious circle. If one abandons the idea that theories or models can be rejected by an experiment then one is left changing the domain of the regularity in response to every negative test.

Models increasingly become arbitrary collections of regularities with lists of exceptions. Models (and indeed the regularities created by the tests) lose value even according to the limited role of organising one's ideas. If models are never rejected then there is no incentive to find a better model.

It should be noted that this problem does not apply to Cartwright's (1999) methodology. Cartwright still allows for the testing of regularities within experiments as experiments in her view are heavily shielded "nomological machines" which are set up to generate such regularities. The testing of these regularities in turn indirectly tests the capacities of the components of the system. If the nomological machines fail to produce the regularities specified by the supposed capacities then this gives a signal that the proposed capacities or the design of the nomological machine is incorrectly specified.

The result of Guala's methodology is that models (together with their list of confounding factors) will become more complex while the domain of applicability is drawn with increasing numbers of exceptions. This runs against the most common requirement for scientific modelling- that explanations should be parsimonious. An experimentalist using a methodology of applicability may start with a parsimonious explanation for the data but this will degenerate as more applicability exceptions pile up. Of course one can find out *why* there is such an exception as a low level hypothesis (as Guala suggests) but unless one creates a new, parsimonious model which allows for such exceptions[7] then one is still left with the same problem.

Finally, even Guala's inductive criterion for allowing exceptions falls foul of the applicability methodology he espouses. By requiring that any posited exception to a model must require a testable hypothesis, Guala quite rightly limits the scope for general refutations of specific experiments or of experimentation in general. However, by allowing for the possibility that this hypothesis in turn may be inapplicable, Guala effectively reduces the impact of this proposal. If such a hypothesis need not apply everywhere and any possible situation may be an exception then it need not apply anywhere except in the experiment where it was first proposed. This means, at the limit, one can suggest an exception to the rule without any necessity for this exception to hold in a future experiment. Ad hoc explanations of anomalous experimental results would become perfectly valid.

[7] Of course, if the exception was integrated into a new, parsimonious model then it would no longer be an exception.

Conclusions

From this one could legitimately ask why this matters to experimentalists. Part of the reason is because of its implications for the status of experiments. If Guala is right and experiments are merely types of models with a small dash of realism then this demotes experimentation from being an empirical technique to being a subsidiary type of theorising. Experiments can no longer give final answers to essential questions about how the economy operates but instead have to wait for data generated by systems in the economy. Experiments therefore are stuck as intermediaries in the process of assessing the truth of models.

Even worse for experimentalists, large classes of experiments, those that do not have analogous counterparts in the external world, would have little influence on the content of economic science (Jones 2006 deals with such experiments at length). Such experiments would not act as intermediaries but simply as subsidiaries- waiting for a proper intermediary type experiment to incorporate its insights.

There are also severe problems for experiments if one accepts the idea that the main function of experiments (and indeed all types of empirical work) is to draw the boundary lines of the domain of applicability of a model or theory. The result would be a weakening of the interaction between experiments and theorising. Finding "applicability" is not the same as finding out *why* a theory or model has gone wrong. It does not allow for a search for explanation and, ultimately, a revision of the theory.

This means that one must ask serious questions about the *point* of experimentation under Guala's characterisation. If experiments do not *test* models or theories but simply act as a peculiar type of model then why should one undertake the time and cost of carrying them out in the first place? Guala claims that experiments are worth doing because they include elements- such as real people- which are excluded from mathematical models and simulations. However, this doesn't seem sufficient. According to his methodology this does not (and cannot) make the experiments more externally valid so the use of "real people" doesn't have an impact on that level.

One could argue (as Guala seems to do) that experiments produce phenomena that the modellers have failed to think up. However, this is a critique of modeller's imagination, as it is hard to see why a sufficiently imaginative modeller would not be able to incorporate such phenomena into his models without using experiments. Even if one accepts experiments as idea-sources then there is still no need to do experiments- one could get one's ideas from the vast amount of research done in psychology.

Guala's characterisation of economic experiments therefore would weaken the argument for carrying them out in the first place. However, as this paper has argued, there is no need to accept this characterisation. His arguments relating to external validity, in particular, are critically flawed. It follows that the rejection, made in this paper, of this type of argument restores experimentation as an empirical technique that has the power to force changes in theories (or models). We can also carry out experiments which investigate phenomena independently of theories/models and still be confident that we are saying something meaningful about the external world. It does not mean that we have to accept the metaphysics of theories, the dogma of "crucial experiments" or ignore the problem of external validity. One can admit that these are wrong- headed positions and still reject Guala's version of experimentation.

This is not to say that there are no correct points in Guala's methodology. The closeness of models to experiments has been remarked on by Uskali Mali (2005) amongst others and in their construction there are obvious similarities. His distrust of theory- testing as the primary aim of experiments is, in my view, probably correct. Likewise for his emphasis on the value of surprising phenomena in experiments and his emphasis on the idea that experiments test causal relationships. There are many other ideas, particularly in his book, which are valid comments on experimental economics. This paper simply aims to challenge those parts which do not go through.

Finally, once one eliminates the distinction between experimental and non-experimental sciences as a meaningful philosophical division then it can be seen that there is a general unity amongst experimental techniques. Economics experiments, while difficult and involving many factors, including human intelligence, not tackled in other sciences, are using essentially the same techniques. From this point of view there is no fundamental difference between the experimental sciences and economics can happily take its place amongst them.

Bibliography

Camerer C. (2003) "Behavioral Game Theory: Experiments in Strategic Interaction" Russell Sage Foundation Princeton NJ Princeton University Press.

Cartwright N. (1999) "The Dappled World: A study of the Boundaries of Science" Cambridge University Press Cambridge

Chu Y. P. and Chu R. L. (1990) "The Subsidence of Preference Reversal in Simplified and Marketlike Experimental Settings: A Note" *American Economic Review* vol. 80 p. 902 -11

Fodor J. A. (1987) "Psychosemantics" Cambridge MA MIT Press

Franklin A. (1986) "The Neglect of Experiment" Cambridge; Cambridge University Press

Friedman M. (1953) "The Methodology of Positive Economics" in "Essays in Positive Economics" Chicago University of Chicago Press

Guala F. (2005a) "The Methodology of Experimental Economics" Cambridge University Press, Cambridge.

Guala F. (2005b) "Economics in the lab: Completeness vs Testability" *Journal of Economic Methodology* vol. 12(2) p. 185-196

Guala F. (2003) "Experimental Localism and External Validity" *Philosophy of Science* vol. 70 p. 1195-1205

Guala F. (1999) "The problem of external validity (or "parallelism") in experimental economics" *Social Science Information* vol. 38 (4) p. 555-573

Guala F. (1998) "Experiments as Mediators in the Non-Laboratory Sciences" *Philosophica* vol. 62 p. 57-75

Hacking I (1983) "Representing and Intervening" Cambridge University Press Cambridge

Hacking I. (1992) "The Self- Vindication of the Laboratory Sciences" in A. Pickering (ed.) "Science as Practice and Culture" Chicago University of Chicago Press p. 29-64

Hawley J.F. & Holcomb K.A. (1998) "Foundations of Modern Cosmology" Oxford Oxford University Press

Jones M.K. (2006) "On the Autonomy of Experiments in Economics" Dundee Discussion Papers in Economics no. 194

Kincaid H. (1996) "Philosophical Foundations of the Social Sciences" Cambridge: Cambridge University Press

Mäki U. (2005) “Models are experiments, experiments are models” *Journal of Economic Methodology* vol 12(2) p. 303-315
Morrison M.C. and Morgan M.S. (1999) “Models as Mediators” Cambridge: Cambridge University Press p. 10-37
Plott C.R. (1991) “Will Economics Become an Experimental Science?” *Southern Economic Journal* vol. 57 p. 901-919
Simon H. (1969) *The Sciences of the Artificial*. MIT Press, Cambridge, Mass,
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
Stearns S.C. & Hoekstra R.F. (2000) “Evolution: An introduction” Oxford Oxford University Press