Published in Journal of Economic Methodology Vol. 15, No. 2, June 2008, 197–204

What Experimental Economics Teaches Us About Models

Anna Alexandrova

Experimental economics, as Francesco Guala's book shows, is rich in its implications for philosophy of science. I shall focus my commentary on its implications for our conception of economic models. This is, of course, only one aspect of the book, but a significant one. The sort of models that experimental economics engages, i.e. rational choice models, are currently spreading from economics throughout political science and sociology. Similar model are already central to parts of biology. Philosophers of science have long recognized the puzzle of such models: they are deeply simplified, often poor predictors, and yet appear to be useful. In particular, we have found it tricky to give a good account of the role of these models in empirical successes such as explanations and policy interventions. The field of experimental economics (hereafter EE) has obviously made important progress with respect both to developing explanations and to justifying policy interventions. So it is natural to look to how EE uses models, in order to extract philosophically useful lessons.

Guala's book gives us a rich description of the practice of EE. I think this description makes possible the testing of different philosophical accounts of such models against the evidence from EE. Here I sketch a way of doing so. In the book, Guala mentions various existing accounts of models but does not settle on any particular one as the correct account. Furthermore at times he suggests that no single account of how models are applied is possible (156). I maintain that Guala's account of the methodology of EE does allow us to weed out certain accounts of models, and also suggests features of a new account.

So what is the relation between models and experiments in EE? Clearly economists interpret models as sources of some kinds of causal claims that can then be tested in experiments and used to justify various institution designs. For example, that such and such rules of an auction (for example, first vs second price rule) cause such and such bidding patterns (lower or higher bids) under such and such conditions that could describe information access, payment terms etc. What is the relationship between models and these claims?

Guala tells us that these hypotheses come from manipulating models, that is from asking 'what if' questions and then answering them by changing the appropriate assumptions of the model (those corresponding to the 'what if' question at hand) and following through the deductive consequences in the model (210). Take, for instance, a simple game theoretical model of a private-value first-price auction with two identical bidders. Under a particular set of assumptions (continuous and uniform distribution of valuations, risk neutrality, Bayesian Nash Equilibrium etc.) the model predicts that bidders will bid half of their valuation. However, switching from the first-price to a second-price rule, while holding all other assumptions fixed, implies that they will bid the amount equal to their valuation. The hypothesis then, following this suggestion, must be "the first price-rule causes bidding below true valuation under the conditions described by the assumptions". This is, allegedly, the hypothesis that gets tested by an experiment. Of course, this is not the only claim that is being tested. A number of assumptions are required in order to translate the hypothesis deduced from the model into a specific prediction that can be shown to be true or false by the experiment.

Note that on the story above the hypothesis comes with a description of at least one set of the assumptions under which it holds. However, throughout the book Guala shows the ways in which models by themselves do not specify empirical claims that can be put to test. In particular, models do not come with a complete specification of the domain of application of their predictions. There are a number of ways in which models can fail to do so. Once all these ways are specified we shall see that in many cases the story above cannot be right.

Assumptions may fail to describe the conditions under which the causal relationship suggested by the model holds in at least the ways listed below:

1. Assumptions may not give us the full detailed description of all the conditions that need to be in place to realize the causal relation of the hypothesis. For example, no game theoretical model tells us the properties of the software that need to be in place for bidding to proceed in the way that game theory predicts. Guala describes this circumstance in detail in the chapter on auction design. In the FCC case, experiments were crucial precisely because they allowed researchers to study which environments counted as satisfying a particular theoretical description (for example, which operational environments count as information-concealing given that blinks of a computer screen can reveal information). This insight is part of, for example, Nancy Cartwright's view of models via her distinction between abstract (often theoretical) and concrete (that is implementation-specific) descriptions (Cartwright 1989). This is the least controversial way in which models might fail to specify the environments in which they apply. No new account of models is necessary to make sense of this possibility.

2. Assumptions may also fail to give an exhaustive and exceptionless list of conditions (even in abstract vocabulary) of all the conditions which realize a causal hypothesis. This is the idea that the causal claim may hold in conditions narrower or broader than those described in the assumptions of the original model, because these assumptions operate at the level of concepts that are not fundamental. They supervene on concepts that are studied by more fundamental sciences. Guala takes this option into account when criticizing "universalist" views of theories (153). This exclusion is also part of Dan Hausman's specification that economic theories have a vague CP clause (Hausman 1992).

3. Alternatively, assumptions may fail to tell us the conditions under which the causal hypothesis is true when the assumptions are violated. For example, a first price rule may still cause bidding below true valuation even when the bidders in question are not perfectly rational or do not play Bayesian Nash Equilibrium. Such circumstances are explicitly contradicted by the model, but the causal relation (one certainly hopes) may still hold. Guala discusses this possibility on pages 215-216.

4. Finally, we may consider the most controversial possibility. Models may fail to describe even one set of empirical conditions under which the causal relation in the hypothesis holds. This may happen if some of the model's assumptions are such that they cannot, as far as we know, be realized in the real world. Some mathematical assumptions detailing the nature of utility functions, or probability spaces may be candidates for that. It would be foolhardy to claim that assumptions such as these cannot refer *because* they are mathematical. But it is a possibility that we may want to leave open.

While every one of these is a real possibility, the first three are actual features of application of economic models. However, not all of these are compatible with all accounts of models. In fact I claim that 3 (and 4) cannot be sustained if the hypotheses that models supply are thought to be deductions formulated by manipulating the model, as Guala appears to imply.

Here's what I mean. Take again our standard auction model described above. Clearly, the answer to the question "what if the rules of the auction change to second price?" is, "bidders start bidding their true valuation". But what exactly is the hypothesis?

Presumably a hypothesis that then gets tested in an experiment comes in the following form: "Feature(s) F(s) cause behavior(s) B(s) under condition(s) C(s)". In our example, the feature F in question is the auction's rules, the behavior B is the proportion of valuation that the player bids. What are the conditions Cs? It is natural to say that the conditions are just the assumptions which we use to deduce Bs from Fs. Indeed this is the natural reading of Guala's claim that hypotheses are obtained by manipulating the model. However, this reading is false for all cases in which possibility 3 above applies.

In fact Guala seems to agree and later in chapter 10 he explains why: experiments need not instantiate the model's assumptions exactly (i.e. in our language the assumptions need not be read as specifying Cs), because if they were there'd be nothing to learn from an experiment. We already know what must happen when all the assumptions are satisfied! Here Guala seems to be saying that something like possibility 3 is not merely possible but that it is actually necessary for experiments to be informative. But if so, then Cs are not just the assumptions of the model.

Note the claim here is not that economic models do not make claims about contingent causal relations *because* its conclusions are entailed logically from its premises. The fact that the conclusions of a model follow logically from its premises should not be thought as a reason to either deny or grant models' derivations the status of causal claims. The relation of logical entailment holds between sentences, while causal relations hold, on most theories of causality, between events. So whether sentences that describe these events stand in a relation of logical entailment is entirely orthogonal to whether or not the relation between these events is causal.[1] Rather, Guala's claim is that the exact sort of causal relations that a model might describe are not the causal relations that experiments investigate.

There are other, more pragmatic, reasons not to treat assumptions as a guide to Cs. These reasons are particularly relevant when economic models and experiments are used for engineering institutions such as spectrum auctions, as described by Guala in chapter 8. Sometimes it is simply not known whether or not some assumption essential for deriving a particular effect in the model can be satisfied by the target system economists are constructing. These may be facts about their system that experimentalists cannot ascertain, for example, the statistical properties of distributions of valuations that are crucial in auction models. The experimentalists' hope is to find some other empirical conditions, not mentioned in the model, under which features F still cause behaviors B.

At other times it is known that some assumption describing a condition C is not satisfied by the target system, but auction designers have no control over this condition and so cannot *make* the target system satisfy this assumption. This may be the case with the assumption of Bayesian Nash Equilibrium. Auction designers know that the flesh and blood first-time spectrum bidders they have to deal with in the actual auction could not be expected to have the sort of rationality that auction models assume. So allowances are made for lack of experience, for the fact that the auction is complex and that conformity with the predictions of models that assume rational choice is just unlikely in a one-shot game with no practice. So a variety of other rules get added to the auction to push the bidders to behave as required to ensure efficiency. Again, some set of conditions C, other than the one specified by the model, needs to be found to make Fs cause Bs.

What then is the hypothesis in relation to a model? That is, where do the Cs come from? If they are not fully specified by the model, they may be specified in other ways, from experimental, observational or other background knowledge. But this implies that models, rather than yielding hypotheses with the help of a "what if" question plus some deduction, actually play a rather different role. My view is that they are best regarded as providing frameworks for formulating hypotheses, or schemas of the following form:

(OF) In a situation x with some characteristics that may or may not include $\{C_1 \ldots C_n\}$, a certain feature F causes a certain behavior B.

I call this schema OF because it is an *open formula*, where x is a variable, F and B are property names, respectively, of putative causes and effects in the model, and $\{C_1 \ldots C_n\}$ are the conditions under which Fs cause Bs in the model. In an open formula, by design, there is no commitment to the existence of x (x is not quantified over) and not yet any claim about any phenomenon since the features of x are not specified. x is a free variable,

[1] I thank Agustin Rayo for making this clear to me.

which needs to be filled in in order for the open formula to make such a claim. Once x is specified, we get a causal hypothesis of the form "an F causes a B in a situation S", where S is characterized by some conditions C. Without closing the open formula by specifying x, (OF) only gives us a template or a schema for a causal claim, rather than a fully fledged causal claim. (OF) is further weakened by the "may or may not" clause which specifies that the causal relation between F and B need not hold under the assumptions C specified by the model.

This is the view that makes best sense of Guala's claim in chapter 10 that when an experiment is designed on the basis of a model, it is not the model itself which is being tested but rather an *application* of this model (219). He defines an application of a model as a "hypothesis stating that certain elements of a model are approximately accurate or good enough representation of what goes on in a given empirical situation" (219). This is one way of defining what an application is, but it is both too vague and uncontroversial and too strong and too controversial at the same time.

It is too vague and too uncontroversial because it employs such notions as approximate accuracy and good enough representation. It is not just that we have no reductive account of "good enoughness" or approximate truth. Rather it is not clear such notions are informative, and that the intricate tests experimental economists design are testing such vague claims. This definition is also too strong and controversial because it states the relationship between the model and the experiment in terms of representation. Do models represent experiments that inspire them? Maybe, maybe not, it depends on what our account of representation is, an issue much more complicated than necessary for our purposes. It is much less controversial to say that models yield hypotheses that, if confirmed, explain empirical phenomena and/or that justify interventions and policy measures.

So far this might look merely like a friendly amendment, but I would go further and claim that once we view models as open formulae and take the lessons from the practice of EE on board, we must reject other existing views of model application, in particular, the *satisfaction of assumptions* (Hausman 1992) and the *concretization of capacities* (Cartwright 1989, 1999 etc.) views as incomplete. This claim is inspired mostly by Guala's chapter 8, which tells the story of economic engineering.

On the capacity reading of models, inspired by JS Mill and developed by Nancy Cartwright, models make capacity claims. Capacities are stable features that produce regular effects in the absence of disturbing factors, that is in idealized circumstances (Mill 1836). However, even when other capacities or disturbing factors are present capacities are still exercised, which is what allows us to export our knowledge of tendencies from the idealized and controlled conditions of the laboratory or a model to the real world (Cartwright 1989 chapter 4). Crucially, capacities are characterized by their stability. For example, when we say that negatively charged bodies have the capacity to make other negatively charged bodies move away, we do not just mean that they make others move away in certain ideal or near ideal conditions, but rather that the capacity is exerted across a number of environments, even when it fails to manifest itself by actually moving the other body away.

What is the relation between capacities and the models in question here? According to early Cartwright (1989, 1999a), we build models such as those in economics in order to investigate the canonical behavior of a capacity. For example, on the basis of the model I mentioned above we may conclude that the first-price auction has the capacity to lower bids under the conditions of private values. In models we idealize away the disturbing factors to allow the capacity of interest to manifest its 'pure' behavior (Cartwright 1999a, chapter 4).[2] We then apply models to a phenomenon under explanation by the process of *concretization* – whereby we add together all capacities we think are at work in this phenomenon using rules that describe how capacities combine.

Does this picture fit the practice of experimental economics? It think that it doesn't. And this is particularly clear from Guala's discussion of external validity. He argues repeatedly that external validity of an experiment or a model is an empirical matter to be established bit by bit. When extrapolating from one environment to another their similarity needs to be established carefully. It pays to be conservative. It pays not just to assume the stability of laboratory or theoretical results in the face of potential disturbing factors. And yet without stability capacities lose their power. As mentioned above capacities, at least within a certain range of circumstances, are supposed to have stability in the face of other factors. This means that within this range there are no interactive effects, in other words that the contribution of one cause does not change when some other cause is introduced (Cartwright 1989, 163). This is precisely what enables exportability. On the picture Guala paints, this is certainly not the experience of economists who design auctions. They find interactions between causal factors more often than not, and believe that the stability of causes is a poor working hypothesis.

Experimental economist Charles Plott, whom Guala often cites, himself made numerous observations of interactions in the context of testing auction designs. To use just one example, there is much evidence, both theoretical and experimental, that open- rather than sealed-bid auctions defeat what economists call the 'winner's curse' – a phenomenon whereby the winning bid in an auction where the object for sale has an uncertain value is the one that most overestimates the object's true value thus resulting in a loss for the winning bidder. By allowing bidders to observe each other's bids, thus reducing uncertainty about valuations, an open auction is thought to counteract this result (McMillan 1994). Did this mean that it would have this effect in the FCC set up? Writing in 1997 Plott remained entirely agnostic. In commenting on the winner's curse he said: "How this [the winner's curse] might work out when there is a sequence of bids and complementarities is simply unknown. No experiments have been conducted that provide an assessment of what the dimensions of the problem might be" (1997, 626). Here Plott is making two claims: first, that the alleged capacity of the open auction to reduce uncertainty, if it exists, might be neutralized by sequential bidding and complementary values, and, second, that we just do not know how much that might affect the overall auction. He expressed similar skepticism about many other features of an auction that

[2] A similar reading of economic models is given by Maki 1992.

seem reasonable in isolation, but that once combined in one institution have unpredictable cascading effects (Plott 1997).

Of course, Cartwright does not invite us to assume capacities without good reason. To establish a capacity is major task. Sometimes as in the Newton-Goethe debate about optics (Cartwright 1999) one really well constructed experiment is enough, but that is presumably an exception. However, once a capacity is established we are entitled to assume its stability within a certain range of environments. But in the case of experiments, and even less in the case of models, EE does not have the luxury of treating its findings as claims about capacities.

Of course, it is open to Cartwright to respond that perhaps capacities in experimental economics are a lot more specific and environmentally sensitive than those in physics. Perhaps, but we don't seem to have evidence for that type of capacities either as Cartwright points out in more recent articles.[3] A more plausible response is that although we may not yet know many capacities in economics, that does not mean that they are not there. I fully agree with the latter claim. But our focus here is methodology not metaphysics. As far as methodology is concerned we must concede that although capacities may be a normative ideal, they are not the reality of EE.

What about another major account of model application, that defended by Dan Hausman? I believe that this treatment is also in conflict with Guala's description of EE.

Hausman develops a version of the semantic view of models. On his view, models do not by themselves make empirical claims. Rather, they supply definitions (Hausman 1992, 74-77). In order to relate to the world, models need to underwrite empirical claims, not just trivially true claims or definitions. On Hausman's view this is done via a *theoretical hypothesis*, which is a claim that some relevant class of the model's assumptions are satisfied by the target system (Hausman 1992, 77).[4] Once a model is supplemented with a theoretical hypothesis we arrive at an empirical claim of the form "Φ is true *ceteris paribus"* where Φ refers to a claim about an economic phenomenon made in the model. The assumptions of the model spell out the *ceteris paribus* (CP) conditions, or the conditions under which Φ is true (though they may not be exhaustive). I won't have anything to say here about the *ceteris paribus* reading of models and Hausman's story so far. Rather, I will concentrate on his account of application.

On Hausman's scheme, to use a theoretical hypothesis for explanation it is necessary to identify the set of the model's assumptions that must be satisfied by the target system. Which assumptions are these? Should we require that *all* the assumptions necessary for the derivation of a given result be satisfied by the target system? This seems much too strict. No actual auction satisfies *all* the assumptions of a game theoretical model, e.g. the assumption of perfect rationality. We need a criterion for distinguishing the relevant assumptions from the irrelevant ones.

[3] In Cartwright 1999b, 2007, Cartwright argues that economic models should not be read as making capacity claims because they are overconstrained.
[4] A similar view of application appears to be endorsed by Morgan 2002.

One such criterion can be supplied by the *de-idealization* approach – ignore those assumptions that can be replaced with more realistic assumptions while preserving (to some degree) the predictions of the model relevant to explaining the target phenomenon. On this approach, discussed by Ernan McMullin (1985), to be explained by a model a system has to satisfy all the assumptions of this model that cannot be de-idealized. If it does, then the model provides a causal explanation of some behavior of the target system.

The trouble is that once this view is applied to economics we are forced to admit that economic models do not genuinely apply either to the outcomes of the experiments or to the operation of the institutions that they inspired. This is because by and large the experiments, or institutions such as the spectrum auctions, do not satisfy all the non-de-idealizable assumptions of the models on which they are based. On Hausman's view the hypothesis the model makes just is the deduction from this model. Above, I have already discussed the problems with such a view. These problems are particularly vivid in the case of the spectrum auction design discussed in detail by Guala.

Auction models have many assumptions that were patently unrealistic of the Simultaneous Multiple Round Ascending auction adopted by the FCC. To give a few examples, bidders were not perfectly rational while all models assumed so; hundreds of licenses were on sale while there were hardly any multi-unit auction models at the time; models assumed no budget constraints while real bidders most probably had those, etc. In none of these examples was the de-idealization technique feasible. It was simply not possible, at least at the time, to build a model incorporating the more realistic assumptions and to check the effect of these on the model's predictions. Indeed there was no *one* theoretical model that was supposed to represent the actual auction, even at a very abstract level. This was known very well by the auction designers: "The setting for the FCC auctions is far more complicated than any model yet, or ever likely to be, written down" (McMillan, Rotschild and Wilson 1997, 429).

Perhaps Hausman is not bothered by the conclusion that model application, as used in experimental economics and in economic engineering, does not fit his account. He might say that models *inspire* experiments or institutions, but they do not genuinely apply to these particular targets. This is a plausible reply. But the trouble is that it is not clear that Hausman's account applies to *any* cases of model use in EE and related disciplines. It certainly does not apply to any of the cases Guala discusses in his book.

Let us take stock. Guala has shown a number of interesting implications of EE for philosophy of science. I invite him to be even bolder and to use evidence from EE to extract lessons about the nature and role of economic models. These lessons point to an account of models and their application that is substantially different from the leading existing accounts. I have only sketched such an account here. But we should already applaud Guala for enabling us to make progress on this thorny issue.

References

Cartwright, N. (1989) Nature's Capacities and Their Measurement, Oxford: Oxford University Press.

Cartwright, N (1999a) The Dappled World, Cambridge: Cambridge University Press.

Cartwright, N. (1999b) "Vanity of Rigour in Economics" in Discussion Paper Series, Centre for the Philosophy of Natural and Social Science, LSE, 1-11.

Cartwright, N. (2007) "Why Economic Models are a Bad Basis for Induction to the Real World", Manuscript, London School of Economics.

Giere, R.N. (1988) Explaining Science Chicago: University of Chicago Press.

Guala, F. (2005) Methodology of Experimental Economics. Cambridge: Cambridge University Press.

Hausman, D.M. (1992) The Inexact and Separate Science of Economics, Cambridge: Cambridge University Press.

Maki, U. (1992) "On the Method of Idealization in Economics" *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26: 319-354.

McMillan, J.R. (1994) "Selling Spectrum Rights," *Journal of Economic Perspectives,* 8/3: 145-62.

McMillan, J., M. Rothschild, and R. Wilson (1997). "Introduction." *Journal of Economics and Management Strategy* 6(3): 425-430.

Mill J.S. (1843) System of Logic, London: Parker.

Morgan, M.S. (2002) "Model Experiments and Models in Experiments" in Model-Based Reasoning: Science, Technology, Values, eds. L. Magnani and N.J. Nersessian, Kluwer Academic/Plenum Publishers, New York.

Plott, C.R, (1997) "Laboratory Experimental Testbeds: Application to the PCS Auction" *Journal of Economics and Management Strategy* 6/3 605-638.