Economics and Philosophy, **27** (2011) 247–271
doi:10.1017/S0266267111000204

EXTERNAL VALIDITY AND LIBRARIES OF PHENOMENA: A CRITIQUE OF GUALA'S METHODOLOGY OF EXPERIMENTAL ECONOMICS

MARTIN K. JONES
University of Dundee, UK
m.k.jones@dundee.ac.uk

Francesco Guala has developed some novel and radical ideas on the problem of external validity, a topic that has not received much attention in the experimental economics literature. In this paper I argue that his views on external validity are not justified and the conclusions which he draws from these views, if widely adopted, could substantially undermine the experimental economics enterprise. In rejecting the justification of these views, the paper reaffirms the importance of experiments in economics.

1. INTRODUCTION

Recently the methodology of experimental economics has become a flourishing subdiscipline within experimental economics, attracting both practicing experimentalists and philosophers to the debate (see the special issue of *Journal of Economic Behaviour and Organisation*; Barkley Rosser and Eckel 2010). This debate has resulted in an in-depth examination of the assumptions and methods used in the area. As experiments become more accepted within economics so 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 many debates on the subject is Francesco Guala. In a series of papers (Guala 1998, 1999, 2003, 2005b) and a book (Guala 2005a), he has put forward a set of wide-ranging philosophical claims on experimental methods in economics. In particular he has proposed some radical ideas on the external validity of experiments. For example he has claimed that:

> Experimental evidence can help only at an intermediate stage of confirmation. It cannot completely bridge the gap between the real world phenomenon and the hypothesis under test. (Guala 2005a: 193–194)

This quote is part of a prominent theme in his work: that experiments cannot be externally validated by experimental means. This is a radical stance to take, particularly in light of the existence of field experiments and the idea of the 'ecological validity' of experiments in psychology. Even more radical is Guala's main justification for the usefulness of experiments in the research process:

> Experimental economists [. . .] help the applied scientist by compiling a library of phenomena: a list of mechanisms, effects and biases that may be relevant in concrete applications. (Guala 2005a: 230)

This is a dramatic come-down from claims made elsewhere (e.g. Smith 1982; Plott 1991; Starmer 1999). Experiments, according to Guala's view, are not used for empirically testing either models or theories but instead act as filters that knock out inapplicable ideas. They act in a similar way to models in that they provide interesting insights for future empirical investigation. Experiments are seen as specific applications of models albeit with large material elements that are used for generating ideas (Guala 2005a: 212–222).

The implications of this point of view are also radical. If experiments cannot be used to directly test for applicability in the external world 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 may even be avoided altogether, 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 ideas about phenomena which may, or may not, be useful in the external world.

The aim of this paper is to criticize this point of view and to demonstrate that it relies on too-sharp distinctions and a false picture of how experimentalists should go about their work. We will look at particular claims that seem to be central to Guala's characterization of experimental economics, particularly in relation to external validity. It will be shown that these claims do not help us to understand the worth of experimental economics and why it is an important innovation in economic method.

The paper is structured in the following way. The next section surveys work in this area by a contemporary philosopher of science, Nancy Cartwright, who agrees with many of Guala's conclusions but disagrees with the idea that experiments cannot be externally validated by experimental means. This acts as a point of comparison for the rest of the paper. The section after that gives an exposition of Guala's ideas relating

to external validity and this is in turn followed by a critique of those ideas in four sections. The paper then concludes with a general discussion.

2. PRELIMINARIES – COMPARABLE WORK

Guala's 2005 book, as well as the papers that preceded it, introduced into the methodology of experimental economics a rich set of ideas from contemporary philosophy of science. Some of these ideas are novel in economic methodology and it is necessary to explain some of them even though this paper does not take issue with them. In doing this it will be found that Guala shares some legitimate concerns with other philosophers of science but these concerns do not lead on to his more radical conclusions.

It would be difficult to survey all of the relevant literature in the philosophy of experimental science so we will focus on just one author who will be a particularly useful comparison. Specifically, we will look at a philosopher of science who has shown an interest in the methodology of economics, namely, Nancy Cartwright. Superficially, her position in *The Dappled World* (Cartwright 1999) seems similar to Guala's but, in fact, it differs in many respects. This similarity between Cartwright and Guala is useful as it allows us to demonstrate that many of Guala's ideas can quite reasonably be held without endorsing his more radical conclusions.

Cartwright argues strongly against the idea that the aim of experimentation is to test 'laws of nature' or universal theories. In her view this is a methodological mistake based on a metaphysical misunderstanding as so-called 'universal' theories are almost never genuinely universal. There are always exceptions to the rule and so scientific theories only hold *ceteris paribus*. Experiments are some of the few occasions that laws can be said to hold since they tend to have many factors deliberately controlled and so are heavily 'shielded' from the external world.

This shielding is important as it allows one to test the network of causes underlying phenomena in the real world. The shielding consists of experimental controls that suppress the actions of certain causes while allowing others to operate. This mimics the *ceteris paribus* clauses in theories and so allows experiments to replicate their operation. Another aspect of shielding is that it allows the creation of phenomena that would not naturally exist. By suppressing certain causes others take on more importance, which makes the resulting phenomena different from those that usually exist.

Cartwright puts forward an alternative metaphysical picture to the picture of universal theories – that of *capacities*. An entity in a theory has the capacity to do something if there is a tendency for an effect to occur in a range of different circumstances. 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 generalize across circumstances, including those in the external world and those in experiments. 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 regularly 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. Generalizations 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 generalized as it is shown to operate across different experimental situations.

If an experiment is a nomological machine then models are *designs* of appropriate types of nomological machine. A nomological machine links various capacities together and a model demonstrates how this is done in the abstract. Models also illustrate how regularities can operate in certain circumstances by specifying the factors that are necessary for the regularity to operate (Cartwright 1999: 58). However, because of the large number of factors involved in economics, regularities tend to be few and far between. In a similar way to physics, some of these regularities can be highlighted in experiments but, in general, economists (just like physicists) tend to work with models and experiments constructed to test these models.

The idea that theory testing is a poor reflection of how science actually operates is reflected in work by Hacking (1983), Kincaid (1996: chapter 3) and Morrison and Morgan (1999). All these authors argue that methodological discussion about theories and theory testing bears little resemblance to what scientists actually do. Scientists primarily build *models* rather than theories and then test these models using experiments or field data. Models also act as mediating instruments since they represent how the world works and allow the modeller to understand it. Models are often constructed *using* theories but they are not purely theory-derived. Modellers borrow from disparate sources, including theories, but also empirical data, and often make arbitrary modelling decisions. Cartwright's insistence on the primacy of models is reflected in the work of these authors.

One can examine an application of Cartwright's framework by looking at a typical economic experiment, in this case an experimental test of decision making by Chris Starmer and Robert Sugden on juxtaposition and event-splitting effects (Starmer and Sugden 1993). This work is interesting in that the authors go to great lengths to illustrate the

theoretical background to the experiment[1] although the focuses of the experiments are the two phenomena in question. Juxtaposition effects occur when the attractiveness of each of two prospects (for given probabilities and prizes) depends on the degree of overlap between the outcomes in which positive prizes are won in the two prospects. The more of an overlap there is, the more subjective weight is given to the prospect with the higher probability of a positive prize. Event splitting effects occur when a subject assigns greater subjective weight to a given outcome if it is subdivided into two separate outcomes rather than one, even if the total objective probability is the same. Starmer and Sugden's experiment was constructed to distinguish between juxtaposition and event-splitting effects. The experiment therefore focused on the *capacity* of aspects of the visual design[2] to influence one's choices between prospects.

The experiment examined two sets of choices, each consisting of two choices. Each choice set was undertaken by one group of subjects. In each choice there were two prospects, where the first prospect (labelled 'R' in the diagrams) had an outcome with a positive prize (£11) greater than the positive prize (£7) of an outcome in the second prospect (labelled 'S' in the diagrams). The first group's choices are illustrated in Figure 1. Their choices were as follows: in the first choice the second prospect, S, had a complete overlap (with an additional 10%) of the probability of its £7 outcome with the £11 outcome in the first prospect, 'R'. The additional 10% probability was portrayed as a separate outcome but adjacent to the main outcome. In the second choice in group 1 the £7 outcome in prospect S and the £11 outcome in prospect 'R' did not overlap although the prospect S still had a 10% larger probability of the outcome with a positive prize of £7 (where the extra 10% was not included as a separate outcome). Choosing 'R' more frequently in the second choice than in the first choice could be attributed to either event-splitting effects or to juxtaposition effects.

In the second group, for prospect 'S' in both choices, the additional 10% of the £7 outcomes were 'split off' into separate outcomes in non-adjacent positions. Otherwise the designs were the same. The second group's choices can be seen in Figure 2. For this group, if there was any systematic difference between the two choices they cannot be attributable to event-splitting effects because both choices have the extra 10% split off in the second prospect. A comparison between the groups therefore allowed one to distinguish between the two effects.

[1] Juxtaposition effects derive from Regret theory (Loomes and Sugden 1982) while Event-splitting effects (ESE) are justified by an appeal to Prospect theory (Kahneman and Tversky 1979) without the editing stage that prevents ESE from happening.

[2] The display used was either a 'strip' display where the prospects are represented as separate strips with events represented as proportions of the area of the strip or a 'matrix'

https://doi.org/10.1017/S0266267111000204 Published online by Cambridge University Press

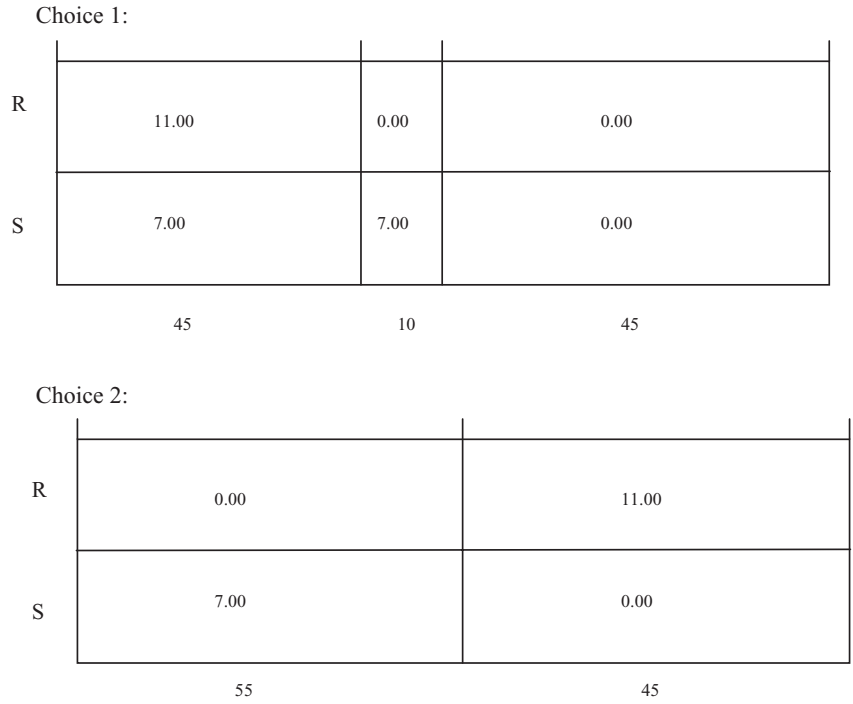

FIGURE 1 Group 1 choices

An examination of this experiment shows how Cartwright's framework fits in to a typical economic experiment. First of all the experiment is heavily shielded. The experimental conditions exclude external effects by holding the experiment in laboratory conditions with no communication between subjects. The prospects are well defined as are the allowable choices between prospects. Contamination between questions was controlled by dividing the subjects randomly into two groups and counter-balancing question orders within those groups.

Secondly, it can be seen how the experimental nomological machine was constructed. It was constructed so that the juxtaposition effect and event-splitting effect could be isolated and differentiated from each other i.e. the main difference between the groups was in the 'splitting off' of the 10% chance of a £7 prize in the second group for both second choices. This isolated the capacity of the diagrams to influence the subjects and produce the desired results. In the external world it would be virtually

display where prospects are illustrated together in one block divided lengthways. Here we focus on the matrix display.

https://doi.org/10.1017/S0266267111000204 Published online by Cambridge University Press

Choice 1

R	11.00	0.00	0.00
S	7.00	0.00	7.00
	45	45	10

Choice 2:

R	0.00	11.00	0.00
S	7.00	0.00	7.00
	45	45	10

FIGURE 2 Group 2 choices

impossible to find a coincidence of events that would allow such a test. Such a situation would very rarely exist and it is the experimental set-up that allows the test to succeed. However, the experiment does say something about the capacity of the framing of choices to influence human decisions in general.

Finally, it can be seen that Starmer and Sugden are testing models rather than theories. The notion that these experiments test models may seem strange in light of the fact that Starmer and Sugden claim to have a theoretical justification for their tests. Both the juxtaposition effect and the event-splitting effect are justified by supposedly general theories: regret theory and simplified prospect theory respectively. However, what is actually tested is something a lot narrower because Starmer and Sugden take account of a variety of *ceteris paribus* conditions, discussed above, that limit the range of choices that are made by their subjects. In creating their experimental design they have in fact created a model of how subjects should behave in an environment where they are autonomous agents, making incentive compatible choices, with clear options and

https://doi.org/10.1017/S0266267111000204 Published online by Cambridge University Press

well-defined probabilities. The testing is confined to the two potential types of behaviour that are allowed within this design.

The methodology outlined in this section deviates significantly from that which is common in experimental economics (see Smith 1994; Starmer 1999) which still revolves around the idea of theory testing. Instead it emphasizes the importance of capacities in building models and in designing experiments. It also emphasizes the role of models (and experiments) in representing the real world, while tackling the problems encountered as a result of *ceteris paribus* clauses. Finally, it casts light on the links between experimentation in the natural sciences and economics because it uses a methodology that applies to both.

3. GUALA'S METHODOLOGY OF EXPERIMENTAL ECONOMICS

Guala's work is primarily associated with experiments in economics rather than other sciences as he believes that the former are qualitatively different from the latter. For this reason, one of the main distinctions he makes is designed to explain why economics experiments are different from those in physics. Guala uses a claim made by Ian Hacking about laboratory sciences: 'those whose claims to truth answer primarily to work done in the laboratory' (Hacking 1992). He then defines a 'Non- laboratory science' as one 'whose claims to truth do *not* answer primarily to work done in the lab and that are aimed at studying phenomena that normally occur spontaneously outside laboratory walls' (Guala 1998, 2005a: 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 idealized circumstances where the entities observed are often unnaturally pure and are often heavily shielded from the external world. Of course, this 'shielding' exists in economics experiments too but in physics it doesn't matter as most theories are concerned with pure entities and unnatural environments. Theories in physics answer primarily to the evidence gathered in experiments and, when one does experiments, one is dealing with one's target phenomena.

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 the relevance of a theory to the external world in an *economic* experiment becomes dubious. 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 for external validity in the laboratory if one completely specified

the *ceteris paribus* clauses for the theory being tested. This would allow a complete specification of the domain of applicability of the theory and would make such an experimental test of external validity legitimate. This Guala (2005a: 150–156; 2005b) refers to as the necessity for 'completeness'. However, completeness is a very difficult thing to achieve as it requires that, in all experiments, there should be no unaccounted confounding factors. If such a factor is found then the experiment is no longer complete.

This problem is magnified when one realizes that completeness does not just apply to economic factors. Many 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). 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 these more fundamental laws and an accounting of the possible confounding factors found at this level. As Guala puts it: 'The neoclassical economist, for example, would have to abandon her models of rational economic agents and engage in a much deeper analysis of human psychology. The laws of psychology being incomplete in character, one would have to move one step further down the ladder of microfoundations to, for example, neurophysiology.' (Guala 2005a: 154). This leads to a chain of reduction all the way down to physics which Guala hypothesizes as being genuinely complete and 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 the absence of such confounding factors but does not explicitly state what they are.

Guala, in common with other authors, acknowledges that experimenters use models in experiments rather than testing theories. However, Guala's ideas about the distinction between theories and models are different from those of Cartwright. While Cartwright sees models as being designs for nomological machines linking together capacities, Guala does not endorse the capacity metaphysics. Models, for Guala, are specifications of theories or, alternatively, theories are sets of models. They manifest themselves as artificial systems constructed for ease of analysis. They tend to be highly simplified, only including small numbers of relevant mechanisms (Guala 2005a: 155–156, 207). Guala, therefore, sees little difference in the problems attached to *ceteris paribus* clauses as these problems apply to models just as much as to theories since the former are simply specifications of the latter.

For these reasons, there is no hard definition of the domain of applicability for economic models and 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 assured 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.[3] As Guala says:

> but establishing that a certain explanation is the right one in the (artificial) domain X does not prove that the same process lies at the origins of a similar pattern of data in the target domain Y. In order to be convinced that this is the case, one needs some further independent evidence from the target domain of application – the real-world phenomenon one is interested in understanding in the first place. (Guala 2005a: 194)

One implication of this view is that one cannot test external validity by bringing external factors into the laboratory as suggested by Starmer (1999)[4] and Jones (2008). Any test carried out in the laboratory that tests the effect of potential confounding factors is a test of the *robustness* of a phenomenon (Guala 1999; 2005a: 228–229). A test of robustness is a generic experimental test of a phenomenon across different conditions. However, it cannot be construed as a test of external validity because, as has been discussed, the experiment cannot include enough factors to be deemed complete. 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 (Guala 2005a: 194–196).

A classic example of what Guala may have in mind is research into the 'Status Quo bias' (Kahneman and Tversky 1979). Status Quo bias is the tendency to see one's current position as a reference point to judge other situations. The result is that gains and losses relative to the status quo are judged asymmetrically. This means that subjects in experiments demonstrate risk aversion for choices involving gains and risk-loving behaviour for choices involving losses. The value function

[3] Guala (2008) in replying to Hausman (2008) confirms this picture. Hausman accuses Guala of having a 'voodoo doll' picture of economic experiments i.e. that the experiments are simply 'stand-ins' for the external world. Hausman argues that, by contrast, experiments involve actual choices with real people and so are not 'stand-ins'. Guala however confirms that the 'voodoo doll' picture is an accurate view of his work.

[4] As will be discussed later on, Starmer's analysis is more subtle than that portrayed here. Starmer sees the experimental method as a *process* whereby different potentially confounding factors are tested in a series of experiments. Not all factors can be tested in one experiment but they can be tested across a series of experiments. (A similar argument is made by Jones 2008.)

is also steeper for losses than it is for gains. This bias has been shown to result in significant differences in contingent valuation experiments between Willingness to Accept (WTA) and Willingness to Pay (WTP), where they should in fact be roughly the same (Knetsch and Sinden 1984).

A study by Hartman *et al*. (1991) moved out of the experimental lab and used a survey by an electricity company on the reliability of electricity supplies to assess whether Status Quo bias existed. In each case the customers were asked for their WTP for improvements or to indicate their WTA in lieu of improvements. The paper analysed the results of the survey and came to the conclusion that Status Quo bias did indeed exist in the external world.

This seems to illustrate Guala's message very well. A theory[5] (prospect theory) is used to construct a model that is tested in an experiment. This experiment is done in a laboratory so a further test is done as a field study to establish external validity. An analogy is made between the lotteries used in the experiments and the variability of electricity supply. Paying for the lottery ticket or being paid money in exchange for a lottery ticket can be seen as an analogy for paying for improvements to electricity supply or accepting money in lieu of such improvements. If the analogy is fairly tight then Knetsch and Sinden's experiment can be seen to be externally validated by that of Hartman *et al*. while Kahneman and Tversky are indirectly externally validated.

The question then arises as to the status of experiments in economics. It may seem from Guala's analysis that the experiment has been stripped of its empirical role of testing models for conformity to the external world. Guala endorses this change because, in his view, experiments are best seen as being similar to models in many respects. This is because both are artificial systems that can be manipulated by the modeller or experimenter. 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 and are more specific than models. Models have a formal similarity with the external world while experiments' similarity is material as well. This does not mean that experiments are made more externally valid by these material elements, however, as both models and experiments abstract from the outside world.

4. CRITICISMS OF GUALA'S VIEWS

4.1 Introduction

It can be seen that Guala's ideas do have some resemblance to those of Cartwright. This can be seen in his distrust for the role of theories

[5] Although prospect theory is actually based on psychological results derived from elsewhere.

https://doi.org/10.1017/S0266267111000204 Published online by Cambridge University Press

and theory-testing 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, there are also substantial differences. Cartwright does not use the distinction between laboratory and non-laboratory sciences. As far as she is concerned there is no distinction and the problems involved in one apply to the other as well.

Cartwright also emphasizes the necessity for *ceteris paribus* conditions attached to theories and the impossibility of finding a 'pure' universal theory. However, she does not think this means that one cannot use experiments to do direct tests of external validity. Instead she sees this as evidence for a different metaphysical picture using capacities, nomological machines and models 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 does not mean that experiments are fundamentally *different* from what is happening in the external world. Guala, by contrast, sees economic experiments as artificial environments where their results cannot be seen as similar to the outside world without additional evidence from that world. Guala therefore sees shielding as a fundamental problem for experiments while Cartwright sees shielding as a process by which genuine causal tendencies can be isolated by experiments.

Given that Guala's work on external validity is not necessarily supported by Cartwright's ideas, despite some similarities, does it stand up on its own merits? I would argue not, as Guala relies on a series of distinctions which are debatable. I will divide the objections into three parts. First will come a series of interrelated arguments about completeness that are central to Guala's position. Then will come two arguments, firstly, relating to Guala's distinction between laboratory and non-laboratory sciences and, secondly, to his use of analogy.

4.2 The problem of completeness

We have already seen how it is possible to accept 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 that exclude confounding factors. However, experiments do contain the same capacities as the external world and we can derive knowledge about these capacities from experiments. Guala, however, does not agree and claims that the need to specify all the conditions for applicability means that the experiment cannot be used for a test of external validity. In effect, any experimental

nomological machine will always lack enough capacities to test for external validity.

However, this insistence on the necessity for specifying *all possible* conditions of applicability in order to test for external validity is strange in the context of modern experimental practice and philosophy of science. 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 that are *plausible*[6] (Franklin 1986).

Logically, there are a large number of possible factors that could conceivably influence a phenomenon. However, it cannot be the case that all of these factors are equally strong. Indeed, if this were the case then science would become impossibly complex. Instead there are some factors that have more influence on a phenomenon than others and it is these that are incorporated into models and empirical tests. Of course, as Cartwright (1999: 56–57) points out, this does not mean that we can include all *possible* cases. However we can, in many circumstances, generalize across the bulk of cases. In itself, the lack of completeness merely suggests that research is necessary to distinguish those factors that do have a significant influence from those that do not.[7]

In their experiment it is noticeable that Starmer and Sugden, in fact, manage to control for a large number of possible influences. For example, it controls for the possibility that subjects obey the independence axiom of expected utility theory. Juxtaposition and event-splitting effects are both effectively demarcated while sampling effects are partially controlled by allocating subjects at random to the two groups. However there are certain elements that are not controlled. To take three trivial examples, one possibility is that large numbers of subjects may have gone to the casino the previous night and played a similar game to that in one of the groups. Alternatively, the size of the room may influence decisions or it may even happen that play is influenced by the flapping of butterfly wings outside the experimental laboratory. None of these possibilities is controlled.

[6] Guala refers to the necessity for completeness as including all 'relevant' factors (p. 153). However, Guala's use of the term is different from my use of 'plausible'. He is referring to all possible factors covered by the *ceteris paribus* conditions. Here we mean those factors that have a significant effect on the phenomenon being explained.

[7] One possible additional 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 strength 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.

The reason for this is that these possible confounding factors are seen as being either unlikely, in the case of the casino, or causally irrelevant in the case of the flapping of butterfly wings or room size. There is no evidence that the flapping of butterfly wings or the size of the room influences such choices. It is unlikely that casinos would have versions of such games or that large numbers of subjects would be playing them. It is also unlikely that any of these things could be an influence in the outside world except in extremely unlikely circumstances. While *logically* the possibility exists that these could be an influence it is merely a logical possibility. There is *no good reason* for thinking of them as significant causal influences.

This notion of plausibility can be related to a far more general methodological point. This is the principle of parsimony (also known as 'Occam's razor'). Guala's insistence on completeness is a claim that we should incorporate all possible confounding factors. However, the principle of parsimony pushes in the opposite direction; we should not be aiming for completeness but for the minimum number of factors that will explain a phenomenon and make an experiment externally valid. Parsimony enforces plausibility. If a possible confounding factor is not necessary for external validity then it should be excluded from the experiment.

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* economic empirical research and not just experiments. *Any* empirical test of a phenomenon or a theory will involve data where there are large differences between contexts. An econometric test of the labour market, for example, may look at a large group of individuals, all of whom have an enormous number of differences between them according to their background and the context in which they work. A regression equation is tested over a given sample, often from one segment of the labour market, and is then extrapolated to the rest of the workers. Applying Guala's methodology, we can see a similar problem emerging. Not all possible factors can be incorporated into the model and the subjects' background and 'context' cannot be guaranteed to be like those outside the sample.

Why might we think that results from one segment of the labour market carry over to another? This is because the econometrician identifies variables that are general across the labour market such as sex, age and education but also controls for specific contextual variables (e.g. manual/ non-manual worker, industry sector etc.) that might confound the initial explanatory variables. Of course, these are often not explicitly stated by the economic model being tested and they are put in as a result of empirical knowledge of probable confounding factors from previous studies. However, these econometric models do not include *all* possible confounding factors; a selection is always made from the infinite number of possibilities.

Guala, therefore, is claiming too much in insisting on completeness. There is no reason why a study of one context, in this case a particular labour market, should have any fewer difficulties in being 'externally validated' than an experiment. In neither case can a researcher aim for completeness by including all possible confounding factors and nor should they. Instead they should (and indeed do) aim to find an externally and internally valid explanation for a phenomenon using the fewest causes necessary.

An attempt *contra* Guala to incorporate potential confounding factors is emerging in economic experiments where the use of 'experimetrics' (Camerer 2003) is becoming increasingly common. However, in all these cases there are no attempts to claim that *all* imaginable factors have been included even though there is an attempt to include as many likely factors as possible. In general, if the effects of a potentially confounding factor on a main dependent variable are thought to be minimal then that factor is not included in the analysis.

An example of the success of this approach is in a paper by Viscusi and Magat (1992) devoted to experimentally investigating ambiguity aversion. This experiment is interesting because Viscusi and Magat used a wider sample of subjects (shoppers in a shopping mall) and a less abstract format (environmental risks leading to disease) than typical experiments in this area. They also gathered some demographic variables on education levels, employment, income, knowledge of the relevant diseases and a variety of other factors. Knowledge of these factors, as well as the broader demographic spread, meant that many more factors were controlled including many that could plausibly influence subjects' attitudes towards environmental risk and ambiguity. These factors were joined together in an econometric model to examine the level of ambiguity aversion.

It may be thought that Viscusi and Magat missed out some factors that could influence the level of ambiguity aversion. This may indeed be the case. However, this is simply a claim that one should find out which factors have been missed out and then include them in another experiment. There is no need to specify all imaginable *ceteris paribus* conditions. Not all factors will be plausible causes and not all should be included in their model.

One possible objection to this approach is that, in many cases, one cannot construct an experiment that includes all plausible confounding factors. By allowing various factors to operate, the experimenter may loosen too many controls in the experiment and may not be able to locate the causes of a phenomenon. In effect, there is a trade-off between increased external validity and decreased internal validity. This is perfectly true as far as one experiment goes. However, as Starmer (1999) (see also Jones 2008) points out, research programmes do not revolve around just one experiment. Instead research programmes consist of a

series of mutually reinforcing experiments, each of which investigates the consequences of incorporating one confounding factor or another into the experimental design. In Viscusi and Magat's experiment it would be possible, for example, to run further experiments using different subject pools (football fans for example) or other frames (the risk of injury during a football game for example).

Another argument used by Guala is the fact that many factors that influence phenomena in economics come from outside the domain of that discipline. However, it is debatable as to whether these confounding factors have to be found in lower level sciences as Guala claims. According to Guala (2005a: 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 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 perfectly reasonable so there is no necessity for reductionism if one is attempting to locate confounding factors. A desire for completeness or even a desire to include all *plausible* factors does not necessarily lead one to endorse reductionism.[8]

Even Guala's argument for the necessity of reductionism to achieve completeness is suspect. A belief that all plausible factors be accounted for in an experiment does not imply that one should have any belief in reductionism at all. The fact that some characteristics in the domain of one discipline supervene on characteristics in another (Guala 2005: 153) is not sufficient. Supervenience simply implies that certain characteristics in the domain of the first discipline cannot vary without certain characteristics in the domain of the second discipline also varying. It does not follow that one has to reduce one to another or that the only way to complete the list of (plausible) *ceteris paribus* conditions is to do such a reduction.[9] Indeed, it is quite plausible to assume supervenience but still believe that explanations in higher-level disciplines such as economics (or indeed sociology) are better than explanations at lower levels.

There are some signs that a debate on reductionism is taking place within the economics discipline. In a recent special issue of *Economics and Philosophy* on neuroeconomics (2008), alongside many voices in favour of the reductionist programme in neuroeconomics (e.g. McCabe 2008), there was some reaction against this viewpoint (Wilcox 2008). Wilcox's point

[8] Curiously, in a note (Guala 2005a: 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.

[9] Indeed the whole point of supervenience is that it is invoked when reduction is seen as impossible or too complicated.

was that neuroeconomics ignores the fact that a large part of cognition is actually distributed among external cognitive artefacts and other agents. Most innovation and growth within economics can be seen to be the result of social learning and the structure of groups of people. Explanation of a phenomenon therefore will not go 'downwards' into neurophysiology but 'across' into social psychology. While neuroeconomics is undoubtedly popular, there is no reason to believe that reductionism is a necessity within economics.

Another claim is that economic models are only meaningful within a given institutional context and that, since experiments do not have the correct institutional context, there is always going to be a problem testing their external validity. This claim presumes that the simplification involved in creating experiments inevitably means that they cannot be identical to the target system. As a result an analogy is needed to transmit findings across to the external world.

However, if experimentalists have included in their experiment those confounding factors that are believed to be the most plausible for a particular situation then there is no reason why an experiment (or a series of Starmer-style experiments) should not be seen as an empirical test in itself. If one follows Cartwright's methodology then an experiment is a nomological machine that is supposed to generate a regularity. If there is a similar 'linking up' of capacities in the external world then it will produce a similar regularity. 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 a 'purified' nomological machine but the 'shielding' does not necessarily make the experimental environment fundamentally different from that of the real world. (See also Hacking 1983.)

Guala makes a strict division between testing in the laboratory for robustness, which he sees as a valid procedure, and testing in the laboratory for external validity, which he sees as invalid. This is because external validity can only be assured if there is an analogy to a specific, concrete target system and experimenters cannot exactly reproduce such a system in the laboratory because of a lack of 'completeness'. Therefore one cannot test for external validity in the laboratory and any such test involving external factors is actually a test of robustness. However, once completeness is no longer an issue and one looks for plausible causes instead, this problem vanishes. It vanishes because a test of robustness (in this sense), if it specifies the correct causal factors, also acts as a test of external validity.

This can be seen by realizing that the aim of experimentation is to search for the causes of a behavioural pattern or some other effect (see Guala 2005a: 71–83 for a discussion on this). If an experiment does not correspond to the external world then this must be because its (plausible) causal relationships do not correspond to the correct external

causal relationships. Hence the causes of an experimental phenomenon are mis-specified. In principle, one could carry out a further experiment to ascertain whether other possible causal relationships hold or one could carry out more 'experimetrics' on the current experiment. Since, as we have seen, the supply of plausible factors is usually not infinite, this is not an impossible task. Such a test would be a genuine test of external validity rather than merely a test of robustness.

4.3 Laboratory and non-laboratory sciences

One significant problem with Guala's characterization of economic experiments comes from his conceptual distinction between laboratory and non-laboratory sciences.[10] This 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 this split is not as sharp or as innocuous to non-economic sciences as Guala 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 cosmologists have built models, used astronomical observations and 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 cosmologists have used experiments as the sole target of their models because they are using their models to describe the universe as well. Also 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 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 biological models but it is seen as being immediately relevant to the external world without the need for field evidence.

It seems that a '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

[10] It should be emphasized 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.

'fundamental sciences' which he was 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 laboratory effects (as Hacking 1983 pointed out) cannot be observed in the external world.[11]

4.4 Analogy

My final criticisms relate to how external validity can be established given that one cannot do it experimentally. Guala's solution is that of an analogy between the experimental system and the target system in the external world. Each of the causal elements of the experiment has to be matched up to those of the target. When a correspondence is established then so is external validity. This is admitted to be a fallible process but does not answer two fundamental questions: first, how general is this process of analogical reasoning in economic research? Is it solely confined to experimental external validation? Second, is analogy genuinely different from experimental external validation in its application?

To answer the first question, take the example, given earlier, of the Status Quo bias. It is obvious, as has been explained, that Knetsch and Sinden's contingent valuation experiments are similar to the real-world analysis by Hartman *et al*. (1991). However, Hartman *et al*.'s survey has its own 'context' in the external world so there is the question of how this survey on the reliability of electricity supplies would relate to other such situations. Another survey on the reliability of electricity supplies would be a different survey, using different subjects at a different time. The subjects may be in a different location or served by a different company with different electricity connections. In order to make any kind of use of the original survey beyond the original subjects one will have to make a match between the conditions of the original survey and of the target population.

However, this can only be done by abstracting away from incidental features of the original survey: the electricity company used, the location of the survey etc. The more generally one wants to apply the results of the survey, the more general the matching and the more context is stripped away. This process of abstraction from a particular context is precisely the type of process involved in experiments when, following Guala, one is making an analogy to the external world. The experiment has its own context but the fundamental causal mechanisms are, in the

[11] 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).

abstract, supposed to match up with similar causal mechanisms in the world outside the experiment.

However, this leaves us with a problem. The external world, under this view, is split up into several different contexts. A field study may cover a different context from another part of the external world that a researcher may be interested in. This could mean that a field study does not strengthen the external validity of experimental results for another part of the field because the external validity of an experiment is impeded by the effects of different contexts on behaviour. If we change context then we still have to overcome the effects of the new context. The original analogy between the experiment and the field study tells us nothing about the other part of the external world without bringing in another analogy.

It is difficult to see how a field study is useful in establishing external validity if it doesn't overcome this problem of context for the external world outside the field study sample. One could argue that it is a matter of degree; experiments are more abstract and field contexts are closer to each other. However, this needs a convincing argument as the notion of contexts being 'close' is very vague. It could also be argued that this is simply an argument for trying to bring the experimental conditions closer to that of the relevant field context rather than going for a completely different field study.

Even if we ignore this problem, then this still leaves us with our second question of how different analogy is from experimentation. To start with: how we can tell whether an analogical test has been successful? Guala claims:

> Remember that external validity inferences are inferences to circumstances we know to be different in some respects from the experimental situation. In order to make such inferences reliably, we must ask (and check) whether the differences between the experimental and the target system can confound the external validity inference or not. (Guala 2005a: 197).

Why is such a mapping between experiment and target in any way sufficient for an external validity test? According to Guala, the causal relations specified in the experiment cannot be complete because there are always unspecified *ceteris paribus* conditions. Since one cannot complete the experimental design sufficiently for an experimental test of external validity then why don't these *ceteris paribus* conditions also eliminate an external validity test via analogy? By assumption, incompleteness means that the inference must be imperfect. To state that one can 'check the differences' ignores the argument about the large number of confounding factors. Why can't one 'check the differences' in an experiment? Guala claims that analogical comparisons are fallible empirical tests and so one cannot expect a complete match. However, experimental tests are also fallible but are not given the leeway to check differences that analogy is

given. Guala seems to give a free ride to his method of analogies to field studies but imposes impossible barriers to experimental tests. In reality the same problems seem to apply to both methods.

5. DISCUSSION AND CONCLUSIONS

My objections to Guala's position can be classified into three types. First of all, there are objections to the arguments against the experimental testing of external validity. Suppose one wanted to test external validity in the manner suggested by Starmer (1999) by bringing in supposed additional causal factors (or excluding others). This is not blocked by Guala's arguments since there is no necessity for completeness. Instead what is required is knowledge of the *plausible* factors influencing phenomena. Rather than trying to establish completeness, it is good scientific practice to minimize the number of factors in a model or experimental design to those that are plausible. The institutional context and the fact that experiments have to be shielded do not prevent one from testing external validity by experimental means as these are all factors that, in principle, one could test. It follows that experiments that introduce plausible external elements are not necessarily just tests of robustness but can also be said to be tests of external validity.

Next, the division into laboratory and non-laboratory sciences cannot be sustained. Most so-called 'laboratory sciences' have results that are supposed to apply to the external world even if they only exist under conditions of shielding in the laboratory. They, strictly, come under the heading of 'non-laboratory sciences'. Guala's ideas, if accepted, would have to apply to all of these non-laboratory sciences and would block some significant scientific results from ever being categorized as externally valid.

Guala's use of analogy as a solution to the external validity problem is also problematic. Analogy requires abstraction and the picking out of plausible characteristics. However, these two requirements are also the same processes that can overcome Guala's objections to the experimental verification of external validity. If, by contrast, we follow Guala and require completeness, we also eliminate both experimental and analogical verification of external validity. Furthermore, there is no reason for analogy to only apply to the relationship between experimental and external contexts and not to two external contexts. If an analogy is required between the field study and another part of the external world then this suggests that the analogy to the field study from the experiment is of dubious relevance. Indeed, there seems to be little reason for requiring an extra layer of empirical work, if the field study does little to enhance an experiment's external validity with respect to another part of the field.

Given this, Guala's ideas on external validity would not seem to bear scrutiny. However, there has been some discussion of the problem of external validity in the literature and it might be legitimately asked whether Guala's ideas find support elsewhere. Nick Bardsley (2005) has also cast doubt on the ability of some experiments to be externally validated. Bardsley focuses on the 'artificiality critique' of experiments and points out that, in certain circumstances, this results in experiments that cannot be externally valid. An example of this is the fact that some social relations simply cannot be introduced into the laboratory. One cannot, for example, do an externally valid experiment on tax avoidance because one cannot persuade subjects that the experimenter is a tax authority. This objection is similar to that of Starmer (1999: 11) who argues that, in practice, there may be problems with the external validity of experiments for similar reasons.

The crucial aspect of these problems is that they are *practical* rather than general logical problems. It is true that some social relations cannot be introduced into the laboratory. However, this is more a result of the technology available to experimenters rather than a feature of experiments in general. Currently we cannot persuade subjects that they are in certain types of social relation in the laboratory. However, this problem is by no means general. Sometimes these social relations are not relevant (as in choice under uncertainty) or they can be approximated in the laboratory (for example in the roles of buyer and seller in a simple experimental market). The fact that some social relations cannot be introduced into an experiment does not mean that all cannot be introduced. Therefore Bardsley's critique, while legitimate, is not supportive of Guala's thesis.

Having explored Guala's arguments on external validity and objections to it, one could legitimately ask why this matters to experimentalists. Partly it is because of its implications for the status of experiments. If Guala is right and experiments are merely types of models with a dash of realism then this demotes experimentation from being an empirical technique to a subsidiary type of theorizing.

Even worse, 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 2008 deals with such experiments at length). Such experiments would not act as intermediaries but simply as subsidiaries – waiting for 'proper' intermediary experiments to incorporate their insights. Starmer and Sugden's experiment on juxtaposition and event-splitting effects are a case in point. There is virtually no straight analogue to these experiments in the external world and so, under Guala's characterization, they are not applicable in the external world.

However, the argument of this paper is that there is no need to accept this. Starmer and Sugden's experiment captures certain characteristics of

human behaviour that are important and interesting from the point of view of individual choice. Furthermore, the causes and effects of this behaviour are genuine, *real* characteristics that carry over into the real world. People may not behave exactly the same as in the experiment but this is because of the heavy shielding involved; other causes are excluded. However, this merely presents a challenge to find out how the excluded causes do influence behaviour as well as the causes identified in the experiment.

Another question concerns the *point* of experimentation under Guala's characterization. If experiments do not provide externally valid tests of models 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, that 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 thus creating a 'library of phenomena'. However, this must be seen above all as a critique of modellers' imagination as it is hard to see why a sufficiently imaginative modeller, without using experiments, would not be able to incorporate such phenomena into his models. One would have to agree with Hausman (2008) in his review of Guala's book that:

> We would be no worse off with respect to our knowledge of how people behave outside the laboratory if we stopped the experiment and instead developed models in which people are irrational or in which they do not care only about their own monetary payoffs.

Guala's characterization 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 characterization. His arguments relating to external validity are flawed. It follows that rejecting these arguments restores experimentation as an empirical, externally validating technique that has the power to force changes in economic ideas.

Finally, once one eliminates the distinction between experimental and non-experimental sciences as a meaningful philosophical division, it can be seen that there is a general unity amongst experimental techniques. Economics experiments are difficult and involve many factors, including human intelligence, not tackled in other sciences. However they are using essentially the same techniques as those employed by experimenters in other disciplines. From this point of view there is no fundamental

difference between the different experimental sciences and economics can happily take its place amongst them.

REFERENCES

Bardsley, N. 2005. Experimental economics and the artificiality of alteration. *Journal of Experimental Methodology* 12: 239–251.

Barkley Rosser, J. and C. Eckel (eds) 2010. JEBO special issue on 'Issues in the Methodology of Experimental Economics. *Journal of Economic Behavior and Organization* 73: 1–132.

Camerer, C. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton. NJ: Russell Sage Foundation Princeton University Press.

Cartwright, N. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Fodor, J.A. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.

Franklin, A. 1986. *The Neglect of Experiment*. Cambridge: Cambridge University Press.

Guala, F. 1998. Experiments as mediators in the non-laboratory sciences. *Philosophica* 62: 57–75.

Guala, F. 1999. The problem of external validity (or 'parallelism') in experimental economics. *Social Science Information* 38: 555–573.

Guala, F. 2003. Experimental localism and external validity. *Philosophy of Science* 70: 1195–1205.

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

Guala, F. 2005b. Economics in the lab: completeness vs testability. *Journal of Economic Methodology* 12: 185–196.

Guala, F. 2008. The experimental philosophy of experimental economics: replies to Alexandrova, Hargreaves Heap, Hausman and Hindriks. *Journal of Economic Methodology* 15: 224–231.

Hacking, I. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.

Hacking, I. 1992. The self-vindication of the laboratory sciences. In *Science as Practice and Culture*, ed. A. Pickering, 29–64. Chicago, IL: University of Chicago Press.

Hartman, R.S., Doane, M.J. and Woo, C.-K. 1991. Consumer rationality and the status quo. *Quarterly Journal of Economics* 106: 141–162.

Hausman, D. 2008. Experimenting on models and on the world. *Journal of Economic Methodology* 15: 209–216.

Hawley, J.F. and Holcomb, K.A. 1998. *Foundations of Modern Cosmology*. Oxford: Oxford University Press.

Jones, M.K. 2008. On the autonomy of experiments in economics. *Journal of Economic Methodology* 15: 391–407.

Kahneman, D. and A. Tversky 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 2: 263–291.

Kincaid, H. 1996. *Philosophical Foundations of the Social Sciences*. Cambridge: Cambridge University Press.

Knetsch, K.L. and J.A. Sinden 1984. Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value. *Quarterly Journal of Economics* 99: 507–521.

Loomes, G. and R. Sugden 1982. Regret theory: an alternative theory of choice under uncertainty. *Economic Journal* 92: 47–62.

McCabe, K.A. 2008. Neuroeconomics and the economic sciences. *Economics and Philosophy* 24: 345–368.

Morrison, M.C. and M.S. Morgan 1999. *Models as Mediators*. Cambridge: Cambridge University Press.

Plott, C.R. 1991. Will economics become an experimental science? *Southern Economic Journal* 57: 901–919.
Simon, H. 1969. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.
Smith, V. 1982. Microeconomic systems as an experimental science. *American Economic Review* 72: 923–955.
Smith, V. 1994. Economics in the laboratory. *Journal of Economic Perspectives* 8: 113–131.
Starmer, C. 1999. Experiments in economics: should we trust the dismal scientists in white coats? *Journal of Economic Methodology* 6: 1–30.
Starmer, C. and R. Sugden 1993. Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty* 6: 235–254.
Stearns, S.C. and R.F. Hoekstra 2000. *Evolution: An Introduction*. Oxford: Oxford University Press.
Viscusi, W.H. and W.A. Magat 1992. Bayesian decisions with ambiguity belief aversion. *Journal of Risk and Uncertainty* 5: 371–387.
Wilcox, N.T. 2008. Against simplicity and cognitive individualism. *Economics and Philosophy* 24: 523–532.