Making Models Count*

Anna Alexandrova†‡

What sort of claims do scientific models make and how do these claims then underwrite empirical successes such as explanations and reliable policy interventions? In this paper I propose answers to these questions for the class of models used throughout the social and biological sciences, namely idealized deductive ones with a causal interpretation. I argue that the two main existing accounts misrepresent how these models are actually used, and propose a new account.

1. Introduction. What sort of claims do scientific models make and how do these claims then underwrite empirical successes such as explanations and reliable policy interventions? In this paper I propose answers to these questions for a specific class of models originally developed in economics and now used all over the social and biological sciences. The models I have in mind are deductive and typically come with a causal interpretation; such that some of their premises are treated as describing a putative cause and some of their deductive consequents an effect. No systematic empirical data goes into the construction of these models. However, they characteristically describe familiar entities such as agents with beliefs and desires, or populations and organisms, while also imposing various idealizing assumptions on their behavior.

Philosophers have discussed such models extensively, because they have so many different roles: they represent phenomena, they function inde-

*Received July 2006; revised August 2008.

†To contact the author, please write to: Department of Philosophy, University of Missouri, St. Louis, 599 Lucas Hall (MC 73), One University Blvd., St. Louis, MO 63121-4400; e-mail: alexandrovaa@umsl.edu.

‡ I owe many thanks to Nancy Cartwright, Francesco Guala, Robert Northcott, Jay Odenbaugh, Julian Reiss, Till Gruene-Yanoff, Harold Kincaid, Don Ross, and many others including my teachers and friends at UC San Diego and audiences at the Philosophy Department colloquia at New York University, University of Missouri Columbia, Washington University in St Louis, Center for Philosophy of Science at Pittsburgh, London School of Economics, the 2007 meeting of the American Philosophical Association Pacific Division, and the 2007 meeting of the British Society for Philosophy of Science.

Philosophy of Science, 75 (July 2008) pp. 383–404. 0031-8248/2008/7503-0006$10.00
Copyright 2008 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

pendently of theories, they aid in theory development, etc. (van Fraassen 1980; Giere 1988; Morgan and Morrison 1999; and many others). In this paper I concentrate on only one role of these models, that is their role in explanations of particular phenomena and interventions by creating technology. This is how models may guide policy. I consider two prominent accounts of how models do that. On the first account these models are treated as licensing ceteris paribus claims and their application consists in satisfaction of some set of the model's assumptions by the target phenomenon. On the second, models are read as claims about tendencies or capacities, and they apply if they capture the main causes of the target phenomenon.

On what grounds might we evaluate philosophical accounts of models and their application? In the spirit of practice-based philosophy of science, we may judge them on their ability to make sense of one of the most successful and representative uses of economic theory for empirical purposes—the design of reliable incentive-compatible institutions. The theory I have in mind is a branch of game theory known as auction theory, and the particular institution is the electromagnetic spectrum auctions constructed on the basis of this theory and now used across Europe and North America to distribute spectrum licenses to firms. I shall argue that the existing accounts of models and their application do not fit the role of models in generating the sort of knowledge that underwrites institution design. According to these accounts, the use of game theory for auction design simply does not count as theory application. I take it to be an unwelcome result that a philosophical theory of scientific models is unable to account for a case that scientists themselves consider to be a paradigm example of model application. I thus propose a new account, one that makes models count. Essentially this account adopts a *weaker* reading of model claims, not as ceteris paribus or capacity claims, but as *open formulae* in the sense to be explained in Section 6.

In Section 2 I give an example of a model in auction theory and briefly describe the institution which such models are supposed to help to design. Sections 3 and 4 consider how the existing philosophical accounts interpret the claims of such models and the application of these claims to the explanation of target systems. In Section 5 I explain the use of theory in auction design and argue that the standards of theory application assumed in the existing accounts are not satisfied by the case in question. Section 6 proposes a new account of models in terms of *open formulae*, and Section 7 of their application in terms of *material realization*.

2. Private-Value First-Price Auction Model. A typical deductive model with a causal interpretation can be found in a branch of game theory known as auction theory. Auctions are strategic situations in which bid-

https://doi.org/10.1086/592952 Published online by Cambridge University Press

ders' behavior and beliefs are a function of what they expect other bidders to do and of the rules of the auction. These are precisely the features that can be built into a game theoretical model, which is why game theory so revolutionized auction theory. Nowadays auction theory is a collection of models that explore the effects of different rules, information distribution and beliefs on bidding behavior, prices and revenues. Theorists solve games by assuming, along with many other things, Bayesian Nash Equilibrium, and varying conditions such as auction rules (open or sealed-bid, second- or first-price, with or without reserve price, etc.) and types of information known to bidders (e.g., whether or not they receive it from the same source).[1]

A sealed bid auction with two buyers, for example, might be modeled in the following way. Two players, Player 1 and Player 2, are competing for a good in a first-price sealed-bid private-value auction. These assumptions mean respectively that the winner pays an amount equal to the highest bid, that players observe only their own and not the other player's bid, and that each player is certain about how much they value the good for sale.[2] Thus Player 1 knows her valuation v_1 of the good and observes her bid b_1, and Player 2 knows her valuation v_2 of the good and observes her bid b_2.

Assume that both v_1 and v_2 are uniformly and continuously distributed in the interval between 0 and 1,000. This allows us to establish that $\Pr(v_i < x) = x/1{,}000$.

How should the players bid? On the assumption of Bayesian Nash Equilibrium, rational players will attempt to bid in a way that maximizes their expected utility given their beliefs about what the other player does.

Suppose Player 1 believes that Player 2 will bid $b_2(v_2) = av_2$ where a is some constant. And suppose Player 1 considers bidding x, in which case her payoff is $(v_1 - x)$ if she wins and 0 if she loses. Given a further assumption that the bidders are risk neutral, Player 1's expected utility (EU_1) can be expressed as follows:

$$EU_1 = (v_1 - x)\Pr(b_2 < x) + 0\Pr(b_2 > x).$$

Since $b_2 = av_2$,

$$EU_1 = (v_1 - x)\Pr(av_2 < x).$$

1. For an introduction to auction theory see Klemperer 1999.

2. The assumption of private values is appropriate, for example, when buyers compete for a work of art that they value for its beauty rather than its resale value.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

Knowing that $\Pr(v_i < x) = x/1{,}000$ we can rewrite Player 1's payoff function as

$$\mathrm{EU}_1 = (v_1 - x)(ax/1{,}000).$$

To find the optimal value of x we solve the first order condition:

$$d/dx\mathrm{EU}_1(x) = 0,$$

$$x = v_1/2.$$

So Player 1 should bid half of her valuation, or $b_1 = v_1/2$. Since the players are identical (save possibly for their valuation), Player 2 should also bid half of her valuation or $b_2 = v_2/2$. Since both players are maximizing their expected utility given their beliefs about the actions of the other, this is a Bayesian Nash equilibrium.

This model, found in intermediate economic theory textbooks (e.g., Osborne 2003) is typical of the class of models in social and biological sciences this paper is about. Firstly, these models commonly concern behavior of familiar-sounding entities such as bidders and sellers, governments and firms, organisms and populations. However, no systematic empirical data about the behavior of these entities goes into construction of the model. Instead an abstract entity called, say, 'seller' is defined by means of idealizing assumptions that enable claims about the behavior of this entity to be *deduced* from the model. Secondly, scientists often interpret these deductive structures as supporting *causal* claims, for example, that the *first*-price rule causes bids *below* true valuation.

Institution design only begins with models such as the above, however. Its end goal is a reliable incentive-compatible institution in the real world. An example of the latter is the type of auction used by the Federal Communications Commission in the United States to distribute electromagnetic spectrum licenses to telecommunication firms, called the Simultaneous Multi-Round Auction (SMRA). In it, all available licenses are open for bidding at the same time and increasing bids can be submitted in successive rounds. Bidders can see the results of each round on their computer screens and must bid a minimum predetermined amount in order to maintain their eligibility. The SMRA is specially designed to handle such features as complementarity of spectrum licenses (when the value of a license to a firm depends on what other licenses it owns), lack of experience on the part of the bidders, uncertainty over licenses' value and future business potential, etc. When it functions reliably, such an auction gives the licenses to those firms that value them most, generates competition and raises much income for taxpayers.

Models such as the above inform auction design heavily. Moreover,

https://doi.org/10.1086/592952 Published online by Cambridge University Press

their features are routinely invoked in the explanation, not just prediction, of the behavior of real auctions. So how do we get from game theoretical models to the SMRA? In order to answer this question we first need to answer two more basic questions: What are we warranted to conclude on the basis of such a model? And how does that then figure in the explanation of the outcomes of actual auctions?

3. Application as Satisfaction of Assumptions. The first answer to these questions comes from an elaboration of the semantic view of theories defended by Daniel Hausman. On this view, models supply definitions of abstract terms rather than empirical claims (Hausman 1992, 74–77). For example, the model in Section 2 should be read as making the following claim: "rational bidders are such that in a first-price private value auction with such and such valuation distribution the bids are lower than their true valuation."

These definitions relate to the world via a *theoretical hypothesis*. A theoretical hypothesis is a claim that some class of the model's assumptions is satisfied by the target system (Hausman 1992, 77).[3] This hypothesis takes us from a definition to an empirical claim. For Hausman, the claim seems to be of the sort "in a first-price auction with rational bidders, bids are below the true valuation of bidders, ceteris paribus" (or more generally, "Φ is true ceteris paribus" where, using our example, Φ refers to a claim about how rational actors bid in such and such situation). The ceteris paribus (CP) conditions, or the conditions under which Φ is true, are spelled out by the model's assumptions (though not exhaustively). Note that the CP clause includes more than just the assumptions under which I derived the result Φ in Section 2. Other sets of assumptions can be used to derive Φ. For example, we do not need to assume that there are only 2 bidders, that their valuations have a uniform distribution, etc. So the CP clause refers to a complete set of sets of assumptions which imply the result Φ.

Thus to use a theoretical hypothesis for explanation we must identify some set of the model's assumptions that the target system must satisfy. One such assumption is that actors are rational. Another is that an auction is a first-price auction. If they are not satisfied, the model is simply not *about* the sort of auctions we are interested in explaining. But that is not enough, for not *all* real world first-price auctions and their behaviors will be explained by this model. Must therefore *all* the assumptions necessary for the derivation of a given CP claim be satisfied by the target system for this claim to explain its behavior? Surely not, because no actual auction

3. A similar view of application appears to be endorsed by Morgan 2002.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

does satisfy all the assumptions, and so then no actual auction could be explained by such a model. So what distinguishes those assumptions that need to be satisfied from those that don't?

One way to identify the needed set of assumption is by *de-idealization*—the 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 do *not* need to be satisfied by the target phenomenon.[4] This approach, discussed by Ernan McMullin (1985) and endorsed by Hausman (1994), may be summarized as follows:

> (**HM**). A model applies (for explanation or intervention) to a phenomenon if the phenomenon satisfies all the assumptions of the model that cannot be de-idealized.

If it does, then the causally interpreted ceteris paribus claim provides a causal explanation of the target behavior. If "first-price rule causes bids below true valuation" is true in a model and some target auction satisfies the relevant assumptions of the model, then the fact that this auction is first-price explains the presence of bids below true valuation.

Following the de-idealization technique, we relax those assumptions that we have a reason (from theory or from background knowledge) to believe unrealistic of the phenomenon to be explained. Imagine a model of a common value auction which we hope to apply to explain some feature of a real world common value auction (say, an auction for drilling rights in an oil field). If we have reason to believe that the real bidders are risk averse, while our model assumes risk neutrality, then we add this feature into our model and check whether the result still holds.

The point of de-idealization is to check that the relationship observed in the original model (for example, that between the first-price rule and bids below true valuation) would still hold once some of the assumptions of the original model were no longer satisfied. When de-idealization is possible, we have the warrant to move from a claim in a model to a claim about a real world target system. The fundamental problem with this

4. Another version of this criterion is known as *robustness analysis*, on which we replace false assumptions with other possibly false assumptions with the aim of arriving at a *robust theorem*, and without attempting to make the model more realistic (Weisberg 2005). On this view, the fact that no particular assumption is necessary for the derivation of a given result is evidence that their falsity does not matter. However, even the advocates of robustness analysis do not believe that robustness of a result in the above sense *by itself* shows that this result is therefore explanatory of the phenomenon it models. Correct prediction of the phenomenon's behavior is also necessary. Since robust theorems are rare in economics and not central to institution design, I omit discussion of them here.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

strategy, however, is that, at least in economics, it is applicable only in a limited range of cases.

For instance, the SMRA operated under conditions vastly different from those assumed by the auction models. Some clearly unrealistic assumptions included perfect rationality of bidders, the number of objects for sale, absence of budget constraints on the part of bidders, etc. The technique of de-idealization could not be applied to any of these. The impossibility, then or now, of building a de-idealized model that incorporates all of the more realistic assumptions and of checking their effect on the models' predictions, was well recognized 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, Rothschild, and Wilson 1997, 429). Even at an abstract level there was no *one* theoretical model that was supposed to represent the actual auction.

The important point for now is that the Hausman/McMullin account of model application tells a different story from what is actually going on in institution design. Whatever else we may say about the view of models as definitions and the claim that models are applied by making a theoretical hypothesis, if we want to understand institution design we need an account that goes beyond satisfaction of assumptions and de-idealization. Of course, the fact that this account does not fit the case of auction design, does not by itself speak against the view of models as ceteris paribus claims, nor does it discredit the technique of de-idealization—we do know how to de-idealize some albeit not all assumptions. I merely want to point out that de-idealization is not sufficient to understand the role of models in auction design.

4. The Capacity Account of Application. The second account makes use of the notion of *capacity*. It was introduced by Nancy Cartwright in order, among other things, to describe the point of idealized models such as the ones in question here.

What is a capacity? In Cartwright 1989 the notion is contrasted with that of a causal law: while a causal law holds only in a given population, a capacity is stronger in that, once established, we are justified in expecting it to obtain across different kinds of populations (1989, 146–147). Later, in Cartwright 1999a, a capacity is characterized by three features. Firstly, capacities are *causal*, rather than merely associational; secondly, they have *potentiality*, that is, a capacity produces a certain canonical manifestation when unimpeded; thirdly, capacities are *stable*. By stability Cartwright does not mean that a capacity, when present in a situation, always has the same manifestation (to allow that it wouldn't is the whole point of replacing regular associations with capacities). Negatively charged bodies have the capacity to move other negatively charged bodies away; but

https://doi.org/10.1086/592952 Published online by Cambridge University Press

when other forces are present, these bodies may end up closer together. Rather, capacities are stable in the sense that the sort of contribution they make does not change from situation to situation. Negatively changed bodies always *try* to move other negatively charged bodies away.

Cartwright contrasts this picture to John Maynard Keynes's vision of the world which she summarizes as follows: "not an atomistic but a wholistic world in which the behavior of any particular feature—the very contribution it makes to the effect from one occasion to another—depends on the setting in which it is embedded" (Cartwright 1989, 157). In such a world there are no capacities and the corresponding methodology does not apply.

What is this corresponding methodology in model-based sciences such as economics? For Cartwright, we build models such as those discussed above in order to investigate the canonical behavior of a capacity. For example, on the basis of the model in Section 2 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 1999b, Chapter 4).[5] However, the stability of a capacity means that it is not limited by the ceteris paribus clause in the model.

How does the model then get applied? The process of application, Cartwright argues following Polish philosopher Leszek Nowak, is the process of *concretization* (Cartwright 1989). Claims made by models are abstract in the sense of describing the behavior of a capacity "out of context" (1989, 187), that is, without stating the ceteris paribus conditions that accompany it in the model. Concretization involves adding back the factors (i.e., other capacities and disturbing factors) omitted in the model but present in the concrete situation we wish to explain. We might summarize the Cartwright/Nowak view of application as follows:

> (**CN**). A model applies to a phenomenon (for explanation or intervention) when the model plus corrections and additions to it identify the relevant capacities and disturbing factors underlying this phenomenon.

Different theories teach us how to concretize the model by providing us with *rules of composition* for capacities: "In game theory various concepts of equilibrium tell us how the self-interested tendencies of different agents will combine to produce a series of moves" (Cartwright 1995, 179).

Note that unlike the Hausman/McMullin account of application, the Cartwright/Nowak account does not require that a model's assumptions

5. A similar reading of economic models is given by Maki 1992.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

be satisfied or even approximately satisfied by the target system in order for the model to be of explanatory value. Given that capacities are stable enough to allow us to move from what is true in a model to what is true in the real world, models can provide explanatory insight even when their assumptions fail to be satisfied. Of course, the various factors we introduce during concretization must correctly describe the disturbing factors.

On this view the application of models requires knowledge of capacities and of the material conditions under which capacities are manifested in a particular way (these may include corrections, approximations, etc.). Some of this knowledge comes from theory—in particular, theory gives us abstract statements of capacities and some of the conditions that are relevant to their operation. But knowledge that allows us to *engineer* systems such as lasers in which capacities are "harnessed," to use Cartwright's term, for particular purposes also includes 'low-level' causal facts (say, about how the different materials that make up a laser react to each other) that are not part of any theory. On Cartwright's view, de-idealization can be understood as a technique for adding other capacities and disturbing factors into the model. However, we do not depend on whether or not it is possible to de-idealize the model by changing its assumptions and tracking derivations. De-idealization has a place but is not a necessary part of concretization. Instead we take whatever models we have and correct them in accordance with this knowledge of capacities and their composition. Unlike on the previous account, this correction can proceed by ways other than construction of a great big deductive model. Indeed, at some point we should expect theoretical tools to run out.

However, the advantages of this view come at a price. That price is the need to demonstrate that we indeed know some genuine capacity claims; in particular, that economic models can be read as supplying such claims. But this is a tough test to pass for a number of reasons.[6] Here I concentrate on the third characteristic of capacities, their stability. It means that within a certain range there are no *interactive* effects, in the sense that the contribution of one cause does not change when some other cause is introduced. That is, the world of economics cannot be "wholistic" (or at least generally "wholistic"), as Keynes is thought to have hypothesized. The experience of economists who design auctions does not conform to this picture. Interactive effects between causes are more common than not. So the stability of causes is a poor working hypothesis.

One researcher who made numerous observations of interactions was

6. Other reasons include the dearth of rules of composition in economics. Although economists know how to add some factors into their models, they lack this knowledge in many other significant cases, as the spectrum auction case illustrates. See also Cartwright 1999c for more reasons.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

Charles Plott, a Caltech experimental economist whom the FCC hired to test various auction designs. Perhaps the most telling example is his attitude to 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. There is theoretical and experimental evidence that open- rather than sealed-bid auctions defeat this curse (McMillan 1994; Kagel and Levin 1986). Intuitively, in an open auction bidders can observe each other's bids, and so the uncertainty about the value of the object for sale is reduced. Does this mean that it will have this effect in the FCC set up, as it would if this were a capacity? In a 1997 article summarizing his role in the FCC auction design Plott remained entirely agnostic: "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" (Plott 1997, 626). Here Plott is pointing out two kinds of uncertainty. Firstly, for all we know, the alleged capacity of the open auction to reduce the winner's curse, if it exists, might be neutralized[7] by sequential bidding and complementary values. Secondly, we do not know how much that might affect the overall auction. He expressed similar skepticism about many other features of an auction that seem reasonable in isolation, but that once combined in one institution have unpredictable cascading effects (Plott 1997). They are "cascading" precisely in the wholistic sense described above.

Of course, Cartwright's metaphysics allows the possibility of interactions. Capacities need not be universal. Outside their range of stability they fail: when acid and base combine, they neutralize each other's powers and "neither can operate to produce any effects at all" (Cartwright 1989, 163). However, the methodology of concretization that goes along with the metaphysics of capacities does not apply when interactions kick in.[8] If we do not have enough evidence that models make capacity claims, then adding other causes into the model will not produce the effect that is produced in reality when causes combine.

Furthermore, a Cartwrightian might deny the extent of interactive wholism that I claim obtains in economics. This would be an empirical issue to be settled by looking at economic reality. My examples from the spectrum auction case, at the very least, put the ball in the Cartwrightian's court.

7. Neutralized is not the same as counteracted. See the acid and base example in the next paragraph.

8. Cartwright explicitly acknowledges that concretization applies only where there is evidence of capacities in her response to Humphreys in Cartwright 1995.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

My claim is not that there are no capacities in the economic world, or that economists should not be aiming at discovering them. Rather my claim is that the sheer extent of interactive effects that experimental economists discover, is an indication that the method of concretization is unlikely to work in general. Interestingly, in a series of more recent papers Cartwright herself questions whether economic models are a good source for knowledge of capacities (Cartwright 1999c, 2006). She worries that the models might be too constrained to yield principles that fit the high standard of capacity claims. However, she does not explain how, given her new view, economic models can still be applied for the purposes of explanation and intervention. I propose to look for just such a method in the successful practices of institution design.

5. The Role of Theory in Auction Design. The problem with the two accounts discussed above is not that de-idealization or extrapolation on the basis of capacities are faulty techniques. They are not, and both are used by economists who design institutions on the basis of game theory. Models are often de-idealized to incorporate factors such as risk-aversion, uncertainty about valuations, various asymmetries in bidder's beliefs, etc. Similarly, interactive effects are not universal and there are certain stable causal facts known to economists (e.g., that the length of intervals between stages does not affect the price-generating properties of auctions, Plott 1997). The problem lies in the fact that there are also *other* methods that scientists often need to use in order to apply models. And these other methods do not fit naturally under either de-idealization or concretization.

One such method is derived from the field of experimental economics. Experimental economists create simple but controlled laboratory environments to study the behavior and decision making of individuals. Plott, whose team carried out many of the experiments for the FCC auction, calls these objects *experimental testbeds*, and treats them as prototypes of the more complex real life economic situations (Plott 1997). Experimental testbeds are crucial in auction design in many ways. They enable the testing of theoretical predictions, discovery of new facts, development and testing of the auction software, etc. (Guala 2005). For our purposes here, the most interesting role of experiments is to incorporate theoretical models into auction design even when both de-idealization and knowledge of capacities are unavailable. How is this accomplished?

Auction theory is a wide portfolio of models such as those discussed in Section 2 They explore the effects of uncertain valuations, of reserve prices, of complementarities between goods on auctions' efficiency, income generation, etc. For example, there are models showing that first-price auction can cause bids below true valuations, that open auction can defeat winner's curse, that individual rather than package bids can secure efficient

https://doi.org/10.1086/592952 Published online by Cambridge University Press

distribution, etc. The task for an auction designer is to figure out what rules an auction should have in order to achieve the goals desired by the FCC, such as efficiency and revenue generation. But how to find out the effect of features such as open auction rules, complementarities and simultaneous bidding when many of the assumptions of the models that study the effects of these features are not satisfied in the real world environment that the FCC auction designers anticipated?

Recall that it is often impossible to build a theoretical model of the auction by de-idealization for three reasons. Firstly, because it would be much too complex. Secondly, necessary theoretical tools may be lacking—for example, it is only very recently that auction theorists found a competitive equilibrium in some models involving complementarities (Milgrom 2004). Thirdly, auction designers cannot even ascertain whether some of their models' assumptions are satisfied or not. Auction models make assumptions about the shape of the distributions of bidders' valuations, but there is no way to gauge facts about these distributions, since potential bidders reasonably would not volunteer such information. Nor can researchers rely on the Cartwright-style concretization. It is not possible to combine the putative components of the auction and calculate the result on the basis of knowledge of capacities and their rules of composition, because, among other reasons, the effects of different features of the auction on its outcomes are not stable.

The FCC auction designers instead proceeded by trial and error, or rather, by a combination of informed guesses *and* trial and error. Guesses informed by theory and experimental knowledge were used to put together one auction technology (one set of rules and software) in an experimental testbed, and then checked, by running many experiments with Caltech undergraduates, if that technology produced the right result. If it didn't, they would put together a new combination and test its performance again, and so on and so forth. The method is not pure guesswork precisely because auction designers have reasons to believe some guesses to be better than others. Yet it is not concretization either, because on this method we test the auction *as a whole*.[9] The role of the experimental testbed is precisely to reveal one combination of rules and software, such that if this environment were instantiated the auction would run as required. Some features of this environment such as the open rules, public information, and bids on individual licenses, figured in models of auction theory. For example, there was theoretical reason to believe that open rules increase revenue when valuations do not have independent distri-

9. Another noncapacity-based method in quantum chemistry is described by Paul Humphreys (1995).

https://doi.org/10.1086/592952 Published online by Cambridge University Press

butions,[10] or that open rules reduce the winner's curse. However, since the models from which these results follow are not treated as making claims about the stable causal contributions of these features, it is the aim of the experiments to find an environment in which these features *altogether* bring about competitive bidding, speed, revenue generation etc. The system thus found is not treated as *decomposable*, that is, contrary to the method of concretization, it is not known what the effects of each of these features are except in this one environment. So models get applied even though economists find themselves in a wholistic Keynesian world.[11]

The experiments testing a number of different combinations of rules and software result in a very specific piece of knowledge: an open auction for goods such as spectrum licenses leads to efficient, speedy and revenue-generating distributions provided a particular list of conditions holds. These conditions include the detailed rules the FCC developed, the right software, the right legal and economic environment, and that certain interventions in the auction are possible. The conditions are intricate and specific. When it comes to generalizing this particular design beyond the 1994 auction, researchers are extremely careful (Klemperer 2002b). Nevertheless, auction design is widely regarded as one of the most successful areas of application of economic theory and the FCC auction is its article of pride. These auctions raised much more revenue than anyone predicted and, to the extent that economists can judge, distributed licenses efficiently (Crampton 1998). The FCC auctions look even more successful when we compare them to the spectrum auctions run in other countries such as Australia and New Zealand in the 1990s and Holland and Switzerland in the year 2000 (McMillan 1994; Klemperer 2002a). These auctions are widely acknowledged as expensive failures, despite the fact that the expertise of game theorists was available to their designers. The FCC designers with their special mix of theoretical and experimental methods did something right, and philosophers of science need to be able to explain this.

What would an account of models that makes sense of this example look like? Here are the desiderata. Firstly, while allowing for de-idealization, such an account should explain how models can be applied when

10. This property of valuations is called affiliation. When bidders use the same information to form their estimates of the good's value, this is a reasonable assumption to make.

11. Auction designers do know some perturbations of this environment that still preserve the outcomes (e.g., that it is safe to run this auction with telecom executives as well as Caltech undergraduates). So it is fairer to say that the auction was treated as minimally decomposable. For more on how the external validity of the auction technology was tested, see Guala 2005.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

their de-idealization is not feasible. Secondly, while allowing for concretization, it should explain how insights from different models can be combined when concretization is inapplicable. Thirdly, it should make sense of the methodology used in institution design, as described above, since this is the primary arena for the application of the models in question. In the next section, I outline such an account.

6. Models as Open Formulae. If experimental economics does not generally treat models as either capacity or de-idealizable ceteris paribus claims, how should we understand the role of models in this discipline? My proposal is to adopt a *weaker* reading of models: in the context of experimental and policy economics, models function as frameworks for formulating hypotheses, or as *open formulae*.

When scientists design economic experiments to test a prediction of a model, they do not attempt to produce an environment in which all the assumptions of the model are satisfied. That is not only practically impossible when experimental subjects are real flesh and blood people, but, as Francesco Guala points out, it is also pointless, since we already know from the model what *must* happen when all its assumptions are satisfied (Guala 2005). Rather models are used as suggestions for developing causal[12] hypotheses that can then be tested by an experiment. One natural causal hypothesis suggested by the model in Section 2 is: when values are private and some other conditions obtain, first-price auction rules cause bids below true valuation. This hypothesis has a form of "feature(s) *F*(s) cause behavior(s) *B*(s) under condition(s) that include *C*(s)." Note that not all the model's assumptions go into the specification of conditions *C*(s) under which *F*(s) cause *B*(s). They could, but they often don't. We fail to list certain assumptions *not* just out of economy of time or space or mental effort. Rather we *mean* to omit them, because we make a judgment about which assumptions are more salient than others for our purposes. I may be interested in the effect of auction rules on the bids, so with this prejudice in mind I formulate such a hypothesis on the basis of the model. I may, on the other hand, be interested in the effect of valuation distribution on the bids, in which case the same model is read as suggesting a different causal hypothesis (say, when there are two bidders, the continuity of valuations' distribution causes bids to be discounted by half). Some, such as the assumption of private valuations, often get retained in the hypothesis as a part of *C*(s); but others, for example the assumption that actors play Bayesian Nash Equilibrium, may not become part of the hypothesis. There is a distinction between such a causal hy-

12. Here I assume that the explanations of interest in experimental economics and institution design are causal.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

pothesis and the model proper. The model plays an important role in formulating such a hypothesis by providing some of its categories, but it does not fully specify it.

When a model is used in this way, it is natural to interpret it as an *open formula*. I use the term open formula in the technical sense used in first-order logic: a well-formed formula with a free variable, i.e., no quantifier. Until the open formula is closed either by bounding the variable with a quantifier or by specifying a singular term, it is not a sentence but rather just a framework for one. The open formula in our case takes the form:

[1] In a situation of type[13] x with some characteristics that may include $\{C_1 \ldots C_n\}$, a certain feature F causes a certain behavior B.

Where x is a variable, F and B are property names of, respectively, putative causes and effects in the model, and $\{C_1 \ldots C_n\}$ are the conditions under which F's cause B's in the model. This is to be distinguished from another claim:

[2] There is a situation of type S with characteristics $\{C_1 \ldots C_n\}$, such that in this situation a certain feature F causes a certain behavior B.

In an open formula [1], unlike in an existential claim [2], there is a free variable x, which needs to be filled in before the open formula can make any claims. Once x is specified, we get a causal hypothesis of the form "an F causes a B in a situation of type S," where S is characterized by some conditions C. Without closing the open formula by specifying x, [1] only gives us a *template* or a *schema*, or a *recipe* for a causal claim, rather than a fully fledged causal claim like we see in [2].[14]

I intend [1] to be weaker than [2] in two ways. Firstly, by virtue of being an open formula [1] is not committed to existence of a situation in which the causal relations suggested by the model hold. Secondly, [1] is not committed to the particular set $\{C_1 \ldots C_n\}$ as the conditions under which F's cause B's. This is why it specifies that $\{C_1 \ldots C_n\}$ *may* be these conditions but do not *have* to be.

Each of these weakenings is crucial for the purposes of accommodating

13. I owe this amendment to Thomas Cunningham.

14. An open formula interpretation is to be distinguished from a Ramsey-sentence of a model in which the predicates are defined in terms of their relations with other predicates and their observational consequences. In an open formula, by contrast, there is no attempt to preserve this kind of structure. Hypotheses formed on the basis of an open formula of a model should be able to deviate from its Ramsey-sentence.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

the role of models in experimental economics. What exactly are the advantages?

Note, first of all, that the open formula [1] is more general than the existential claim [2] in the sense that it does not exclude the possibility expressed in [2], that is, that models can come with a specification of the situation in which the causal hypothesis is true. We may, after all, use the assumptions of the model to describe one set of conditions *C* under which *F*'s cause *B*'s.[15] But [1] does not commit us one way or the other on the question of whether or not models specify even one set of conditions under which the hypothesis is true.

Furthermore, [1] is more liberal than [2] in that it is explicit in *not* restricting the causal claims to be made on the basis of a model to those claims defined by the model's assumptions, and in *not* privileging any of those assumptions over other conditions under which these relations might hold. The models that I am discussing are often evaluated on their tractability, elegance and, of course, deductive closure. These are all reasons, over and above the statement of the empirical conditions under which the result holds, to include certain assumptions in models. So while assumptions *may* be treated as also specifying the situations in which the causal relations the model suggests hold, they do not have to be. This gives us license to go ahead and build many different causal claims on the basis of one model. This freedom is particularly important in the circumstances that often obtain in institution design:

1. **Not knowing whether an assumption is satisfied.** For example, assumptions about the statistical properties of distributions of valuations (uniformity, continuity, etc.) can often be essential to the derivation of a particular result about bidding. But auction designers who deal with actual bidders have no way of ascertaining these facts simply because companies wish to keep their valuations secret. That is when auction designers cannot use the assumptions of the model as a guide to specifying *C*'s. So they hope to find some other empirical conditions, not mentioned in the model, under which features *F* cause behaviors *B*.
2. **Knowing that an assumption is not satisfied.** Take a case in which an assumption describing a condition *C* is not satisfied by the target system. Moreover, auction designers have no control over this condition and so cannot *make* the target system satisfy this assumption.

15. This is possible if all the assumptions describe conditions that are potentially realizable in the real world. If this is not the case, then models cannot be read as correctly specifying even one situation in which a given causal relation holds. This eventuality is allowed for by the open formula view.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

> For example, most models assume Bayesian Nash Equilibrium, but flesh and blood people who represent companies seeking airwaves cannot be expected to obey this assumption. Some companies hired their own private game theorists to advise them on bidding strategy, so their bidding may be highly sophisticated. Nevertheless auction designers cannot just make bidders rational in the sense required by their models. So they had to make allowances for lack of experience, for the auction's complexity and for the fact that conformity with rational choice is just unlikely in a one-off interaction with no practice. Various other rules were added to move the bidders to behave so as to maximize the auction's efficiency. Again, some set of conditions *C*, other than the one specified by the model, needs to be found to make *F*'s cause *B*'s.

These two challenges indicate that economists often cannot rely on their models' assumptions to determine any conditions *C* under which the features *F* cause the effects *B*. If so, then it is necessary to find other knowable material conditions that researchers can control and that can realize their models' hypotheses. According to the competitor views, under the conditions described above models cannot be applied. There is clearly no possibility of assumption satisfaction or de-idealization and, plausibly, no evidence for capacity claims. The open formula conception, on the contrary, is specially geared to cases such as these.

However, such flexibility comes at a cost. On the two existing accounts if a model establishes a causal fact under particular assumptions, and if the rules these accounts set up for moving from the model to some real world situation can be used and if we follow them correctly, then we can be sure that this causal fact also holds in this situation. On the Hausman/McMullin account the successful de-idealization of a model with respect to some real world situation gives us warrant to assert that if the original model established a causal claim, then the de-idealized one does too. Similarly, on Cartwright's account if the model makes a justified capacity claim, then this capacity claim would still be true outside the model provided that we correctly identified the situation in which this capacity is exercised and no factor neutralizes it. In each case, if the original model tells us that an *F* causes a *B* in a situation *S*, then if the model applies to a situation S' according to the rules of the account, then the warrant to claim that an F' causes a B' *travels* from the model to S'. But on the open formulae view, such preservation of warrant cannot be sustained. Once we treat models as open formulae from which we can build causal hypotheses about the world, we must keep in mind that, unless we can use de-idealization or have evidence for a capacity claim, the causal hypothesis we have constructed has to be confirmed in some other way.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

There is thus a trade-off—on the open formulae view we have free reign to pick and choose what assumptions of the model will figure in our causal hypothesis. But by doing so, we commit to finding an independent confirmation of this hypothesis. This must be reflected in the account of how open formulae come to form explanations. However, I think it is a price worth paying. That is because the ways in which causal truths are preserved when we move from the model to the world on the two existing accounts are unlikely, for the reasons discussed above, to be available in many cases of theory application in special sciences.

7. From Open Formulae to Explanations. So, how do we apply open formulae? A causal explanation of a real-world phenomenon requires that we make a justified claim about a causal relation that obtains in it. On the two accounts I have discussed this is done by assessing the relationship between the causal relations in the model and those in the phenomenon in question. On the account I am proposing, the model is not assumed to supply us with any proposition at all. So we proceed differently, in three steps:

1. Identify an open formula on the basis of a model by picking from the model's premises and conclusions the *F*'s and *B*'s of interest. These presumably will correspond to the putative causes and the putative effects in the real world phenomenon in question.[16]
2. Fill in *x* so as to arrive at a closed formula or a causal hypothesis of the from "*F*'s cause *B*'s under conditions *C*" where *F*'s and *B*'s match some aspects of the target situation.

For example, if we wish to explain some art auction which uses the first-price rule and generates bids lower than expected, we may use the model in Section 2 to make a hypothesis "the first-price rule causes bids below true valuations under such and such conditions *C*" where *C*'s may or may not be described by the original model's assumptions.

3. Confirm the causal hypothesis. We do that by finding a *material realization* of the causal relation in the hypothesis. A material realization is a material environment such that if it obtains, then an *F* causes a *B*.[17]

16. What if the *B*'s in the world are not matched by the *B*'s in the model? In this case, I would argue that we do not have a model-based explanation of the phenomenon, though we may have some other explanation, which is outside the scope of this paper.

17. The realization has to be material in the sense that the causal relation in question has to be one that actually happens in the world. This does not mean that it may not be represented by a computer simulation or an equation.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

That such a material realization exists tell us that the causal hypothesis inspired by the model is true, and thus that we have a causal explanation of whatever aspect of the situation the material realizer describes. One way of specifying such a material realizer is to de-idealize the model, another is to extrapolate from a model to the material realizer on the basis of knowledge of capacities and disturbing factors present in the situation in question. But if neither method is available, we need to find the material realizer by just examining the features of the situation in which the causal relation of interest holds. The features of this environment in which the causal hypothesis is true are what allow us to fill in the causal hypothesis, that is to specify the conditions *C* under which an *F* causes a *B*. If in an art auction, for instance, we find that the first-price rule indeed causes bids below true valuation, then we examine this particular material situation to specify *C*'s. Part of this specification may match the assumptions of the original model. For example, we may find that our art auction is indeed private rather than common value or approximately so. We may also find that other assumptions of the model are *not* satisfied (say, the bidders are not perfectly rational), in which case some condition other than perfect rationality figures in our specification of *C*'s.

So it is by filling in models' open formulae that models can be used for explanation. To summarize the view proposed here:

> (**OF**). A model applies to a phenomenon (for explanation or intervention) if that phenomenon materially realizes a hypothesis based on the model's open formula.[18]

A specification of a material realization does not just blindly list every single detail about the material situation in which the causal relation we are interested in operates. We would not, for example, typically list the eye color of bidders in an art auction as part of the *C*'s. Not just because explanations are sometimes evaluated on their conciseness, but rather because we seek to include in our specification of *C*'s only the factors that are causally relevant in that one case, i.e., the factors the presence of which makes *F*'s cause *B*'s in this particular situation. In the FCC auction, many more factors turned out to be relevant than was thought at first. Even small details of the software which implemented the auction were found to be crucial for achieving the auction's economic objectives (Plott 1997). So the safe option is to be very specific in one's character-

18. There is thus a disconnect in fields such as economics between theory and the empirical successes to which it contributes. That naturally has implications for how we should understand scientific progress in such fields, discussed at length in Alexandrova and Northcott (forthcoming).

https://doi.org/10.1086/592952 Published online by Cambridge University Press

ization of *C*'s. Just what the conditions *C* must be like for *F*'s to cause *B*'s is typically subject to much debate. We test the causal relevance of *C*'s to *F*'s and *B*'s by whatever methods we normally use to test for causal relations. Of these there are a number: randomized controlled trial, natural experiment, Mill's methods and variations on them, mark methods, Bayes Nets methods, and so on. Whatever story philosophers of science tell about causal inference in various settings will help to clarify the methodology of filling in models' open formulae. But this story does not have to be part of an account of model application.

The account of models in terms of open formulae and their application in terms of material realization allows us to make sense of institution design. In particular, unlike other accounts, it explains why the use of models in the design of institutions such as the spectrum auction is a good methodology, despite the fact that neither de-idealization nor concretization was always available. Models of auction theory supplied researchers with a number of partially filled-in open formulae: "First-price auction causes bids below true valuations under conditions . . . ," "Open auction defeats the winner's curse under conditions . . . ," "Individual rather than package bids do not hinder efficient distribution under conditions . . . ," and so on. So, at the beginning, there was a wide range of "F_i cause B_j in x" claims. What researchers ended up with was an explanation of the actual auction in the form:

> A set of features $\{F_1 \ldots F_k\}$ causes a set of behaviors $\{B_1 \ldots B_m\}$ under a set of conditions $\{C_1 \ldots C_n\}$,

where *F*'s stand for features of rules, *B*'s for aspects of the auction's outcomes (i.e., its revenue generation, license distribution, speed, etc.) and *C*'s for the material conditions such as the software that auction designers could control. The *B*'s were partially specified by the government requirements. The trick was to find the combination of *F*'s and the *C*'s that would bring about the *B*'s. Some of the *F*'s, *B*'s and *C*'s figured in the models of auction theory, others came from different sources of knowledge. The experimental testbeds allowed auction designers to find one combination of rules and software that did indeed cause a speedy and efficient distribution of licenses.

8. Conclusion. While making room for the legitimate methodologies described by previous accounts, a full philosophical account of models must also be able to make sense of paradigm empirical successes such as the spectrum auction design. For many economists, if microeconomic theory is applied at all for empirical purposes, it is applied in the ever-growing instances of institution design and experimental economics. The account of models proposed here catches up with the best actual science.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

REFERENCES

Alexandrova, Anna, and Robert Northcott (forthcoming), "Progress in Economics: Lessons from the Spectrum Auctions", in Don Ross and Harold Kincaid (eds.), *Oxford Handbook of Philosophy of Economics*. Oxford: Oxford University Press.
Cartwright, Nancy (1989), *Nature's Capacities and Their Measurement*. Oxford: Oxford University Press.
——— (1995), "Reply to Eels, Humphreys, and Morrison", *Philosophy and Phenomenological Research* 55: 177–187.
——— (1999a), "Capacities", in John Davis, Wade Hands, and Uskali Mäki (eds.), *The Handbook of Economic Methodology*. Northampton: Elgar, 45–48.
——— (1999b), *The Dappled World*. Cambridge: Cambridge University Press.
——— (1999c), "Vanity of Rigour in Economics", in *Discussion Paper Series*, Centre for the Philosophy of Natural and Social Science, London School of Economics, 1–11.
——— ([2006] 2007), "Why Economic Models Are a Bad Basis for Induction to the Real World", presentation at the Philosophy of Science Association 2006 meeting in Vancouver. Reprinted in Nancy Carthwright, *Hunting Causes and Using Them*. Cambridge: Cambridge University Press, 217–235.
Crampton, Peter (1998), "The Efficiency of the FCC Spectrum Auctions", *Journal of Law and Economics* 41: 727–736.
Giere, Ronald (1988), *Explaining Science*. Chicago: University of Chicago Press.
Guala, Francesco (2005), *Methodology of Experimental Economics*. Cambridge: Cambridge University Press.
Hausman, Daniel M. (1992), *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
——— (1994), "Paul Samuelson as Dr. Frankenstein: When Idealizations Escape and Run Amuck", in Bert Hamminga and Neil de Marchi (eds.), *Idealization in Economics*. Poznan Studies in the Philosophy of the Sciences and the Humanities. Amsterdam: Rodopi, 229–243.
Humphreys, Paul (1995), "Abstract and Concrete", *Philosophy and Phenomenological Research* 55: 157–161.
Kagel, John, and Daniel Levin (1986), "The Winner's Curse Phenomenon and Public Information in Common Value Auctions", *American Economic Review* 76: 894–920.
Klemperer, Paul (1999), "Auction Theory: A Guide to the Literature", *Journal of Economic Surveys* 13: 227–286.
——— (2002a), "How (Not) to Run Auctions: The European 3G Telecom Auctions", *European Economic Review* 46: 829–845.
——— (2002b), "What Really Matters in Auction Design", *Journal of Economic Perspectives* 16: 169–189.
Maki, Uskali (1992), "On the Method of Idealization in Economics", *Poznan Studies in the Philosophy of the Sciences and the Humanities* 26: 319–354.
McMillan, John (1994), "Selling Spectrum Rights", *Journal of Economic Perspectives* 8: 145–162.
McMillan, John, Michael Rothschild, and Robert Wilson (1997), "Introduction", *Journal of Economics and Management Strategy* 6: 425–430.
McMullin, Ernan (1985), "Galilean Idealization", *Studies in History and Philosophy of Science* 16: 247–273.
Milgrom, Paul (2004), *Putting Auction Theory to Work*. Cambridge: Cambridge University Press.
Morgan, Mary S. (2002), "Model Experiments and Models in Experiments", in Lorenzo Magnani and Nancy J. Nersessian (eds.), *Model-Based Reasoning: Science, Technology, Values*. New York: Kluwer Academic, 41–58.
Morgan, Mary S., and Margaret Morrison (1999), *Models as Mediators*. Cambridge: Cambridge University Press.
Osborne, Martin (2003), *An Introduction to Game Theory*. New York: Oxford University Press.

https://doi.org/10.1086/592952 Published online by Cambridge University Press

Plott, Charles (1997), "Laboratory Experimental Testbeds: Application to the PCS Auction", *Journal of Economics and Management Strategy* 6: 605–638.
van Fraassen, Bas (1980), *The Scientific Image*. Oxford: Oxford University Press.
Weisberg, Michael (2005), "Robustness Analysis", *Philosophy of Science* 73: 730–742.

https://doi.org/10.1086/592952 Published online by Cambridge University Press