CBESS Discussion Paper 11-08

Implementing theoretical models in the laboratory, and what this can and cannot achieve

by Stefania Sitzia and Robert Sugden*

* School of Economics and Centre for Behavioural and Experimental Social Science University of East Anglia, Norwich NR4 7TJ, United Kingdom
email: s.sitzia@uea.ac.uk; r.sugden@uea.ac.uk

Abstract
We investigate the methodology used in a significant genre of experimental economics, in which experiments are designed to test theoretical models by implementing them in the laboratory. Using two case studies, we argue that such an experiment is a test, not of what the model says about its target domain, but of generic theoretical components used in the model. The properties that make a model interesting as a putative explanation of phenomena in its target domain are not necessarily appropriate for such tests. We consider how this research strategy has been legitimised within the community of experimental economists.

JEL classification codes
B41, C90

Keywords
Experiments, models, methodology

CBESS
University of East Anglia
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/ssf/cbess

Traditionally, economics has investigated its domain of enquiry primarily by means of theoretical models. Most sciences use both models and experiments, but experiments were extremely rare in economics until the 1980s, and not widely used until the 1990s. It is therefore only recently that economists have needed to think about the relationship between models and experiments. Economic methodologists have discussed this relationship in general terms, noting many similarities between the roles of experiments and models as 'mediators' between the investigating scientist and the world (Guala, 1998, 2005: 203–230; Mäki, 2005; Morgan, 2005), but there has been little consideration of how models are actually used *within experimental economics*.[1] In this paper, we investigate the methodology used in a genre of experimental economics, in which experiments are designed to test theoretical models by implementing them in the laboratory. This is a significant genre in two respects. First, it accounts for a large part of current experimental research. Second, some influential insider-written accounts of the methodology of experimental economics treat model-testing as one of the core activities of the sub-discipline and argue that, in experiments of this kind, there is no requirement that the laboratory environment resembles the target domain of the model (Plott, 1982, 1991; Smith, 1982; Croson and Gächter, 2010). We must stress, however, that this is only one of the genres of experimental economics. Our paper is not about experimental economics as a whole.

Given the central role that models have traditionally played in economics, it is perhaps not surprising that experiments are often viewed primarily in relation to models rather than in direct relation to the world that those models represent. In this paper, however, we will ask whether implementing models in experiments is a good investigative strategy. Are such experiments informative and, if so, what are they informative about?

In trying to answer these questions, we will focus on two specific examples of experiments that implement models – an investigation of price dispersion by John Morgan, Henrik Orzen and Martin Sefton (2006), and an investigation of information cascades by Lisa Anderson and Charles Holt (1997). We have deliberately chosen experiments that have been carried out by highly-regarded experimental economists and published in leading journals (the *American Economic Review* and *Games and Economic Behavior* respectively). Further, the models that these experiments implement are generally viewed as significant contributions to economic theory. Anderson et al's experiment implements the model of information cascades developed by Sushil Bikhchandani, David Hirshleifer and Ivo Welch (1992) and published in the *Journal of Political Economy*; Morgan et al's experiment

implements Hal Varian's (1980) model of sales, published in the *American Economic Review*. The latter model has particular interest for methodologists, because Varian has referred to it both in a philosophical discussion of the role of models in economics (Gibbard and Varian, 1978) and in a more practical essay on model-building, addressed to research students (Varian, 1998).

Why might one doubt the value of implementing models in experiments? The source of our concern is the nature of economic models – or at least, of the important subclass to which the models of Varian and Bikhchandani et al. belong. As we explain in Section 1, although such models are precisely specified, and although their creators claim to be offering some kind of explanation of real-world phenomena, these claims are not given any concrete formulation. This makes the whole idea of *testing* a model problematic. In Sections 2 and 3, we summarise the two models that feature in our case studies.

In Sections 4 and 5 we examine how Morgan et al. and Anderson and Holt design their model-testing experiments. In each case, the experimenters create a laboratory environment that closely resembles the model itself. The only important difference between the experiment and the model is that, whereas the model world contains imaginary agents who act according to certain principles of rational choice, the laboratory contains real human beings who are free to act as they wish. The decision problems that the human subjects face are exactly the problems specified by the model. We argue that such an experiment is not, in any useful sense, a test of what the model purports to say about the target domain. Instead, it is a test of those principles of rational choice that the modeller has attributed to the model world. Those principles are not specific to that model; they are generic theoretical components that are used in many economic models across a wide range of applications. In relation to the experiment, then, the role of the model is not to represent particular features of its target domain; it is to specify an abstract laboratory environment in which the relevant components can be tested.

In Section 6, we ask whether this is the most productive way to test generic theoretical components. We argue that the properties that make a model interesting as a putative explanation of phenomena in its target domain are not necessarily appropriate for controlled tests of theoretical components – although models can sometimes serve both purposes effectively.

In Section 7, we consider how the research strategy of implementing models has been legitimised within the community of experimental economists. We show how, in their methodological pronouncements, pioneers of the field emphasised the theory-testing role of experiments and downplayed issues of external validity. Ironically, these arguments were made in defence of a research programme in which experiments *did* relate directly to their target domains; but their acceptance as folk wisdom has legitimated the practice of implementing models. We end, in Section 8, by drawing some practical lessons for experimental economists.

1. Theoretical models in economics

In this paper, we are concerned with laboratory implementations of a particular type of theoretical model that is common in economics. A model of this type describes a self-contained model world, created by the modeller but resembling some aspect of the real world. The mechanisms described by the model induce 'results' in the model world. No definite hypotheses are offered about the relationship between the model and the real world; but the modeller refers to features of the target domain that resemble the results and suggests that in some unspecified way the model explains these.

Various methodological accounts have been offered to explain what such models are, how they connect to the real world, and how they might help in understanding real phenomena. In *instrumentalist* accounts (the most famous of which is that of Milton Friedman, 1953)[2], a model connects with the world only through its results, which are understood as falsifiable empirical hypotheses; the model itself is merely a procedure for generating such hypotheses. In *realist* accounts, such as those of Daniel Hausman (1992), Uskali Mäki (1992) and Nancy Cartwright (1998, 2002), a model isolates a specific mechanism that works in the real world; thus, despite its apparent lack of realism, the model can be presented as a true description of a highly refined form of reality. In *fictionalist* accounts, such as that of Robert Sugden (2000), a model describes a fictional world that is credible or truthlike in something like the way that the events of a realistic novel are; the model connects with the real world by relations of similarity. In *mechanistic* accounts, a model describes a general mechanism which, in principle, could operate in many different settings and so might prove useful in explaining real phenomena, but the modeller does not claim to be representing any actual property of the world, or to be explaining anything in

particular; the truthlikeness of the model world is merely an aid to understanding the model's potential usefulness.[3] In this paper, we will not explore the differences between these accounts. For our purposes, it is sufficient to notice a feature that they all share – the idea that it is only is some indirect sense that a model can be understood as purporting to describe the world. In different ways, these accounts warn us that the concept of 'testing' a model is far from straightforward.

2. Varian's model of sales

Varian's paper begins:

> Economists have belatedly come to recognize that the 'law of one price' is no law at all. Most retail markets are instead characterized by a rather large degree of price dispersion. The challenge to economic theory is to describe how such price dispersion can persist in markets where at least some consumers behave in a rational manner. (1980: 651)

This passage identifies the broad target domain of Varian's model – price dispersion in retail markets. It also identifies the theoretical literature to which Varian intends to contribute. This literature, relatively new in 1980, consists of a family of models of markets with rational but imperfectly informed consumers, in which (contrary to the 'law of one price' of traditional neoclassical economics) there can be equilibria with price dispersion. The final sentence of the quotation suggests that, for Varian, it is fundamental to the explanations provided by economic theory that they assume at least some degree of individual rationality. Hence, the challenge to economics is to find an explanation of price dispersion that is consistent with rationality.

Varian immediately turns to the theoretical literature. He points out that in most existing models of price dispersion, and specifically in Steven Salop and Joseph Stiglitz's (1977) well-known model of 'bargains and ripoffs', price dispersion is 'spatial'. He interprets these models as describing markets in which 'some stores are supposed to *persistently* sell their product at a lower price than other stores' (italics in original). Price dispersion persists because there are uninformed consumers who, prior to visiting specific stores, do not know what prices are posted there. Varian sees this feature of spatial models as a weakness: 'If consumers can learn from experience, this persistence of price dispersion seems rather implausible' (1980: 651). Here Varian seems to be suggesting that in order for a

model to be plausible as an explanation of real phenomena, the model world must meet some standard of credibility or truthlikeness. Because real consumers can learn from experience, a model which works only by assuming they do not is unsatisfactory.

Varian responds to this problem by modelling 'temporal' price dispersion – that is, markets in which each store intentionally varies its price over time. If this variation is random, consumers cannot use experience to predict which store is posting the lowest price at any given time. Varian supports this proposal by using what he and Gibbard call 'casual empiricism' (Gibbard and Varian, 1978):

> One does not have to look far to find the real world analog of such behaviour. It is common to observe retail markets where stores deliberately change their prices over time – that is, where stores have *sales*. (1980: 651)

This passage explains the title of the paper ('A model of sales') and reveals the specific target domain of the model: sales in the sense of temporary price reductions by retailers. The implication is that the model will be in some way informative about real sales.

Informative in what way? In the only further reference to the real world before the modelling begins, Varian says:

> In the model to be described below, firms engage in sales behaviour in an attempt to price discriminate between informed and uninformed costumers. This is of course only one aspect of real world sales behaviour. (1980: 652)

We take him to be saying that he will describe one particular mechanism that operates (or perhaps: that could operate) in the real world, alongside others which will not appear in the model, and that this mechanism is at least a potential explanation of real sales.

The main body of the paper – everything between the introduction and the final 'Summary' section – is concerned with specifying and analysing the model. The model has a large number of consumers, each of whom wants to buy at most one (discrete) unit of some good. Each consumer has the same reservation price r. Consumers are of two types, *informed* and *uninformed*. Informed consumers buy from the store that posts the lowest price (provided this is less than r). Each uninformed consumer picks a store at random and then buys if and only if that store's price is less than r. Stores are profit-maximising firms with identical cost functions; they are in monopolistic competition with one another in a market with free entry and exit. Varian shows that, in the price-setting game played by firms, there

is no pure-strategy Nash equilibrium, but there is a symmetric mixed-strategy Nash equilibrium (MSNE) in which each firm randomises its price.

Varian's modelling is mainly directed towards providing a mathematical characterisation of this equilibrium. There are only two points in this characterisation where he refers to properties of the equilibrium that might be thought to have significant counterparts in real retail markets. The first is where he shows that (under special assumptions about costs) the equilibrium density function of prices is U-shaped. The second is part of his surprisingly brief review of the comparative statics of the equilibrium. He reports that 'the signs are mostly as expected', but notes one 'interesting feature' – that as the absolute number of uninformed consumers increases, the price paid by informed consumers falls (1980: 656–7).

How is this model intended to relate to the real world? All that Varian has to say about this, apart from the passages in the Introduction that we have already discussed, is contained in the final 'Summary'. We quote this in full:

> I have shown how stores may find it in their interest to randomize prices in an attempt to price discriminate between informed and uninformed consumers, and have solved explicitly for the resulting monopolistically competitive equilibrium in randomized pricing strategies. The form of the resulting pricing strategy as given in Figure 2 [i.e. the U-shaped density function] does not seem out of line with commonly observed retailing behavior. Large retailing chains such as Sears and Roebuck and Montgomery Ward sell appliances at their regular price much of the time, but often have sales when the price is reduced by as much as 25 percent. However, we rarely observe them selling an appliance at an intermediate price. Although this casual empiricism can hardly be conclusive, it suggests that the features of the model described here may have some relevance in explaining real world retailing behaviour. (1980: 658)

The first sentence, which summarises the *theoretical* contribution of the paper, is crisp and precise. But when Varian tries to say how the model is informative about the world, he is extraordinarily cautious and vague.

3. Bikhchandani et al.'s model of information cascades

Bkihchandani et al. (1992; henceforth BHW) offer a 'theory of fads, fashion, custom, and cultural change as information cascades'. They begin by reminding the reader of the

pervasive tendency for human behaviour to be characterised by localised conformity. Previous theories have tried to account for localised conformity in various ways, but none of the mechanisms that has been proposed can explain why conformity is *fragile* – that is, subject to sudden changes brought about by apparently small shocks. BHW mention some examples of such changes, including the collapse of Communism in Eastern Europe in the late 1980s. Their model is presented as 'an explanation not only of why people conform but also of why convergence of behavior can be idiosyncratic and fragile' (1992: 993–4).

As with Varian, it is important for BHW that their model is based on rationality assumptions:

> Although the outcome [of an information cascade] may or may not be socially desirable, a reasoning process that takes into account the decisions of others is entirely rational even if individuals place no value on conformity for its own sake. Imitation is, of course, an important social phenomenon, as has been documented by numerous studies in zoology, sociology, and social psychology. Our contribution is to model the dynamics of imitative decision processes as informational cascades. (1992: 995)

In this passage, BHW seem to be claiming that their model describes a mechanism that operates alongside others in the real world, and which contributes to the explanation of real phenomena of conformity. The hint is that an economic model of conformity should be based on assumptions of rational choice, even if non-rational mechanisms are also implicated in real conformity.

In the main part of the paper (Sections II and III), BHW set up and analyse a formal model, beginning with the stripped-down version which we now describe. There is a sequence of individuals I_1, I_2, ..., each of whom decides in turn whether to *adopt* or *reject* some behaviour, having observed the decisions of everyone who preceded her. Adopting has a utility cost of 0.5. The gain from adopting has a utility value of V, the same for all individuals. With equal probability, either $V = 0$ or $V = 1$. Each individual observes a private signal, which takes the value H (high) or L (low). Signals are identically and independently distributed. H is observed with probability $p \in (0.5, 1)$ if $V = 1$ and with probability $1 - p$ if $V = 0$. All of this is common knowledge.

In presenting the initial stripped-down model, BHW do not state any explicit rationality assumptions; they merely describe what individuals do in the model, leaving it to the reader to work out why this behaviour is rational. When they present the general model, they simply state 'We use the concept of perfect Bayesian equilibrium', without giving any

further explanation or justification (1992: 999). Presumably, BHW treat it as self-evident that the agents in an economic model act in accordance with principles of ideal rationality, as defined in conventional decision and game theory.

Suppose that I_1 observes H (the reasoning is symmetrical in the case where L is observed). So I_1 adopts. If I_2's signal is H, he adopts; if it is L, he is indifferent between adopting and rejecting and so (BHW assume), he adopts with probability 0.5. If I_1 and I_2 have adopted, it is rational for I_3 to adopt irrespective of her signal, and similarly for everyone who follows: this is an information cascade. If I_2 rejects, I_3's position is essentially equivalent to I_1's. Thus, an information cascade (either for adoption or rejection) is likely to form very quickly. Because cascades can be precipitated by very few signals, 'incorrect' cascades (universal adoption when $V = 0$, or universal rejection when $V = 1$) are quite probable if individual signals are noisy. However, precisely because a cascade has so little information content, it can easily break up if new public information is released, or if some individuals' signals are more informative than others'. In this sense, cascades are fragile.

In marked contrast to Varian, BHW devote a considerable amount of space to the discussion of real-world 'illustrative examples', drawn from such diverse areas as politics, medical practice, finance and zoology. As economic modellers, BHW are unusually (and admirably) explicit about how they have selected their illustrations. Their examples satisfy two criteria. First, they are consistent with four 'model assumptions' – actions are sequential, individuals have both private information and information about previous decisions, there is no verbal communication, and there are no sanctions or externalities to enforce conformity. Second, they are consistent with three 'model implications' – conformity is local or idiosyncratic, conformity is fragile, and some individuals ignore their private information (1992: 1009–10). BHW do not claim to be *testing* their theory, but the suggestion seems to be that a good test would look for real-world cases in which the four model assumptions were satisfied, and then investigate whether the three model implications were confirmed. The fact that their examples satisfy this test provides informal support for their hypothesis that the mechanism described by the model operates in the world.

Notice that the Bayesian rationality of individuals is *not* included in the list of model assumptions. Correspondingly, when discussing examples concerning the behaviour of voters in US primaries, doctors making decisions about tonsillectomies, investors deciding whether to subscribe to share issues, or sage grouse females choosing between males, BHW offer no evidence that the relevant decision-makers are Bayesian expected utility maximisers.

The implication seems to be that, for BHW, rationality is an explanatory principle, not a domain restriction. By this we mean that (for example) BHW's claim to explain the behaviour of US voters is not prefaced by '*If* voters were rational, *then* ...'. Rather, the claim is that, in relevant respects, the behaviour of US voters *is* similar to that of the rational agents of the model.

4. Morgan et al.'s implementation of Varian's model

Morgan et al. (2006; henceforth MOS) present their experiment as a test of a model of price dispersion that is 'closely related' to Varian's model of sales (p. 137).

The first substantive part of MOS's paper is an exercise in modelling. MOS revise Varian's model by treating the number of firms as exogenous, rather than assuming free entry and exit. They derive two new comparative static results for this *clearinghouse model*. The first ('Proposition 2') is that as the proportion of informed consumers increases, with the number of firms and the number of consumers held constant, the expected prices paid by both informed and uninformed consumers decrease. The second ('Proposition 3') is that, as the number of firms increases, with the number of consumers and the proportion of informed consumers held constant, the expected price paid by informed consumers decreases and *the expected price paid by uninformed consumers increases*. MOS are particularly interested in the italicised part of Proposition 3, which they describe as a 'counterintuitive prediction' (p. 153). From a theoretical point of view, this result is interesting because it has two properties in combination: it is surprising (economists normally expect an increase in the number of firms in a market to lead to a reduction in prices), and it can be derived by standard theoretical reasoning from assumptions which appear to describe a credible if highly stylised economic scenario. However, MOS are no more concrete than Varian in explaining how the comparative-static properties of the model relate to the real world of retail pricing.

Although MOS's development of Varian's model is interesting in its own right, the main emphasis of their paper is on a report of an experimental investigation of the clearinghouse model. As experimentalists, their objective is to 'examine the empirical relevance of [the] comparative static implications [of the model], as well as the fundamental prediction of equilibrium price dispersion, in a controlled laboratory setting' (p. 135).

MOS seem to be signalling that their primary concern is with what (on their interpretation) the model predicts about behaviour *in the laboratory*. Nevertheless, they

present their experimental findings as informative about real retail markets. Summing up the contribution of the paper, they claim to have found 'strong support for the ability of clearinghouse models to predict the comparative static effects of changes in market structures'. Noting that a 'competitive' market is often defined as one with many firms, they say: 'Our results show that increased competition in this sense does not necessarily result in lower prices' (pp. 153–4). The suggestion in these passages is that the clearinghouse model's claim to be informative about the world is strengthened if its results are confirmed in the laboratory. *In this sense* the experiment is informative about the world. But the experiment itself is a test of the model, not of what the model says about the world.

MOS explain the purpose of this test by expressing three possible 'doubts' about the applicability of the comparative static results of the model:

> There are several reasons why one might doubt the empirical validity of these predictions [i.e. Propositions 2 and 3]. First, in a mixed strategy equilibrium, there is no positive reason for a rational player to conform to the equilibrium strategy since she will receive the same expected payoff from any pure strategy within the support of the equilibrium distribution. Second, the equilibrium price distribution is difficult to compute, and so it seems unlikely that subjects will reason their way to an equilibrium. Third, for the parameters we employ in our experiment, the equilibrium is unstable under the class of positive definite adjustment dynamics, and so it is unclear whether subjects could reach the equilibrium through some learning process. All of these factors suggest that an experiment will provide a stern test for the theory. (p.139)

Notice that MOS's doubts are concerned with the role of MSNE in the model. Surprisingly, these doubts are not expressed in terms of the applicability of MSNE to the model's target domain, pricing decisions by retail firms. The doubts are about whether *experimental subjects* will act according to MSNE when placed in a laboratory environment that reproduces the main features *of the model*.

Consistently with the aim of investigating whether these doubts are well-founded, MOS's experimental design implements the model almost completely. The main difference is that an individual human subject (an undergraduate student) was substituted for each of the profit-maximising firms (or *sellers*) of the model; the mechanism by which subjects were paid gave the subject an incentive to maximise the profits of the firm she represented, profits being determined as in the model. There was no corresponding substitution of human

subjects for model consumers (or *buyers*); the behaviour of consumers was implemented in the laboratory by a computer program which exactly replicated the model.

The experiment had two treatments, one with two sellers in each market and one with four, thus allowing an investigation of the effect of a change in the number of sellers. In each treatment, a session involved twelve subjects making decisions in ninety periods. In each period, subjects were randomly and anonymously rematched into groups of two or four sellers facing six computer-simulated buyers. The experiment consisted of three phases of thirty periods each. In the first and third phases, three of the six buyers were 'informed' consumers in the sense of the model, and three were 'uninformed'. In the second phase, five were informed and one was uninformed. This difference between phases allowed an investigation of the effect of a change in the proportion of informed consumers. In each period, each seller chose a price from the set {0, 1, ..., 100}. These decisions were then processed according to the model, to generate each seller's quantity sold and profit. At the end of each period, each subject was told the prices that had been chosen by, and the resulting sales by, all sellers in the experiment; each seller was also told the profit she had earned.

MOS explain the need for thirty repetitions of each pricing problem as a response to the 'complexity of calculating the equilibrium distribution' (p. 142). In other words, they interpret the MSNE hypothesis as referring, not to play in one-shot games, but to the end state of a process in which a game is played repeatedly. On this interpretation, a fair test of MSNE requires adequate opportunities for experiential learning.

The procedure of random and anonymous rematching of subjects is explained as a means of eliminating 'unintended repeated game effects', such as tacit collusion among sellers (pp. 142–3). This argument illustrates how tightly the laboratory environment is being configured to match the model. In a test of MSNE, repeated game effects are indeed a source of contamination; and MSNE is a property of Varian's model. But in the target domain of retail trade, the same firms interact repeatedly in the same markets, with opportunities for tacit collusion. Similarly, MOS argue that one of the advantages of using computer-simulated rather than human buyers is that this makes buyer behaviour 'controlled and known to sellers', thus reducing the strategic uncertainty faced by sellers. That buyer behaviour is known to sellers is one of the simplifying assumptions of the model; it is not something that is obviously true of the target domain.

If one takes the viewpoint of the subjects themselves, there seems to be very little resemblance between the decision problems they face and those by which retail firms set their prices. The connection between the two is given by the model: the subjects' decision problems are like those of the firms *in the model*, and the firms in the model are supposed to represent firms in the world. But the subjects are simply playing a sequence of ninety simultaneous-move two-player or four-player games with 101 strategies per player and monetary payoffs. The only connection with retail sales is that the players are *called* 'sellers' and 'competitors' and their strategies are *called* 'prices'. Interestingly, MOS see even this link with the target domain as optional. They explain that, at the design stage, they considered following the 'standard' procedure in experimental economics of using 'abstract and context free terminology'; they chose to use the language of sellers and prices to reduce the 'complexity of the experimental setting' (p. 143). In other words, it was important to use language that helped subjects to understand the game *in its own terms*; but provided the game replicated the model, there was no need for other similarities between the experiment and the target domain.

MOS investigate whether the comparative static implications of the model – including the 'counterintuitive' Proposition 3 – are found in the experiment, and (in broad terms) conclude that they are. But what are they testing here?

Clearly, if an experiment implemented a model in its entirety, all that it could test would be the *mathematical* validity of the model's results. Provided one were confident in the modeller's mathematics, experimental testing would be pointless. Thus, when an experiment implements *almost* every feature of a model, all it can test in addition to mathematical validity are those features that have *not* been implemented. In the present case, MOS's experiment effectively implements every component of the model apart from its assumptions about how 'sellers' choose between strategies in the 101-strategy games. The model assumes that these strategy choices lead to MSNE. Thus, *the experiment is a test of MSNE in a specific class of games*. As viewed by the modeller, the specifications of the payoff matrices for these games represent pricing decisions in retail markets; but in relation to the test of MSNE, that representation plays no role.

The counter-intuitive nature of Proposition 3 is a property of this modelling representation. What is counter-intuitive is that an increase in the number of firms in a market increases the expected price for some category of consumers, not that the particular implication of MSNE described by Proposition 3 holds in the laboratory game. The fact that

a theoretical result is counter-intuitive does not necessarily make it a suitable candidate for empirical testing. As Schelling (2006: 147–151) has pointed out, even tautologies – one of his examples is the proposition that for every purchase there is a sale – can have economic implications that are deeply counter-intuitive. Such implications can be significant theoretical contributions even though, once they have been discovered, empirical testing would be pointless. So it would be a mistake to conclude that, just because the clearinghouse model has counter-intuitive results, MOS's experiment is a severe test of MSNE in the Popperian sense.

MSNE is what we will call a *generic component* of economic models – a piece of ready-to-use theory which economists insert into models with disparate target domains. Recall that MOS motivate their experiment by referring to three doubts about MSNE. These are specific neither to the model that they are implementing nor to its target domain. The first doubt, that there is no reason for any player to conform to MSNE even if she knows that other players will conform, applies to *every* application of MSNE. The second doubt, that MSNE is difficult to compute, applies to any even moderately complex application. The third doubt, that MSNE is unstable in the game implemented in the experiment, may seem more model-specific; but notice that this instability is described as a property of the particular parameters that MOS have used to configure the laboratory environment, and not of the model in general. We conclude that the experiment is best understood, not as a test of Varian's model as an explanation of retail markets, but of MSNE *as a generic component of economic models*.

5. Anderson and Holt's implementation of Bikhchandani et al.'s model

Anderson and Holt (1997; henceforth AH) report 'a cascade experiment that is based on a specific parametric model taken from [BHW]' (p. 848). They begin their paper with an account of BHW's model and of its potential relevance for explaining real-world instances of conformity. Then, as a motivation for their experiment, they present three 'reasons to doubt that cascades develop in this way'.

Two of these doubts are about the role of perfect Bayesian equilibrium in BHW's model. In an argument that parallels MOS's doubts about MSNE in Varian's model, AH question whether real decision-makers act with full Bayesian rationality. They point to experimental evidence that individuals often fail to make rational Bayesian inferences. They

also note that the agents in BHW's model make inferences not only about impersonal events, but also about one another's rationality; the suggestion is that the latter are more demanding.

AH's third doubt is very different in nature. Apparently referring to some of BHW's illustrative examples, they say:

> much of the evidence offered in support of the rational view of cascades consists of anecdotes about patterns in fashion, papers getting rejected by a sequence of journals, the risk of entering the academic job market too early, etc. Laboratory experiments can provide more decisive evidence on the validity of the rational view of cascades. (p. 848)

In what we take to be a fleshing-out of the idea that experiments can provide more decisive evidence than BHW's examples, AH discuss some alternative explanations of cascades. One possibility is that individuals' preferences are biased towards whatever is perceived as the status quo. If later actors in an adoption game treat earlier decisions as establishing a status quo, status quo bias could induce non-rational cascades. Another possibility is that individuals simply 'derive utility from herding together'. AH argue that laboratory experiments can control for these alternative explanatory mechanisms, allowing sharper tests of the Bayesian explanation of cascades. By controlling information flows, it is possible 'to determine whether subjects tend to follow previous decision(s) only when it is rational'. By maintaining subject anonymity, interpersonal factors such as preferences for conformity per se can be minimised (p. 848).

One might question whether what AH propose to do is comparable with BHW's use of their illustrative examples. As we explained in Section 3, BHW draw a clear distinction between their model and its target domain. Their formal analysis is of the model, but the *theory* they propose is a theory of real-world phenomena – fads, fashions, customs and cultural change. Their illustrations are real phenomena, presented as evidence in support of what the theory says about the real world. In contrast, AH seem to be concerned with testing the model itself.

AH's experiment is an implementation of BHW's stripped-down model. In each session, a group of six subjects interacted for fifteen periods. At the beginning of each period, a monitor rolled a die to pick one of two 'urns' (in fact, envelopes), each containing three marbles. In urn A, two marbles were light-coloured and one was dark-coloured. In urn B, two were dark and one was light. Subjects knew only that the urn used in the experiment was equally likely to be A or B, and that its contents had been placed in an opaque container.

Then, in random order, each subject in turn *privately* drew one marble from the container, observed its colour, replaced it, and *publicly* reported the letter of 'the urn they think is most likely to have been used'. Each subject was paid \$2 if and only if he reported the letter of the urn that had in fact been used.

The two urns correspond with the events $V = 0$ and $V = 1$ in BHW's model. The colour of the marble drawn by a subject corresponds with a private signal in that model, a light-coloured marble being an indicator of urn A in the same way that the H signal is an indicator of $V = 1$. The compositions of the urns implies $p = 2/3$ in BHW's notation. Dollar payments in the experiment are positively and linearly related to utilities in the model.

The main difference between the experiment and the model is that the Bayes-rational agents of the model are replaced by undergraduate students. But there is also a significant difference of framing. BHW's model is described in terms of individuals adopting or rejecting modes of behaviour which have costs and benefits – a description that corresponds with the target domain of social conformity. AH's experiment uses a framing that is taken from statistical theory. Subjects confront random processes, defined in terms of marbles and urns, and express beliefs about those processes. From the viewpoint of the subjects, there is little obvious resemblance between the statistical problems they are asked to solve and, say, the situation faced by voters deciding how to vote in US primaries (to say nothing of sage grouse females choosing between males). It seems that, for AH, this lack of resemblance is a deliberate and desirable measure of experimental control. Recall their argument that their design minimises the effects of preferences for herding. One of the ways in which this is achieved is by removing any suggestion that the subjects are choosing whether or not to engage in some common behaviour. The connection between the experiment and the target domain is *through the model*: the model uses concepts in statistical theory to represent the target domain, and statistical theory provides the structure for the experiment.

AH's experiment effectively implements every formal component of BHW's model apart from its assumptions about the Bayesian rationality of individuals. Thus, *the experiment is a test of Bayesian rationality in a specific game*. As viewed by the modeller, that game represents social environments in which conformity might be observed, but in relation to the test of Bayesian rationality, that representation plays no role. Bayesian rationality, like the MSNE tested in MOS's experiment, is a generic component of economic models. In motivating their experiment, AH express doubts about the use of Bayesian

rationality in BHW's model, but those doubts are generic too: they apply to any model in which individuals make Bayesian inferences about one another's rationality.

The results of AH's main treatment generally confirm the hypothesis of Bayesian rationality.[4] Almost all the decisions made by subjects were consistent *either* with the prescriptions of Bayesian rationality *or* with the relevant subject's private information. In 41 of the 56 cases in which these two decision principles conflicted, the Bayesian prescription was followed. In other words, in answer to the main question that motivated the experiment: subjects tended to follow previous decisions only when it was rational to do so.

6. Testing generic modelling components

Our interpretation of MOS's and AH's model-implementing experiments raise two obvious methodological questions. Is it informative at all to run experimental tests of theoretical principles such as MSNE and Bayesian rationality, viewed as generic components of economic models? And if so, what makes a particular model a suitable or unsuitable vehicle for such a test?

A strict instrumentalist (taking a position that is often attributed to Friedman) might answer 'No' to the first question, on the grounds that tests should be directed only at the predictions of theories and not at their assumptions. But models like Varian's and BHW's, unlike the neoclassical price theory that was Friedman's main point of reference, do not generate well-defined predictions that can be subjected to straightforward tests. In claiming to explain real-world phenomena, the builders of these models are relying on the supposed credibility of the generic theoretical components they are using. We take it as uncontroversial that tests of these components have some bearing on the validity of the corresponding models.

To explain what we mean by this, let us suppose (counterfactually) that behaviour in MOS's experiment was inconsistent with the comparative-static implications of the clearinghouse model. That observation would imply that the MSNE hypothesis had failed when applied to a well-defined game implemented with real payoffs under controlled conditions. Since MSNE is a general hypothesis in the theory of games, there is a prima facie reason to expect that if it is to hold anywhere, it should hold for games of this kind. This is not to say that there *cannot* be a reasonable argument that particular laboratory games lie outside the domain of the theory, but only that, in the absence of such an argument, any

laboratory disconfirmation counts against MSNE in general (and any confirmation counts in its favour).[5] When a specific model uses MSNE as an off-the-peg component, as the clearinghouse model does, it is drawing on the general credibility of that component. Accordingly, any evidence that counts against MSNE in general also counts against that model – unless there are specific reasons to expect MSNE to work in the model's target domain, despite its disconfirmation in the laboratory.[6]

We now turn to the second of our methodological questions. Granted that it makes sense to test generic modelling components, what makes a particular model a suitable vehicle for such a test?

We have already explained why an experiment which tests a generic component by implementing a specific model is not thereby specifically informative about the target domain of that model. Thus, we submit, such an experimental design should not be appraised in terms of what the model purports to say about its target domain. It should be appraised in terms of what it can tell us about the relevant generic component, *considered generically*. When (as in the cases of MSNE and Bayesian rationality) the same theoretical component appears in many different models, an experimenter can afford to be selective in looking for a suitable design for a test. One might expect there to be many economic models which, however interesting, distinctive and counter-intuitive they might be as explanations of real-world phenomena, are not particularly suitable for experimental implementation as tests of the generic components they use.

We suggest that the clearinghouse model implemented by MOS is such a case. Considered simply as a test of MSNE, MOS's experiment uses extraordinarily complicated games. Many of the canonical experiments in game theory use 2×2 games. Depending on the treatment, MOS's games are either 101×101 (for two players) or 101×101×101×101 (for four players). Payoffs to combinations of strategies are determined by a formula which, although perhaps intuitive to an economist (it replicates the demand conditions of the clearinghouse model), might not be easy for a typical subject to grasp. The arithmetic calculations implied by this formula have to be done by the subjects themselves. The hypotheses that are tested are comparative-static implications of MSNE concerning changes in the payoff formula, or comparisons between two- and four-player games. Leaving aside the interpretation of these games as models of retail markets, it is hard to imagine that any experimenter would choose them as vehicles for testing hypotheses about MSNE.

If one considers the specific questions that provided the motivation for the experiment, the limitations of these games as tests of MSNE become even more obvious. Recall that MOS present their experiment as a response to three 'doubts' about MSNE. Since these doubts refer to distinct and orthogonal causal mechanisms, it would surely be an advantage to use a design (or designs) that could investigate these mechanisms independently. MOS's first doubt is about whether fully rational players would choose MSNE strategies. Translating this into empirical terms, one might ask whether MSNE strategies are played by human subjects who fully understand the relevant game. In investigating this, it would be natural to use games that were particularly easy to understand (perhaps variants of Matching Pennies, with story lines about penalty kicks in football or serves in tennis) – and certainly not MOS's games. The second doubt is about whether, even if players *want* to play MSNE strategies, they can compute them in complex games. To investigate this question, one needs to compare games in which the problem of calculating MSNE varies in difficulty, but the concept of MSNE itself is intuitively easy to understand: an experiment in which all games are extremely complex is not particularly helpful. The third doubt is framed in terms of a specific dynamic theory of learning which implies that some MSNE will be reached after repeated play and others will not. This hypothesis might be tested most efficiently by comparing the evolution of behaviour in repeated play of simple games with different payoff structures.

MOS are of course right to say that, by virtue of the separate plausibility of each of these doubts, the experiment is a 'stern test' of MSNE. But that is not to say that it is well-designed to be *informative about* MSNE. MOS's main findings support the MSNE hypothesis, but they do not do much to help us understand why it has worked in this particular environment. More importantly, had Propositions 2 and 3 been *dis*confirmed, we would have learned very little about why MSNE had *not* worked. The failure might have been caused by any of the three mechanisms, or by some entirely different mechanism; the data would not discriminate between these alternative explanations.

We conclude that the clearinghouse model is not a suitable vehicle for testing MSNE. What about BHW's model as a vehicle for testing Bayesian rationality?

In this case, there is quite a lot to be said in favour of the model. If the aim is to test whether people make Bayes-rational inferences about one another's rationality, the game played by AH's subjects is about as simple as it could possibly be. In this game, 'nature' has only two possible moves (the two urn compositions) – the minimum necessary for there to be

a problem of inference. The relationship between events and signals (the draw of one marble from a set of three marbles in two colours) may not be immediately transparent to all subjects, but it is hard to see how it could have been much simpler. Each subject chooses between only two alternative responses (judging A or B to be more likely), reducing to the minimum the information about earlier decisions that each subject has to process. The cognitive demands of Bayesian rationality increase with the number of previous decisions, but the design generates data about decisions made at each step in the sequence, allowing the effects of increasing complexity to be investigated. Because the players act sequentially with full information about previous decisions, Bayesian rationality has sharply-defined implications which do not depend on assumptions about equilibrium.[7]

The basic structure of this sequential game provides a versatile framework for testing hypotheses about rational and non-rational inference, as can be seen from its use in a range of later experiments (Weizsacker [2008] presents a meta-analysis of thirteen such studies). Its versatility is also illustrated by AH's subsidiary *asymmetric* treatment which tests for the potentially confounding effect that subjects use the *counting heuristic*. This heuristic simply counts the number of previous decisions in favour of each urn, adds the individual's own signal, and then goes with the majority. With the parameters used in the main treatment, the counting heuristic has the same implications as Bayesian rationality. In the asymmetric treatment, the composition of the urns is changed so that the two decision rules sometimes point in different directions. In these cases, AH find an exactly equal split between the corresponding decisions, suggesting that the two modes of reasoning are about equally common.[8] Notice, however, that the asymmetric treatment marks a further move away from the target domain of BHW's model. Oddly, AH say nothing to connect this treatment either with that model or with the possible doubts about Bayesian rationality as the explanation of cascades. It is as if AH themselves are unsure how their experiment relates to BHW's model – unsure, that is, whether they are testing an explanation of cascades or testing hypotheses about Bayesian inference.

Our two case studies illustrate how model-implementing experiments can be more or less effective as tests of generic modelling components. However, the point we wish to stress is that the effectiveness of an experimental design in this respect is orthogonal to the success of the corresponding model in explaining phenomena in its target domain.

7. Model-implementing experiments and the folk methodology of experimental economics

Within the experimental economics community, testing models by implementing them in laboratory experiments is widely seen as a worthwhile research strategy. Our case studies illustrate this generalisation. Evidence of a different kind can be found in insider-written accounts of the methodology of experimental economics. A recent example is a paper on 'the science of experimental economics' by two prominent practitioners, Rachel Croson and Simon Gächter (2010).

Croson and Gächter focus on the relationship between experiments and theories. As a 'mental model' of how experiments interact with theory, they offer the following schema:

Theory ---------- Experiment (Lab/Field) ---------- Observational data

They say that this schema displays two purposes of experiments. One purpose, and the one on which they focus, is 'to address theories'. (The other is 'to examine regularities from the field ... in a controlled, abstracted setting'.) Among the ways in which experiments can address theories is by 'test[ing] predictions':

> Theories (models) are, by definition, simplifications of the world. The goal of a theory is to identify and isolate a phenomenon in order to understand its impacts. Ideally, theories yield unique and testable predictions. Economic theories are logical systems whose truth derives logically from the assumptions. Experiments test whether observed behavior corresponds to the predictions of a particular model. (p. 125)

It is not clear what Croson and Gächter mean by 'prediction'. Following the conventions of economics, they treat 'theory' and 'model' as synonyms. In the first two sentences, they seem to be using a realist interpretation of models as descriptions of features of the real world, represented in isolation. In the fourth sentence they interpret models as logical systems, which can say nothing about the empirical world. Applied to modelling exercises like those of Varian and BHW, in which almost nothing is said explicitly about how the model relates to the real world, either interpretation might be defended, but neither allows an obvious explanation of how a model can generate testable predictions.

At the end of their discussion of how experiments test models, Croson and Gächter refer back to their mental model:

> We [reiterate] a point made by Plott (1982) in discussing the role of the experimental lab as a midpoint between the theory and the field. The (well-designed) laboratory experiment gives a theory its 'best-shot' at making accurate predictions. The assumptions of the theory are designed into the lab experiment. For example, if an auction theory assumes that signals are independently drawn from a known and stationary distribution, the lab experiment addressing that theory will involve signals independently drawn from a known and stationary distribution. ...
>
> [E]xperiments are an *existence proof*; for some set of individuals, for some sets of institutions, with some set of parameters, the theory's predictions are observed. ... [I]f under these best-shot conditions the theory's predictions are not observed, this is a strong statement indeed. (p. 126)

The idea here is that, by virtue of its implementing many but not all features of a model, an experiment is intermediate between the model and the target domain. Thus, it is argued, the experiment allows a genuine but relatively weak test of the model.

This argument seems to assume that, in respect of those features of the model that are *not* implemented, the experiment is more like the target domain than the model is. (Without that assumption, there would be no ground for the claim that the experiment is intermediate between the model and the target domain.) However, Croson and Gächter provide no further argument in support of the assumption. As our case studies have illustrated, it is not self-evidently true. For example, one might reasonably ask: Which are more like business firms making pricing decisions – MOS's student subjects playing repeated 101×101×101×101 games for small money prizes, or the game-theoretically rational agents of Varian's model? A case could be made for either answer. Thus, granted that Varian's model uses MSNE as a theoretical component and that MOS's experiment is a test of that component in a laboratory environment that implements many features of the model, it is still an open question whether the experiment is intermediate between the model and its target domain.

It is significant that, in defending model-testing experiments, Croson and Gächter refer to the work of Charles Plott. As Croson and Gächter (p. 123) say, the methodological convictions of experimental economists have been shaped by the writings of Plott and Vernon Smith (e.g. Plott, 1982, 1991; Smith, 1982). Plott and Smith were two of the leading pioneers of experimental economics, and their early and very influential methodological essays had many common features. We shall argue that current views on the status of model-implementing experiments can be traced back to those essays.

In the 1980s, one of the main research programmes in experimental economics, and one in which both Plott and Smith were working, investigated the workings of markets. For this research programme, the seminal work was Edward Chamberlin's (1948) market experiment.

Chamberlin wanted to test one of the core predictions of neoclassical theory: that in a competitive market, trade takes place at the prices and quantities determined by the theoretical concept of competitive equilibrium. In traditional neoclassical models, the process by which equilibrium is reached either is not explained at all, or is explained by introducing some fiction, such as the Walrasian auctioneer or Edgeworth's recontracting mechanism. The model supports predictions about real markets by means of the hypothesis that real markets work *as if* presided over by an auctioneer, or *as if* traders were able to recontract. Chamberlin's experiment was designed to 'illuminate' the problem of 'the effect of deviations from a perfectly and purely competitive equilibrium under conditions (such as in real life) in which the actual prices involving such deviations are not subject to "recontract" (thus perfecting the market), but remain final' (1948, p. 95).

The neoclassical prediction is difficult to test in the field, because individuals' demand and supply functions are not directly observable. Chamberlin realised that a direct test would be possible if demand and supply functions were subject to experimental control. In his experiment, student subjects played the roles of buyers and sellers. The good to be traded was represented by tokens. Reservation values were *induced* by the payoff mechanism of the experiment. (For example, a seller was endowed with a token and told that if she failed to sell it, it would be redeemed at some specified cash value.) From the viewpoint of the experimenter, equilibrium price and quantity were determined by these induced values; but the equilibrium was not known by the subjects themselves. Subjects were brought together in a room and allowed to circulate and engage in bilateral bargaining. When a contract was made, the agreed price was written up on a blackboard. Neoclassical theory was tested by comparing actual trades with those implied by competitive equilibrium.

Chamberlin's experimental design provided a template for a research programme which investigated the workings of different market institutions, different numbers of traders, different supply and demand conditions, and so on. This programme was the main point of reference for Plott's and Smith's methodological essays.

One recurring theme in these essays is that experimental markets should not be thought of as models of markets: they are *real markets*. For example: 'An important message of the paper ... is that laboratory micro-economies are real live economic systems, which are certainly richer, behaviourally, than the systems parameterized in our theories' (Smith, 1982, pp. 923–4). Or: 'The trick is to notice that economies created in the laboratories might be very simple relative to those found in nature, but they are just as real' (Plott, 1991, p. 905). This insistence on the reality of laboratory markets is reflected in Plott's and Smith's use of expressions such as 'in the field' or 'in the wild' to refer to what theorists would call the 'real world'.

Plott and Smith argue that, because laboratory markets are real markets, they are located in the target domain of economic theories of markets and hence allow valid tests of those theories. Because laboratory markets are *simpler* than their equivalents in the field, experimental tests can be more controlled than field tests. If a theory succeeds in the laboratory, one cannot be confident that it will succeed in the field; but if it fails in the laboratory, there is a presumption that it is seriously deficient. Thus:

> Microeconomic theory abstracts from a rich variety of human activities which are postulated not to be of relevance to human economic behaviour. The experimental laboratory, precisely it uses reward-motivated individuals drawn from the population of economic agents in the socioeconomic system, consists of a far richer and more complex set of circumstances than is parameterized in our theories. Since the abstractions of the laboratory are orders of magnitude smaller than those of economic theory, there can be no question that the laboratory provides ample possibilities for falsifying any theory we might wish to test. (Smith, 1982, p. 936)

And:

> General models, such as those applied to the very complicated economies found in the wild, must apply to simple special cases. ... Since the laboratory economies are real, the general principles and models that exist in the literature should be expected to apply with the same force to these laboratory economies as to those economies found in the field. The laboratories are simple but the simplicity is an advantage because it allows the reasons for a model's failure to be isolated and sometimes even measured. (Plott, 1991, p. 905)

The Plott–Smith account clearly views experiments as intermediate between theory and field, as in Croson and Gächter's 'mental model'. Notice also that experimental

economics is viewed primarily in relation to theory, rather than the field. Plott is particularly explicit about this:

> Once models, as opposed to economies, became the focus of research the simplicity of an experiment and perhaps even the absence of features of more complicated economies become an asset. The experiment should be judged by the lessons it teaches about theory and not by its similarity with what nature might happen to have created. (Plott, 1991, p. 906).

Here Plott is using what Bardsley et al. (2010, pp. 54–56) call the *blame-the-theory argument*. The claim is that, if an experiment is designed to test a general theory, the experimenter does not need to address issues of external validity. Provided that the laboratory environment is in the domain of the theory, similarity between experiment and field is not necessary. If some lack of resemblance between experiment and field reflects an unrealistic assumption of the theory, that does not compromise the experiment as a test of the theory. In Smith's words:

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

In the passage to which Croson and Gächter refer, Plott (1982, p. 1520) adds the suggestion that if a theory's simplifying assumptions are reproduced in the experiment, that makes the theory less likely to be disconfirmed – the 'best shot' idea.

We suggest that the Plott–Smith account of the methodology of experimental economics has been internalised by many practitioners, and has been seen as legitimating model-implementing experiments such as those of MOS and AH. Croson and Gächter's discussion of experiments as tests of model predictions is an example of this line of thought. But is it right?

Plott's and Smith's reluctance to address issues of external validity was perhaps understandable, given the status of experimental economics in the pioneering era of the 1970s and 1980s. By insisting that their designs were valid tests of received theories, experimental economists were able to sidestep potential criticisms of what was then a controversial

methodology. Ironically, however, the research programme that was being defended in this way could also have been defended on grounds of external validity.

Consider Chamberlin's classic experiment. This is not an implementation of the Walrasian or Edgeworthian model of competitive equilibrium in the sense that MOS's experiment implements the clearinghouse model or AH's experiment implements BHW's model of information cascades. Certainly, Chamberlin's method of inducing reservation values implements an assumption of neoclassical models, namely that agents have well-articulated preferences over the objects that are being traded. But notice that there is no attempt to implement the Walrasian auctioneer or Edgeworth's recontracting procedure. This is not a problem of feasibility. For example, it would be straightforward to instruct one subject to act as an auctioneer, with incentives to find prices which minimise the absolute value of excess demand. One could then test whether the task performed by the auctioneer in Walras's model was within the competence of a human subject (just as MOS test whether human subjects can compute MSNE). We take it that Chamberlin was more interested in whether real markets work *as if* presided over by an auctioneer.

Recall that Chamberlin's experimental market was designed to investigate how markets work in conditions 'such as in real life' where the assumptions of perfect competition do not hold. The price-determination mechanism of the experimental market is intended to be *more realistic than* that of the neoclassical model. In other words, the experimental market is similar to real-world (or 'field') markets in ways that are not mediated by the model: in this respect, external validity is designed into the experiment. In terms of Smith's characterisation, Chamberlin's experiment consists of a far richer and more complex set of circumstances than is assumed by the neoclassical theory of competitive markets, and so provides ample possibilities for falsifying that theory. But – and this is equally important – the richness and complexity that have been added are similar to features of real-world markets that do not appear in the neoclassical model. Thus, the experiment is not a test of the neoclassical model itself. It is not a model-implementing experiment in the sense that MOS's and AH's experiments are. It is better understood as a test of hypotheses about the real world for which the model has provided support.

Let us make ourselves clear. We are not arguing that the Plott–Smith defence of market experiments as tests of neoclassical theory was invalid, but only that it downplayed the extent to which these experiments were designed to be similar to the target domain of the

theory. In doing so, it founded a folk methodology in which model-implementing experiments are treated as more informative than they really are.

8. Conclusion

We end with some practical advice, addressed to practising experimental economists. Whenever you are planning to base an experiment on a theoretical model, first ask yourself in what respects what happens in the experiment could possibly be different from what happens in the model. These respects set outer bounds to the questions that the experiment can possibly answer. Next, ask yourself whether the questions that you want to answer lie within these bounds. If they don't, the experiment is pointless. If they do, ask whether the model you are planning to use provides the best design for answering those questions. What you should *not* do is implement a model just because it is interesting or insightful or famous, and because you will be the first experimental economist to do so.

We know that summer school attendees are sometimes given exactly the opposite advice by well-established experimental economists. Perhaps, given the folk methodology of experimental economics, being the first experimenter to implement a well-known model *is* an effective recipe for achieving publications. But if it is, that is only because that folk methodology is flawed and because the conviction with which it is endorsed in the profession is misplaced. Our aim in this paper has been to explain why.

References

Anderson, Lisa and Charles Holt (1997). Information cascades in the laboratory. *American Economic Review* 87: 847–862.

Aydinonat, N. Emrah (2007). Models, conjectures and exploration: An analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology* 14: 429–454.

Bardsley, Nicholas (2005). Experimental economics and the artificiality of alteration. *Journal of Economic Methodology* 12: 239–253.

Bardsley, Nicholas, Robin Cubitt, Graham Loomes, Peter Moffatt, Chris Starmer and Robert Sugden (2010). *Experimental Economics: Rethinking the Rules*. Princeton: Princeton University Press.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100: 992–1026.

Cartwright, Nancy (1998). Capacities. In John Davis, Wade Hands and Uskali Mäki (eds), *The Handbook of Economic Methodology* (pp. 45–48). Cheltenham: Edward Elgar.

Cartwright, Nancy (2002). The limits of causal order, from economics to physics. In Uskali Mäki (ed.), *Fact and Fiction in Economics* (pp. 137–151). Cambridge University Press.

Chamberlin, Edward (1948). An experimental imperfect market. *Journal of Political Economy* 56: 95–108.

Croson, Rachel and Simon Gächter (2010). The science of experimental economics. *Journal of Economic Behavior and Organization* 73: 122–131.

Friedman, Milton (1953). The methodology of positive economics. In *Essays in Positive Economics* pp. 3–43. University of Chicago Press.

Gibbard, Alan and Hal Varian (1978). Economic models. *Journal of Philosophy* 75: 664–677.

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

Guala, Francesco (2005). *The Methodology of Experimental Economics*. Cambridge University Press.

Hausman, Daniel (1992). *The Inexact and Separate Science of Economics*. Cambridge University Press.

Mäki, Uskali (1992). On the method of isolation in economics. *Poznań Studies in the Philosophy of Science and the Humanities* 26: 316–351.

Mäki, Uskali (2003). 'The methodology of positive economics' does not give us *the* methodology of positive economics. *Journal of Economic Methodology* 10: 495–506.

Mäki, Uskali (2005). Models are experiments, experiments are models. *Journal of Economic Methodology* 12: 303–315.

Morgan, John, Henrik Orzen and Martin Sefton (2006). An experimental study of price dispersion. *Games and Economic Behavior* 54: 134–158.

Morgan, Mary (2005). Experiments versus models: new phenomena, inference and surprise. *Journal of Economic Methodology* 12: 317–329.

Plott, Charles (1982). Industrial organisation theory and experimental economics. *Journal of Economic Literature* 20: 1485–1527.

Plott, Charles (1991). Will economics become an experimental science? *Southern Economic Journal* 57: 901–919.

Salop, Steven and Joseph Stiglitz (1977). Bargains and ripoffs: a model of monopolistically competitive price dispersion. *Review of Economic Studies* 44: 493–510.

Schelling, Thomas (2006). *Strategies of commitment and other essays*. Cambridge, MA: Harvard University Press.)

Smith, Vernon (1982). Microeconomic systems as an experimental science. *American Economic Review* 72: 923–955.

Sugden, Robert (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology* 7: 1–31.

Sugden, Robert (2009). Credible worlds, capacities and mechanisms. *Erkenntnis* 70: 3–27.

Varian, Hal (1980). A model of sales. *American Economic Review* 70: 651–659.

Varian, Hal (1998). How to build an economic model in your spare time. In Michael Szenberg (ed.), *Passion and Craft: Economists at Work*. Ann Arbor: University of Michigan Press.

Weizsacker, Georg (2008). Do we follow others when we should? A simple test of rational expectations. Forthcoming in *American Economic Review*.

Notes

[1] Bardsley (2005) and Bardsley et al. (2010: 204–214) are exceptions. Our analysis in this paper is broadly consistent with those presented in those texts. We develop Bardsley's original argument with more detailed consideration of case studies and in closer relation to questions about the role of theoretical models in economics. Guala (2005: 203–230) offers a normative account of the role that models *should* play in experimental economics. We agree with many of his arguments, especially about the importance of external validity. However, as our case studies illustrate, the practice of experimental economics does not always follow Guala's prescriptions.

[2] Some commentators have raised doubts about how far Friedman's methodology really is instrumentalist: see e.g. Mäki (2003).

[3] This is a possible reading of Schelling's (2006, pp. 235–248) account of 'social mechanisms', discussed by Sugden (2009). Aydinonat (2007) has defended a similar interpretation of Schelling's checkerboard model of segregation.

[4] This conclusion is qualified by the results of a subsidiary treatment, which we describe in Section 5.

[5] This position is developed and defended by Bardsley et al. (2010: 46–94).

[6] In the case of BHW's model, for example, evolutionary biology provides reasons for expecting some of the properties of rational choice and game theory to apply to animal behaviour in natural environments. Thus it might be argued that non-Bayesian behaviour by subjects in AH's experiment would not compromise BHW's explanation of sage grouse behaviour.

[7] BHW state their rationality assumption as 'perfect Bayesian equilibrium', but this assumption is stronger than they need. Their results can be derived from the assumption that it is common knowledge that all individuals maximise expected utility and make Bayesian inferences.

[8] AH give more prominence to the fact that 'only a third of the deviations from Bayes' rule ... can be explained by counting' (p. 859), but this is not a neutral comparison between the two rules.