CHAPTER 11

PROGRESS IN ECONOMICS: LESSONS FROM THE SPECTRUM AUCTIONS

ANNA ALEXANDROVA
AND ROBERT NORTHCOTT

1. Introduction

The models of microeconomics are famously idealized and have a famously spotty predictive record. Yet recent years have seen some tremendous successes in using these models to construct reliable economic institutions. The focus of this chapter is one such institution, namely, spectrum auctions—now used all over Europe and North America to distribute licenses to telecommunications firms. This has been seen as a great triumph for game theory in particular. We shall explore the implications of this case for a venerable issue in philosophy of science, namely, the status of idealized models. What is the contribution of such models to empirical successes? We then use the lessons drawn from the spectrum auction to shed light, in turn, on a second venerable issue, namely, scientific progress. In particular, we explore the question, Is there progress in economics? And if so, what form does it take?

Previous analyses of these issues have often focused on theory in the abstract, or else on how models are used in academic discussions. But we believe there is much to be gained by examining the dirty details of how theory is actually applied when there is much at stake. Those details turn out to be revealing. They are especially revealing in a case of conspicuously *successful* application such as the spectrum auction, as such a case tells us much about how the gap between idealized model and messy world may actually be bridged.

Several well-developed philosophical accounts of models are now available. Auction design, meanwhile, is an instance of a larger branch of applied economics—institution design or, to use Al Roth's term, design economics. More and more, this is a principal arena for the application of microeconomic theory (Roth 2002), so it is desirable that any account of models speaks to it. Can our case study in design economics arbitrate between the different competing accounts? We believe that it can. In particular, even though the spectrum auction design is now a paradigmatic case of the use of microeconomic theory for policy making, we shall argue that none of the existing accounts can explain the role of models in it. We, therefore, propose a new account—models as open formulas—that, we claim, alone is able to do so.

The plan of the chapter is as follows: We begin in Section 2 by surveying existing work on scientific models, with an eye to the specific case of economics. We review four accounts in particular—the satisfaction-of-assumptions account, the capacities account, the credible-worlds account, and the partial-structures account. In Section 3, we tell the detailed story of the 1994 Federal Communications Commission (FCC) spectrum auction in the United States, highlighting the crucial role of experiment as well as theory. In Section 4, in the light of this case study, we present our own open-formula account of economic models.

We are then finally ready, in Section 5, to turn to the issue of economic progress. Our case study enables us to get clear on exactly what has been progressing, and on exactly what theory has—and has not—contributed to that. We shall conclude that we may *not* speak of empirical progress in economic *theory,* or at least that the success of the spectrum auction provides no warrant for doing so. Rather, progress is better seen as more akin to the worthy but piecemeal variety typical of *engineering.* This in turn has important implications for just what it is about economic theory that we should value.

2. Models in Science: A Brief Survey

The models we have in mind here are the rational choice ones of neoclassical microeconomics. These are idealized in that they posit perfectly rational agents never seen in real life, and in many other ways too. There exist a number of philosophical accounts of such models. (See Frigg and Hartmann 2006 for a general inventory of types of models in science.) The accounts we shall consider here focus on answering two questions: what sorts of claims do these models make? And how do these claims figure in the interventions (and explanations) that we might design on the basis of these models? In other words, what does the spectrum auction tell us about how, in cases of true success, economic models really work? Shortly, we shall review four such accounts.

It is important to distinguish these questions from others that philosophers of science have been concerned about, such as what kind of objects models are, how they represent, how they relate to theory, and more. Take for instance the debate about how models represent the world. Various accounts have been proposed: models have been said to represent by virtue of isomorphism (van Fraassen 1980), partial isomorphism (Da Costa & French 2003), similarity (Giere 1988), by generating inferences (Suarez 2004), or by satisfying descriptions (Frigg 2006). Interesting though the issue is, we do *not* think that a correct account of representation would answer the questions we are concerned about here. To illustrate why not, let us take a closer look at one of the accounts just mentioned—the partial isomorphism view, as recently defended by Newton Da Costa and Steven French (2003).

Partial Structures

According to the classic structuralist view, a model is a structure of the form $\{A, R_i\}_{\iota \in I}$, where A is a set of individuals in a given domain of knowledge and the R_i comprise a family of relations defined on A. In Tarski's schema, models provide interpretations of a language and we may talk about the truth of a sentence of this language by reference to a model that the sentence satisfies. Using this framework, philosophers of science have represented scientific theories by the structures or models that the linguistic formulations of those theories satisfy. Da Costa and French amend this idea in order to incorporate the incomplete and imperfect nature of scientific knowledge. To this end, they propose the notion of a *partial* structure, in which an n-place relation R_i is not necessarily defined for all n-tuples of elements in A. Then, such structures are true only of part of the domain that they model. The relationship between models and the world (the latter understood as a model of the true theory) is understood in terms of "relevant structural relationships, suitably weakened to include the more plausible similarity, rather than strict identity, and suitably broadened to cover similarities in both formal and material properties" (Da Costa & French 2003, 48). They dub such relationships partial isomorphism: One partial structure is partially isomorphic to another when certain relationships in the first structure stand in one-to-one correspondence to certain relationships of the second structure.

Da Costa and French take this framework to provide "an overarching account of models in science" (48). In their view, the notion of partial isomorphism can make sense of the use of idealization and approximation, the representative and heuristic functions of models, the autonomy of models from theories, and scientists' attitude toward models of partial or pragmatic acceptance.

Regardless of whether their account succeeds in these purposes, we do not think it succeeds in answering the questions we are interested in here. If economic models are to be represented by partial structures, then we need some account of how these partial structures are applied for explanation and intervention. More specifically, even if we grant that an auction model is partially isomorphic with reality, that fact is not very helpful in itself; such a model may still be dramatically inapplicable to the real world situations to which it is partially isomorphic, because there may be isomorphisms that do not tell us anything about the key processes involved. We need something extra to license interventions, that is, to distinguish between models that are useful and those that aren't. The critical question is whether a model is isomorphic to the part of the reality that actually matters. Da Costa and French's scheme does not tell us how to find out.

Credible Worlds

Another recent account views models as *credible worlds* (Sugden 2000, Gruene-Yanoff 2007). More particularly, and contrary to the Hausman satisfaction-of-assumptions account (see later), it denies that a model's assumptions describe the conditions under which causal relations hold in the world. Robert Sugden, in particular, worries that Hausman's logic is too restrictive because under it "we end up removing almost all empirical content from the implications of the models" (Sugden 2000, 17). Instead, economists' claims start to look like abstract theorems—*if assumption 1, assumption 2, . . . , assumption n are true, then such-and-such result follows*. But this is not how economists treat the conclusions of their models. Rather, they treat those models as making claims about the causal impact of one variable on another, holding fixed everything else. By itself, this is not a novel reading of models; what is novel is rather the justification that Sugden offers for it.

This justification lies in the notion of a credible world. A model is a sketch of a credible world to "the extent to which we can understand the relevant model as a description of how the world *could* be" (Sugden 2000, 24, original italics). Modelers construct imaginary worlds in which certain causal relations are shown to hold. To the extent that these relations cohere with our sense of how the actual world works, we come to view the world in the model as credible. A model is thus not a simplification of the existing world, but rather a parallel reality.

Unfortunately, the details of the account are unsatisfactorily underspecified. (Sugden notably does not connect it at all with the standard literature on the metaphysics of possible worlds.) Thus, for instance, how do we judge a world's credibility? Presumably, via background knowledge. But if so, judgments of credibility could not take us very far, in particular they could not take us *beyond* our background knowledge. In the case of the spectrum auction, *nobody*'s background knowledge could have told us which design would best distribute spectrum licenses. Of course, we can and do use judgments of credibility to decide which hypotheses we might admit into a pool of hypotheses to consider. But that is very far from showing that a particular piece of model-based knowledge is explanatory or is a reliable justification for policy. In the actual design of the auction, that crucial extra step required active experimental investigation, as we shall see, not mere informal judgments of credibility.

We conclude that the credible worlds view, although inspired specifically by microeconomic models and seeking to describe the conditions under which we can have confidence in them, turns out not to answer the questions that we are posing. It either fails to explain what makes models suitable bases for explanation and intervention, or else does not have that aim in the first place. However, there are other philosophical accounts of models that do clearly have that aim and that do give explicit answers regarding it. So we move on now to those.

Satisfaction of Assumptions

The first answer to these questions comes from an elaboration of the semantic view of theories defended by Ronald Giere (1988) and Daniel Hausman (1992). Hausman, applying the framework to economics, would argue that the equations in auction models, for instance, do not by themselves make empirical claims. Rather, they supply mere definitions, relating one mathematical entity to another. They only relate to the world via an additional hypothesis, namely that a real-world system satisfies some relevant class of the model's assumptions (Hausman 1992, 74–77). (A similar view of application appears to be endorsed by Morgan 2002.)

To see how this works, consider a standard auction theory model—say, one that claims that first-price auctions lead to bids lower than bidders' true valuations. To use this model to explain an actual first-price auction for, say, artwork, first the target system must satisfy the assumption that the auction is first-price. But that alone is not enough, because the model has many other assumptions, too, for instance that bidder's valuations have the right kind of statistical distribution, that bidders play according to Bayesian Nash equilibrium, that bidders are identical save for their valuations, etc. Should we, therefore, require that *all* the assumptions necessary for the derivation of the given result (i.e., bids below true valuation) be satisfied by the target system? This seems much too strict. No actual auction satisfies *all* the assumptions of a game theory model, for example, the assumption of perfect rationality. Rather, we need a criterion for distinguishing the relevant assumptions from the irrelevant ones.

One such criterion is supplied by the *de-idealization* approach (McMullin 1985). An assumption may be handled in two ways. First, the real-world system may satisfy it. Second, if it does not, then de-idealize by replacing the assumption with a more realistic one, while still preserving (to some degree) the predictions of the model relevant to explaining the target phenomenon. For example, if we have reason to believe that the real bidders are risk averse but our model assumes risk neutrality, then we add risk aversion into our model and check whether the derivation still holds. If it does, then the de-idealization process is successful and the model's result applies.

In such happy cases, 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 the strategy, however, is that, at least as far as economics is concerned, such happy cases may be hard to come by. In particular, problems arise when an assumption is not satisfied by the real-world system and yet cannot be relaxed, on pain of the model's derivation failing or on pain of us not being able to solve the model at all. For instance, in the actual spectrum auctions, bidders were not perfectly rational but all models assumed they were; hundreds of licenses were on sale, but there were hardly any multi-unit auction models at the time; models assumed no budget constraints, but real bidders most probably had those; and so on. Yet for none of these assumptions was de-idealization feasible. It was simply not possible, at least at the time, to build a model incorporating more realistic versions of the assumptions and to check the effect of these changes on the models' predictions. Indeed there simply was no *one* theoretical model capable of representing the actual auction as a whole, even at a very abstract level. This was known very well by the auction designers, who, as a result, had to use models in a more piecemeal manner, to be explained later.

Of course, the Hausman/McMullin account of model application may work better in other contexts. And, even in the case of auctions, we do know how to de-idealize *some* assumptions. But the important point for now is that de-idealization does not capture *all* that was actually going on in the spectrum auction design. A piece of the story is still missing.

Capacities

The last account is explicitly causal. On its view, models make claims about *tendencies* or *capacities,* notions proposed originally by John Stuart Mill (1843) and more recently elaborated by Nancy Cartwright (1989, 1998). For example, when we say that negatively charged bodies have the capacity to make other negatively charged bodies move away, we mean that they make others move away even in the presence of disturbing factors. Auction models, on this view, are built in order to investigate the canonical behavior of a capacity, that is, its operation in its pure form in the absence of interferences. (Cartwright 1999). (A similar reading of economic models in terms of isolations is endorsed by Maki 1992.) For example, on the basis of a model,we may conclude that a first-price auction has the capacity to lower bids under the conditions of private values.

How might such models be applied? Following Polish philosopher Leszek Nowak, Cartwright argues that this occurs via a process of *concretization*. This involves adding back the factors (i.e., other capacities and disturbing factors) omitted by the model but present in the real-world situation. Unlike the previous account, concretization does not require that a model's assumptions be satisfied by the target situation for it to explain some feature of it. Capacities are supposed to be stable enough to allow us to move from what is true in a model to what is true in the real world. Of course, the various factors we introduce during concretization must correctly describe the disturbing factors. In particular, this means that de-idealization in the sense described earlier, although admissible, is not necessary. Concretization can proceed by correcting the model in accordance with our background causal knowledge in ways other than by construction of a great big model. Indeed, at some point we should expect theoretical tools to run out. For example, low-level facts that form no part of any theory, such as how different materials react to each other, prove necessary to correct theoretical models when constructing a laser (Cartwright 1989).

As we'll see shortly, the spectrum auction design required much more than theory. Knowledge of extratheoretical practicalities proved crucial, which tells in favor of Cartwright's concretization account. But it must still be demonstrated that economic models indeed supply genuine capacity claims, and this is a tough test to pass. Capacities, at least within a certain range of circumstances, are supposed to have stability in the face of other factors. That is, they are always 'attempting' to manifest themselves even when—because of disturbing factors—they do not actually do so. (For our purposes, we may ignore metaphysical qualms over the nature of this attempting, which is rooted in Cartwright's ontology of causal powers.) Yet, again as we'll see, that was certainly not the experience of the auction designers. Instead, they found interactions between causal factors more often than not, meaning that the postulated capacities were no longer operational, that is, they were no longer even attempting to manifest themselves. Consequently, the designers believed that the stability of causes was a poor working hypothesis.

More generally, in many contexts in special sciences, such as economics and biology, the stability of causal relations is precisely what is in question and cannot simply be assumed. (For more on whether there are capacities in economics, see Reiss, forthcoming.) And yet, even in such contexts, often theory can still be successfully applied. This suggests that something more is going on than is captured by the capacities account. We return to what that might be in Section 4. But first, it is time to tell the story of the spectrum auction in more detail.

3. The Story of the Spectrum Auction

Political Background

The radio spectrum is the portion of electromagnetic spectrum between 9 kilohertz and 300 gigahertz. Spectrum not needed for governmental purposes is distributed via licenses by the FCC. For a long time, most of these licenses were awarded on the basis of hearings in which potential users had to demonstrate the public interest of their proposed enterprise. Because these had tended to become highly politicized, in the 1980s Congress authorized the use of lotteries instead. Then, in 1993, the Omnibus Budget Reconciliation Act gave the FCC the right to use competitive market mechanisms such as auctions.

When taking office in 1992, Vice-President Al Gore viewed communications policy as key to his broader objectives. These were no less than to bring about an information revolution: deregulate the communications market, and thereby jump-start innovative technologies capable of bringing about genuine social change and empowerment. For instance, in one episode, the recently appointed FCC chairman Reed Hundt was in all seriousness lecturing Gerry Adams, the leader of the Irish republican party Sinn Fein (who had come to the United States hoping to have a meeting with Gore but instead had to make do with Hundt), that the Internet and cellular phones could bring peace to Northern Ireland (Hundt 2000). Less idealistically, Gore's team also saw an opportunity to weaken the big communications firms, who traditionally had been large Republican donors, by deregulating in such a way as to let new entrants take some of the industry's market share.

By the time the auctions were finally authorized by Congress, the Clinton administration was under siege, faced with Newt Gingrich's 1994 revolution in the House. The stakes were, thus, extremely high. If the auctions failed, the administration would lose credibility with regard to its communications policy and, with it, lose its ability to resist Congress. The FCC, as well as having to rethink its whole approach to spectrum distribution, would lose its bargaining power to extract concessions from the industry. In addition to the political stakes, there was also the considerable financial stake of the potential billions of dollars that could be raised for taxpayers.

Why Auction Design Matters

The best way to appreciate the importance of good auction design is to see what happens in its absence. In the early 1990s, the New Zealand government adopted a second-price design with no reserve price (i.e., no minimum bid). The results were deeply embarrassing because the high value of the licenses was widely known and publicized and yet the government ended up earning very little. Most notoriously, an Otago university student won a license for a small-town TV station by bidding just $5, actually paying nothing since nobody else submitted a bid. In Australia, the 1993 auction for satellite television licenses was a first-price sealed-bid auction, but with no deposit and no specific payment policy. An unknown outbid the big players such as Rupert Murdoch (which the government initially gave a good spin to), only to default on the payment with no punishment whatsoever (for which no good spin was available). A series of after-auction resales followed, which delayed the introduction of paid television for nearly a year (McMillan 1994).

More recently, Switzerland offered four licenses for sale in 2000 in an ascending auction and initially attracted nine bidders. However, the weaker bidders were put off by the competition from the incumbents, and, in addition to that, the government also allowed joint-bidding agreements, which gave two companies the right to agree on which license they'd each settle without raising the price for each other—"officially-sanctioned collusion" in the words of Klemperer (2002a, 835). In the end, right before the auction, the number of bidders shrank to four, and it was looking increasingly likely that the four bidders would just pay the very low reserve prices. So the government tried to postpone the auction, only to be taken to court because it had not specified beforehand the right to cancel an auction in these circumstances. As a result, valuable licenses in one of the richest countries in Europe went for one-fifteenth of what the government had hoped (Wolfstetter 2003).

In contrast, the FCC's series of seven auctions from 1994 to 1996 were a remarkable success. They attracted many bidders, allocated several thousand licenses, and raised an inordinate amount of money—$20 billion—that surpassed all government and industry expectations (Cramton 1998). Even the first auctions went without a glitch and gave Reed Hundt the photo-opportunity of a lifetime when in front of TV cameras he presented to Bill Clinton a giant check made out to "American taxpayers" (Hundt 2000). The auctions' efficiency is harder to judge, partly because it is hard to observe bidders' valuations and hence hard to ascertain that the licenses went to those who valued them most. However, one positive sign is that similar licenses sold for approximately similar prices, suggesting that they were likely the market prices. Another is that many bidders were able, as desired, to purchase aggregations of licenses consistent with geographic synergies. Finally, there was little resale in the years following the auctions, which suggests that bidders still valued the licenses they purchased (Cramton 1998, 2006).[1] (Experimenters were also able to provide independent evidence of the auction's efficiency—see later.)

The Actual FCC Design

In the academic literature, the idea of using auctions to assign spectrum property rights dates from the 1950s (Herzel 1998, Coase 1998). The literature was revolutionized in the 1980s by a new generation of auction designs based on game theory. Nowadays, auctions have become a standard tool. Economic journals have published special issues on spectrum auctions (*Journal of Law and Economics* XLI (2) October 1998, *Journal of Economics and Management Strategy* 6, 1997), and the FCC hosts conferences in which economists discuss which designs are preferable for what purposes and what environments. The FCC Web site itself hosts records of papers and presentations on the topic by eminent scholars.

There was wide participation of economic theorists and experimentalists in the 1994 auction design. Several months in advance, the FCC solicited public comments. Many academics, hired either by prospective bidders or by the FCC itself, responded with recommendations. The constraints were set by government requirements: efficient and intensive use of the spectrum; promotion of new technologies; prevention of excessive concentration of licenses; and ensuring that some licenses go to favored bidders such as minority- and women-owned companies, small businesses, and rural telephone companies (McMillan 1994). Exactly what rules would reliably produce the desired outcome was a formidable puzzle for teams of economic theorists, experimentalists, lawyers, software engineers and policy makers.

To give a flavor of the intricacy of the final design, consider first that geographically the country was subdivided into 51 major trading areas, which in turn were subdivided into 492 basic trading areas, each of which had four spectrum blocks up for license. The auction mechanism finally selected was a simultaneous multiround auction. It put all licenses for sale *simultaneously,* as opposed to sequentially, and in an *open* rather than sealed-bid arrangement. Bidders placed bids on *individual,* as opposed to packages of, licenses they were interested in, and when a round was over, they saw what other bids had been placed. Then, the next round began, in which bidders were free to change the original combinations of licenses but had to increase their bid up to the level of the highest previous round bid plus a prescribed increment if they wished to hold on to a license. The process continued until no more bids on any license were received.

The bidding was regulated by a number of further rules:

—*Activity Rules:* a bidder had to maintain a certain level of activity during each round, on pain of reduced eligibility in subsequent rounds. However, each bidder also got five waivers that allowed them to take advantage of five opportunities not to place a bid on a particular license in a round.

—*Minimum Bid Requirement:* between rounds, the auctioneer specified the minimum bid increment (between 5 and 20 percent). The exact increment chosen depended on bidders' behavior—the more bidding, the larger the increment.

—*Designated Entities:* companies owned by women, minorities, and/or small businesses, got 10–40 percent credits on specific licenses.

—*Spectrum Cap:* no firm could own more than 45MHz of spectrum in any geographical area, hence could not bid on more than the corresponding number of licenses.

—*Payment Rules:* the FCC required an upfront payment from which penalties were deducted if the bidder withdrew, or which got refunded if the bidder failed to win the license they were bidding for.

This is a very brief summary. The full statement of the auction rules takes over 130 pages (FCC 1994b). By rules, we mean the explicit and public instructions covering entry, bidding, and payment that all participants received and studied before the auction. But, as we shall see, in addition much work also had to be put into perfecting the precise material environment, that is, features such as the software, the venue and timing of the auction, and whatever aspects of the legal and economic environment the designers could control.

The critical question for us is what justified this final complex design—why this design rather than another? It would be convenient if the methodology could be read off uncontroversially from what the actors involved said and did, but matters are not so simple. This is not surprising given that the auctions were a hugely sensitive phenomenon that put at stake the fortunes of the FCC, the Clinton administration, and the telecoms industry, as well as the reputations of many economists. Two main competing accounts emerge—roughly, those of *theorists* and those of *experimentalists*.

The Theorists' View

John McMillan and Preston McAfee were economic theorists, at the time working, respectively, at University of California–San Diego and University of Texas–Austin, who became actively involved in the FCC auction design. McMillan was hired by the FCC itself, McAfee by the potential bidder Airtouch Communications. They are also the influential authors of the first two articles dedicated wholly to the FCC spectrum auctions, which appeared in the *Journal of Economic Perspectives* in summer 1994 and then winter 1996.

What is interesting about their version of the story is its emphasis on the role played by theory in settling the major design questions, and, hence, the conclusion that game theory deserves the major credit for the auction's success. The first paragraph of their joint article "Analyzing the Airwaves Auction" reveals this attitude: "Just as the Nobel committee was recognizing game theory's role in economics by awarding the 1994 prize to John Nash, John Harsanyi and Reinhard Selten, game theory was being put to its biggest use ever." They quote *Fortune* describing the auctions as the "most dramatic example of game theory's new power…It was a triumph, not only for the FCC and the taxpayers, but also for game theory (and game theorists)." Further on, *Forbes* is quoted again: "Game theory, long an intellectual pastime, came into its own as a business tool," as is the *Wall Street Journal:* "Game theory is hot" (all citations from McAfee and McMillan 1996, 159). What exactly was hot about game theory, according to McAfee and McMillan, was that the FCC "chose an innovative form of auction over the time-tested alternatives (like a sealed-bid auction), *because* theorists predicted it would induce more competitive bidding and a better match of licenses to firms" (1996, our italics).[2]

This is not to say that McAfee and McMillan accord *all* the epistemic credit to theory. Rather, on their account, various judgment calls were also required. Some examples of particular design issues will make clear their exact position:

> *Open or Sealed-Bid?* Theory suggests that an open auction, that is, one in which all bids are public, reassures bidders that they have not overestimated a license's value because the other bids give information

about rivals' evaluations. An open auction, therefore, should reduce bidders' fear of the winner's curse—the phenomenon in which the winning bid is the one that most overestimates the value of the object for sale. Therefore, an open auction should raise more revenue. However, other theory argues against open auctions, for example, because their greater information flow also makes undesirable bidder collusion easier and, thus, revenue *lower*. Therefore, theoretical advice was ambiguous, identifying two effects but not specifying which is the strongest. Which way to go, open or sealed bid? In the end, the designers chose an open auction. McMillan explains that the decision required a judgment call rather than a neat theoretical demonstration, but does *not* explain in any detail on what basis that judgment call was made.

Simultaneous or Sequential? Should all licenses be offered simultaneously, or instead auctioned off one by one, that is, sequentially? McMillan explains that the "debate pitted theoretical virtues against practical feasibility" (1994, 153). When several items are up for auction, the usual practice is to sell them sequentially. However, in the case of spectrum licenses, aggregation is important—companies wanted clusters of licenses that would work efficiently together. If these licenses were sold sequentially, a bidder would have to make guesses about the future prices of other licenses when bidding for current ones. Such uncertainty impacts not only on revenue raised but also on whether licenses get distributed efficiently, that is, to those who value them most.

These considerations argue in favor of a simultaneous auction. However, just imagine what it would be like to have literally thousands of licenses all up for sale at the same time! Would bidders be able to track an auction of such gigantic proportions? Would they be able to process that much information simultaneously? What if a small attention slip or clerical error turned into a disaster worthy of a court battle? Finally, would such an auction ever come to a stop?

To deal with these reasonable practical concerns while still holding on to the advantages of the simultaneous mechanism, the auction designers devised stopping rules and penalties for withdrawing bids. As for concerns about complexity, McMillan tells us, auction designers judged the problem manageable and hoped for the best.

Packages or Individual? Should bids for packages of licenses, as opposed to bids for individual licenses, be allowed? One argument in favor is *complementarity.* Two or more licenses are complementary if owning one increases the value of other. For example, a company might value a Minneapolis license higher if it already owns Chicago rather than Atlanta. McMillan explains that it is easier to develop a customer base in adjacent geographical areas, to manage interferences, to establish roaming, and so on (1994, 150). The problem is that theory supports both package and

> individual bids, depending on the extent of complementarities, yet the exact extent of those was unknown. So arguments from sources other than theory must account for the adoption of one design rather than the other. In the end, package bidding was disallowed. (It is still officially considered by the FCC to be an option—it is listed as one of the two official auction designs on its Web site—but in the actual auction it was never permitted.)

Overall, when we look at the details, we can see the strength of McMillan's own admission that "theory has limits" (McMillan 1994, 151). Apart from anything else, none of the relevant models was remotely a model of the complete auction that the FCC was seeking to design. It is not clear that theory did any more than suggest possible issues for designers to take into account. This is no small contribution—but it does still leave one crucial factor unclear. Although we learn about the practical considerations and resultant judgment calls that were in play, we get no sense of what exactly *justified* those particular judgment calls that were eventually made.

The Experimentalists' View

We turn next to a very different perspective. Charles Plott, a prominent experimental economist from Caltech, worked on the FCC auctions from fall 1993 through the fall of 1994. He recounts his experience in Plott 1997. In experimental economics, laboratory environments are used to study people's decision-making processes. Many aspects of these environments, such as the subjects' characteristics, the information they receive, the rules within which they act, and so on, may be very carefully controlled. Such environments—or, as Plott calls them, *experimental test beds*—are treated as prototypes of more complex real-life economic situations. The test beds played many roles in the auction design. One was precisely to provide grounds (albeit retrospective ones) for the sort of judgment calls that McMillan talks about.

Initially, the experimental test beds were used to test broad aspects of the auction rules. Sometimes it was discovered that outcomes were very sensitive to unexpected features. For instance, in a sealed-bid auction in which the price is rising continuously and bidders drop out until only one remains, bidders had a tendency to stay in, rather than drop out, just to drive up the price for the competitor, even though this entails the risk of winning an unwanted item. The interesting fact is that the bubble created in this way is even bigger when bidders have access to information that rivals are still "in." Plott says that there "seems to be no theoretical foundation for this phenomenon, since expectations of the actions of others could cause the same behavior. Nevertheless, in experiments with the information removed such bubbles were less pronounced if they existed at all" (1997, 620, note 1).

In other words, experiment revealed something that could not have been known just from theory. Perhaps even more importantly for our purposes, Plott emphasizes that it also revealed "the sensitivity of the behavioral characteristics of the auction process to the environment in which it might be operating" (1997, 621). For instance, with regard to the bubble behavior just mentioned:

> Even if the information is not officially available as part of the organized auction, the procedures may be such that it can be inferred. For example, if all bidders are in the same room, and if exit from the auction is accompanied by a click of a key or a blink of a screen, or any number of other subtle sources of information, such bubbles might exist even when efforts are made to prevent them. The discovery of such phenomena underscores the need to study the operational details of auctions (Plott 1997, 620)

Another example of these problems was *interactive* effects. Experiments showed that the impact of any particular auction rule tended to be dependent both on which *other* rules were included, and also on the details of the implementation. Theory alone was typically unable to predict these interactive effects, even those with respect just to other rules. So at later stages of auction design, experimental test beds were crucial to overcoming the difficulty. In Plott's phrase, the test beds enabled the "development and implementation of auction technology" (1997, 627). By "technology" he does not just mean the software that implemented the auction electronically. Rather, he means the overall set of rules governing bidding, activity, stopping, and so forth, *and* the material conditions such as the software. The experiments run at this stage enabled researchers to develop one technology, that is, one set of rules plus material environment, designed to generate the outcome desired by the FCC regarding speed, efficiency, income generation, and so on.

Plott's account shows the holistic nature of the auction design process. Individual rules do not have a stable effect across different environments, so the performance of any particular *set* of rules must be tested as a package, and moreover tested anew with every significant change in environment. This process of trying out different "wholes" was accomplished by a three-stage system of testing (Plott 1997, 630–631). First, Caltech undergraduates with experimentally induced preferences participated in auctions using the actual FCC software. Second, to make sure that the success of these mock auctions was not an accident, researchers then hired the same students to look for ways to derail the auction by various devious moves within the software (for example, withdraw but then start bidding again), in a process called debugging The students were paid for keeping diaries about their experience of playing the auction. Third, researchers implemented parallel checking, in which a program was fed the data from the experimental auctions and, on the basis of this data, it performed all the computations that the FCC program was to do. This allowed them to check the accuracy of the FCC programming and to reverse engineer the system when problems were discovered. The result of all this was the perfection of one piece of technology, that is, of one material environment in which the auction rules operated more or less as desired.

Therefore, it was *not* the case that the key design decisions were first made by theorists, and that the software was then developed to implement these decisions only afterwards. Rather, the two issues had to be settled simultaneously because, as the experiments showed, they were not independent of each other. (It is true though that Plott's team did not have total freedom here, because, by the time of the later experiments, the broad aspects of the auction design—for example that it would be open and disallow package bids—were decided.)

Next, the experimenters applied their understanding in the field, that is, at the actual auction itself. Plott was a member of the increment committee created by the FCC whose purpose was to advise on possible interventions into the first auction, held in July 1994. The FCC had reserved the right to intervene, for instance by speeding up rounds. Plott's team knew that the intervals between rounds did not make much of a difference to the efficiency of the design, so the increment committee was free to vary the length of the intervals (1997, 632–633). Thus, another function of the experiments was to teach researchers some of the ways in which it would be safe to perturb the final auction design.

The experimenters' final function was to check that the actual auction was running efficiently. In the laboratory, auction preferences are induced, that is, controlled by the experimenter. Subjects each receive a piece of paper informing them how much they are willing to pay for a license. This means that, at the end of such a mock auction, the experimenters are able to check whether the final price reached is an equilibrium price, that is, whether the winner is the bidder with the highest valuation. But in real auctions, this is impossible because bidders' preferences are unknown to outsiders. Francesco Guala explains how, nevertheless, it was possible to get some indication of whether the auction was efficient (Guala 2005). Prior to the real auction, a number of laboratory auctions were run that approximated its parameters as much as possible. Those parameters included the number of items for sale, rules, valuations (or guesses thereof), the extent of complementarities (or guesses thereof), and so on. The data from these auctions, such as their duration, price movements, existence of bubbles, and such, were meticulously collected. Similar data were then collected, during the real auction, and compared to the laboratory data. Since the price trajectories achieved were very similar and since the laboratory prices were efficient, researchers were able to make an argument that the real auction prices were efficient too.

Lessons From the Spectrum Auction

How well do the four accounts of models face up to this detailed history? As noted earlier, unfortunately it seems that neither invoking a partial isomorphism between model and world, nor invoking the credibility of a model's world, sheds light on how the auction mechanism was constructed, nor on why it rather than some alternative mechanism was preferable. We shall reaffirm now why neither the satisfaction-of-assumptions nor the capacities accounts can do the job either, or at least why neither can give any more than an incomplete picture.

The theorists' and experimentalists' views emphasize different methods for giving warrant to the particular auction design adopted by the FCC. McMillan's narrative is most naturally read as endorsing the method of concretization, while Plott's, at least in part, isn't. Plott's narrative, however, is more successful at showing that the decisions that ended up being taken by auction designers were justified. So concretization does not seem to capture the true methodology of the auction design.

McMillan's history emphasizes the claim that the success of the FCC auctions consisted in the proper deployment of theoretical resources. Theoretical models allowed us to learn that open auctions encourage the flow of information, thereby reducing the winner's curse and increasing revenue; that package bids can be inefficient; that simultaneous auctions favor efficient distributions; and so on. The notion of capacity fits these claims nicely. For McMillan, theoretical results give us facts that are *stable* enough to employ for interventions, explanations, and predictions. At least sometimes, theoretical facts wear their policy implications on their sleeve. Thus the FCC auction design was successful *because* it harnessed the capacities of different design features properly.

Notice first that, contrary to the Hausman/McMullin satisfaction-of-assumptions account, this was not achieved by means of de-idealization. Assumptions in the auction models typically included, for instance, perfect rationality, no budget constraints on bidders, single units (as opposed to hundreds of spectrum licenses simultaneously on sale), and so forth. These assumptions were critical to the derivation of the models' results, yet no one had a model that yielded the consequences of relaxing these assumptions. Thus, as noted in Section 2, de-idealization was not feasible. Moreover, in the case of some other assumptions, even when a de-idealized theoretical treatment was available, the results were often discouraging, for example, proving that no competitive equilibrium is possible. All this, therefore, tells against the Hausman/McMullin account of model application being the full story in this case.

Since the language of capacities fits McMillan's own description of the theoretical arguments, turn lastly to the Cartwright/Nowak method of concretization. In the case of the auction, this would essentially amount to the process of combining together design features (i.e. auction rules) that have different capacities, in such a way as to ensure that the outcome is the one we want. This does seem to fit McMillan's history very well. For instance, simultaneous bidding, which is valued for its efficiency, also has the capacity to generate chaos by being too complex and by causing the auction to last too long. So we supplement it with various bidding and stopping rules that would prevent these effects.

Something like this story probably was the basis for constructing the mock auctions used *initially* by Plott's team. But a reason to include some rule in a preliminary test auction is not the same as a reason to include it in the final auction itself. In the first context we need only a defeasible reason, while in the second we need a solid justification. Plott reminds us again and again that experiments demonstrated interactions between different rules, in turn making it hard to draw any conclusions based only on models that understandably excluded messy real-world features, such as software details and small variations in rules. This is precisely what the experiments were for—they worked to reduce the uncertainty created by the interactions by revealing a single material environment in which the auction worked as desired. These are the experiments that supplied the needed justification for one design over another. Because of the interactions, Plott treats the material environments as *wholes* in which all elements together produce the result. No single feature of these environments on its own, such as the open design, can be said to be responsible for it.

This aspect of Plott's methodology suggests skepticism even about the capacities account. For example, while McMillan credits the open auction design with reducing the winner's curse effect, Plott remained entirely agnostic about how the winner's curse plays out when complementarities are in place (Plott 1997, 626). At the time, no model that incorporated complementarities even existed. In the experiments, more features were added specifically to defeat the winner's curse, namely, flexible minimum increments that made it more attractive to bid when activity was low, and also activity rules that required bidders to submit bids on pain of losing their eligibility. This does not demonstrate that the theoretical result was not taken seriously, but it does show that further justification was needed. Designers could not trust that the capacity of an open auction to defeat the winner's curse was stable enough to be relied on without further testing.

It would not be new to claim that theory alone cannot tell us the effects of features in any actual auction. Work on the role of models by contributors to (Morgan & Morrison 1999), by Cartwright (1983, 1989), and by a number of other philosophers of science, has already shown us how much extratheoretical knowledge is required to apply theory properly. However, Plott's methodology suggests a stronger conclusion: There was not even much knowledge of capacities at his disposal, and thus that the method of combination of causes was inapplicable. Regardless of whether auction designers did know any capacity claims, the methods they used to confirm the right set of auction rules and software were *not* that of concretization. For Plott, theoretical models were merely useful heuristics, generating categories in terms of which to start the design process. A new account of application is needed.

4. Models in Science Revisited: The Open-Formula View

Models as Open Formulas

How exactly does a model apply to a phenomenon, for example, when it yields its explanation? A common feature of the satisfaction-of-assumption and capacity accounts is that the recipe for identifying an explanatory claim involves the model *itself* in some essential way. On the former account, a model applies when some set of its assumptions are satisfied, and the explanatory claim can be obtained from the model by appropriately de-idealizing it. On the latter, a model applies if the causes it describes occur in the target situation. The associated explanatory capacity claim is thus stated by the model. But why should we assume that a model should necessarily provide such information? It might do so, but perhaps it is too restrictive to make that a requirement. We shall now propose that a model sometimes serves a different function, namely, that of a framework or heuristic for formulating causal hypotheses, but where it is those latter hypotheses—distinct from the model—that are explanatory. (For more philosophical detail about this approach than there is space to give here, see Alexandrova, forthcoming.)

In the experimental test beds of the spectrum auction, game theory models of auctions were used as suggestions for developing causal hypotheses that could then be tested by experiment. One such hypothesis might be: When values are private and some other conditions hold, first-price auction designs cause bids below true valuation. This hypothesis has the general form: Feature(s) *F*(s) cause behavior(s) *B*(s) under condition(s) *C*(s). Note that not all the model's assumptions need appear in the specification of conditions *C*(s). Rather, only those deemed salient are included. There is thus a distinction between such a causal hypothesis and the model proper. While the model plays an important role in formulating such a hypothesis, it does not fully specify it.

When a model is used in this way, we propose to conceive of it as an *open formula,* that takes the form:

1. 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*.

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.[3] It is important to distinguish this from another claim:

2. There exists a situation *S* with characteristics $\{C_1 \ldots C_n\}$, such that in this situation a certain feature *F* causes a certain behavior *B*.

An open-formula 1, unlike an existential claim 2, makes no commitment about the existence of *x*, and does not make yet any claim about any situation under which *F*s cause *B*s since the features of x are not fully specified. Rather, *x* is a free variable that 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*, 1 only gives us a template or a schema for a causal claim, rather than a fully fledged causal claim such as 2.

Advantages of the Open-Formula Account

What are the advantages of this new reading of models? First, it still allows us to pursue the old readings—when we so wish. Formally, one set of conditions *C* under which *F*s cause *B*s is just all of the model's assumptions. So if we have evidence to treat the model as a capacity claim or if we know how to de-idealize it, then we can use those methods. But we are also free to ignore some assumptions, and so in effect not to have to privilege the model's formal assumptions over other extratheoretical conditions under which the causal hypothesis might hold.

Thus we now have 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 exist in institution design. In the spectrum auction, sometimes it was simply not known whether some assumption, essential to a derivation in the model, was satisfied in the real world. For instance, many models assumed facts about the statistical properties of the distributions of bidder valuations, such as their shape, uniformity, continuity, and so on. But designers dealing with actual bidders have no way of ascertaining these facts simply because companies keep their valuations secret. In a situation like that, designers could not use the assumptions of the model as a guide to specifying *C*s. So they hoped to find some other empirical conditions, not mentioned in the model, under which features *F* cause behaviors *B*.

Moreover, sometimes auction designers had no control over a condition *C* and so could not *make* the target system satisfy this assumption. This was the case with the assumption of Bayesian Nash Equilibrium, for instance. Auction designers knew that the flesh and blood first-time spectrum bidders they had to deal with in the actual auction could not be expected to have the sort of rationality that the models assumed. So allowances had to be made for lack of experience, for the fact that the auction was complex, and so on. In particular, new rules were added purely to push the bidders to behave as required to ensure efficiency. Formally, again, some set of conditions *C*, other than the one specified by the model, needed to be found in order to make *F*s cause *B*s in the real world.

Moreover, generally microeconomic models are often evaluated on their tractability, elegance, and, of course, deductive closure. But whatever their merits for other purposes, we see now that such criteria do not necessarily give us any assurance that a model explains and in fact may get in the way of explaining. They are thus all reasons that certain assumptions are included over and above the reason of merely stating the empirical conditions under which a result holds. In turn, this is why it often proves useful *not* to include such assumptions when formulating causal hypotheses about the real world.

These challenges indicate that we often cannot rely on models alone to tell us the vital conditions *C*. If so, then we need some other way. The open-formula view is set up precisely to accommodate the necessary contributions from beyond theory. However, such flexibility does not come free. For the older accounts, at least we have explicit procedures by which to guarantee that a causal claim derived from a model holds in the real world. More particularly, on the Giere/Hausman satisfaction-of-assumptions view, a successful de-idealization gives us warrant to assert that if the original model established a causal claim, then the de-idealized one does, too.

Similarly, on the Mill/Cartwright capacity view, if a model makes a justified capacity claim, then that claim will still be true outside the model, provided that we have concretized successfully. In each case, if the original model tells us that an *F* causes a *B* in a situation *S*, then if the model also applies to a situation *S** according to the rules of the account, then the warrant to claim that an *F** causes a *B** automatically *travels* from the model to *S**. But on our open- formula view, such preservation of warrant cannot be sustained. Once we treat models merely as open formulas, the causal hypotheses we construct from them have to be confirmed in some other way.

There is, thus, a trade-off—the open-formula view gives us free reign to pick and choose the assumptions from the model that will figure in the final causal hypotheses we end up with, but the price is that we are left still needing a way to confirm those hypotheses. Nothing about the procedure of their formation guarantees that they will in fact hold; rather, that must be established anew each time. We think that unfortunately this price is in any case unavoidable because neither of the alternatives—de-idealization and concretization—is likely, for the reasons discussed earlier, to be available in many cases. Nevertheless, it does mean that we must now address how open formulas may be applied.

From Open Formulas to Explanations

A causal explanation of a real-world phenomenon requires that we make a true and justified claim about a real-world causal relation. With the open-formula account, we achieve this via a three-stage procedure:

Stage 1: Construct an open formula by picking from a 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 at work in the real-world phenomenon of interest.

Stage 2: Fill in *x* so as to arrive at a closed formula, that is, a causal hypothesis of the form "*F*s cause *B*s under conditions *C*" where *F*s and *B*s match some aspects of the target situation. The conditions *C* may or may not be described by the original model's assumptions.

Stage 3: Confirm the causal hypothesis. We do that by finding a *material realization* of it, that is, a material environment in which an *F* indeed causes a *B*.

Stage 3 is obviously the difficult bit. That is why, in the case of the spectrum auction, so much work was needed in the experimental test beds. When we do achieve a material realization then we are entitled to say that the causal hypothesis inspired by the model is true, and, thus, that we have a causal explanation of *B*. The features of the environment in which the causal hypothesis is true are what allow us to fill in the open formula, that is to specify fully the conditions *C* under which an *F* causes a *B*. Part of this specification may match the assumptions of the original model, but part of it may not.

How are material realizations specified? Answer: in the same way that we normally specify a claim about one set of sufficient causes of a phenomenon. (In the terminology of J.L. Mackie's INUS conditions, a material realization for a causal relation can be understood as one of the disjuncts of the overall INUS formula.) Thus, we do not just blindly list every feature of the material situation, such as the color of the bidders' eyes and so on. Rather, we only specify those conditions that are causally relevant.

Exactly what the conditions *C* must be like for *F*s to cause *B*s can be tricky to establish. We may *test* the causal relevance of *C*s to *F*s and *B*s by drawing as convenient from the usual repertoire of methods—controlled trial, natural experiment, Mill's methods and variations on them, mark methods, Bayes Nets, and so forth. Whatever story philosophers of science and methodologists tell about causal inference generally, will presumably help to clarify how we may fill in models' open formulas.

This account of models, and of their application in terms of material realization, enables us to make sense of institution design. In particular, unlike the alternative accounts, it explains why the use of models in the design of the spectrum auction was justified, even though neither de-idealization nor extrapolation on the basis of capacities was always available. According to this view, the models of auction theory supplied researchers with a number of partially filled-in open formulas: "First-price auctions cause bids below true valuations under conditions…," "Open auctions defeat the winner's curse under conditions…," "Individual rather than package bids do not hinder efficient distribution under conditions…" and so forth. So at the beginning, there was a wide range of "F_i cause B_j in x" claims. After much work, what researchers ended up with was an explanation of the actual auction in the form:

3. 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 that the designers could control. The *B*s were partially specified by the government. The trick was then to find the combination of *F*s and *C*s that would bring about those *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. Crucially, it was the experimental test beds that allowed auction designers to find one combination of rules and software such that when this combination was instantiated, the particular kind of open auction selected did indeed cause a speedy and efficient distribution of licenses.

5. Progress in Economics

At last, we are ready to turn to the question of *progress* in economics. Without the development of modern auction theory, the successful FCC auction design would not have been possible. In addition, there has also accumulated crucial practical know-how, so that the designers of the United States spectrum auction did not repeat the mistakes of their New Zealand and Australian predecessors, for instance. This suggests two distinct senses in which the successful FCC auction reveals progress in economics—first, purely theoretical development; and second, engineering development in how theory might be applied. Can both of these senses be sustained? The details of the auction case study now turn out to be very revealing. In particular, we shall argue, they endorse progress strictly only of the second engineering kind, in a manner to be explained. This has implications in turn for the role played in any progress by theoretical development, and, therefore, for just what it is about economic theory that we should value.

What Economic Progress is not

In order eventually to see the import of the spectrum auction case here, begin first with a brief detour into the literature on scientific progress generally. This has traditionally analyzed progress in terms of scientific theories. In particular, from the beginning it has been motivated primarily by the debate between scientific realism and antirealism. Because it is accepted that most, if not all, of our best theories are not literally true, a satisfactory account of approximate truth (or 'verisimilitude') has been seen as important, perhaps even essential, to the realist position (Putnam 1975; Newton-Smith 1981; Miller 1987; Boyd 1990; Psillos 1999). Scientific progress has then been seen in terms of successive theories achieving closer and closer approximations to the truth—convergent realism. For example, the sequence of Aristotelean, Newtonian, and finally relativistic mechanics is seen as one such convergence.

It has proven notoriously difficult to flesh out satisfactorily the needed notion of verisimilitude. One classic problem is language dependence: For any two false theories, any ranking of them by verisimilitude can be reversed simply by changing the choice of variables on which that ranking is defined (Miller 1974; Miller 1975). The problem is especially devastating because even were we to postulate a privileged one-true ontology—perhaps the natural kinds as revealed by an ideal science for instance—still that ontology would underdetermine choice of variables. For this and other reasons, it is currently dubious whether any authoritative sense can be made of a false theory's closeness to the truth. In turn, this has motivated some to be skeptical of global scientific progress at all (Laudan 1984).

This emphasis on realism and theories reflects the literature's concentration on fundamental physics and chemistry rather than the special sciences. For example, general relativity and quantum mechanics are plausibly taken to be literal descriptions of the world, and thus as potential candidates for being true and for being interpreted realistically. Or at least, so many realists have believed. Other positions, too, such as Van Fraassen's (1980) constructive empiricism, agree that such theories should be interpreted as literal descriptions. Theories in economics, by contrast, are agreed by all to be extremely idealized. Nobody thinks that *Homo economicus* is a literally true description of anyone. So the idea that economic theories could ever be literally true is clearly a nonstarter.

Nevertheless, this still seems to leave open a different sense in which economic theory could be progressing, namely, that it is becoming better at capturing particular aspects of reality. In particular, the thought runs, although auction models, for instance, are idealized descriptions and, thus, are not literally true, nevertheless they do successfully capture real patterns of strategic interaction between bidders. They are, thus, able to track regularities between, say, auction format and size of bids.[4] As we saw earlier, there are various ways in which more flesh has been put on this rather vague initial thought—by viewing economic theory as capturing causal capacities, as idealizations requiring only that their assumptions hold, as positing credible worlds, or as positing partial structures, and so on. We have argued that, in any case, the evidence of the spectrum auction does not support any of these interpretations. But bracketing that conclusion temporarily, how might those interpretations make sense of economic progress?

Unfortunately, a major problem remains even with this more modest conception of progress, in which theory becomes better merely at capturing particular aspects of reality. The problem stems ultimately from the continuing emphasis on theory in isolation. Like all idealizations, economic theory is good at capturing some things but less good at capturing others. For example, auction models capture patterns of strategic interaction between bidders but make no allowance for individual variations in agent psychology; for local details, such as software glitches or physical arrangement of bidders; and so on. How much do these omissions matter? An answer is surely vital to any assessment of how valuable a model is, and hence to whether that model represents progress relative to other models that capture other things. Yet the importance of any given omission will inevitably depend on the particular application in question. In some contexts the omission of variation in agent psychology, for instance, will matter more than in others. The point is that there is no univocal *general* answer about the seriousness of a given omission, independent of the local details. It follows that no context-general sense of theoretical progress can be sustained, for a model may simultaneously have progressed with respect to one application but not to another. An overall score for that model's progress could, therefore, only ever be a crude average of its scores over various particular applications, and would, therefore, carry no more weight than those applications' selection criteria.

For instance, on the capacities account, the importance of the capacities a model captures relative to those it omits will vary, case by case. The problem equally infects the other accounts, too. The degree of ease with which assumptions can be de-idealized, the degree of credibility of a world, and the degree of importance of the portion of the total domain that a partial structure captures—all of these, even assuming rigorous sense can be made of them, also seem bound to vary, application by application. The lesson is that, even when speaking of our more modest conception of it, we can only make sense of economic progress context-specifically. We can, therefore, *not* make sense of it for any theoretical model in isolation, such as a particular auction model, say. Rather, we can, at best, speak of that auction model representing progress with respect to a specific application.

Next, finally, turn to the new twist put on these matters by analysis of the spectrum auctions. Alas, the lesson of that analysis is that to speak even of application-specific progress is to be too optimistic. For, as we have seen, in the case of the spectrum auction at least, empirical warrant does *not* accrue to the theoretical model. The only confirmed causal claim is with respect to the mechanism as a whole (that is, claim 3). *That* causal claim is made only by the mechanism, of course, not by any of the auction-theory models. This is why the theoretical models garner no support here. Rather, support accrues only to the final mechanism used in the actual auction, and, thus, it is only that final mechanism that has a claim to be close to the truth. Because that mechanism was the result of extensive experimental tweaking independent of the theoretical model, it is a mistake to think that the success of the auction implies any progress toward the truth on the part of *theory,* even application-specifically. Thus, the contrasting experiences of the New Zealand and United States auctions, for instance, speaks to progress with respect to final mechanisms but not with respect to theoretical models. (Indeed, much the same theoretical repertoire was available in both cases.)

To sum up so far: The usual notion of scientific progress in the literature has been of successive theories approaching closer to the truth. This notion has proved problematic in general, for instance, because of language dependence. Even if we focus only on an agreed economic vocabulary, still economic theory is recognized by all to be highly idealized and, therefore, not a candidate for literal truth. A weaker claim is that economic theory nevertheless captures aspects or parts of the truth. But then we need to judge the relative importance of those aspects it captures and those it doesn't. Any such judgment must inevitably be application-specific, implying that theoretical models in isolation cannot be said to progress; rather, at best, they can be said to progress in capturing the truth about specific situations. Finally, the spectrum auction case suggests that even this last claim is not supportable, because the empirical warrant from successful applications in fact accrues only to final concrete mechanisms that cannot be derived directly from theoretical models. If we interpret progress in the traditional way as increasing closeness to truth, then, we have no reason to believe that economic theory is progressing.

Some Red Herrings

We shall offer a positive view, and address its implications for economic practice, shortly. Before that though, a little more negative work is required, in particular to dispose of some red herrings. First, a common fall-back defense of economic theory is that even if empirical warrant is elusive, still it provides 'insight' or at least is 'suggestive'. For instance, an auction model illustrates how the winner's curse comes about, or could come about, and thus increases our 'understanding' of that phenomenon. How are we to interpret the quoted terms? Can they carry any philosophical weight?

The most likely candidate for arguing so seems to be the view that theory provides *explanation*. In other words, auction theory explains why the winner's curse comes about, for instance. But on any standard account of explanation, for a theory to do this it must have empirical warrant, and, given the essential role of the experimental test beds, that is just what close analysis of the spectrum auctions casts into doubt. Thus, an auction model does not, in conjunction with initial conditions, entail the successful auction outcome, as would be required on the deductive-nomological view. Neither does it provide the cause or mechanism that produced that outcome, as only the final composite auction procedure achieves that. Neither finally does the theoretical model unify various phenomena, since on our view it does not on its own account for any phenomenon at all, let alone several different phenomena. The necessary empirical warrant is just what the theoretical models did *not* accrue in the case of the spectrum auction. Accordingly, we have no good reason to declare them explanatory of the auction's success. In lieu of some alternative interpretation (on which see later), therefore, and with apologies to Hume, in our view, vague talk of such models providing insight or understanding can contain nothing but sophistry and illusion.

Next, and more generally, it is easy just to declare after the fact that some real-world observation was explained by an economic model. The spectrum auction was different in that, it being a case of institution design, scientists faced the constraint of eventually having to actually run an auction and make it a success. This was, of course, a very different challenge to mere after-the-fact rationalization. And, although it might have looked casually as if, as the contemporary publicity claimed, the theoretical auction models indeed explained the outcome, as we have seen, the reality was more complicated. Judging by the example of the spectrum auction, therefore, there is ample reason to be cautious about attributions—after the fact and from a distance—of explanatory success to economic theory. When push came to shove and an actual working mechanism was required, it turned out not to be so.

A Positive Story—Economists as Engineers

We have seen the difficulty of making sense of any variety of purely theoretical progress. That was the negative story; what is the positive one? For that, consider, again, the second option mentioned earlier, namely, that of understanding progress as progress in engineering know-how, that is, as progress in the practicalities of *applying* economic theory. We believe this turns out to be a much better way to view the matter.[5] Moreover, such a view of economic progress, as well as capturing the success of the spectrum auctions, also turns out to have normative implications for the development of economic theory more widely. (Of course, we do not mean to deny progress also in other applied economic tasks, such as measuring new data.)

A helpful analogy here is with the development of *racing cars,* for instance the work of a Formula One team. The designers in such a team are faced with the challenge of maximizing a car's speed and reliability while constrained by regulations concerning weight, tires, engine size, fuel capacity, and a host of other details. These regulations typically change each championship season. There is of course a lot of relevant theory to know and, for this reason, senior designers are highly trained engineers. Yet theoretical knowledge alone is not enough. All teams also have huge testing programs, analogous to the experimental test beds of the spectrum auction. For example, new chassis designs are tested extensively in wind tunnels, while in-house drivers take each new model for vast numbers of timed laps on private tracks. Engineers on all teams are highly educated and presumably have very similar levels of theoretical knowledge. Yet, of course, the final results of their efforts—that is, each team's cars in the races—are far from very similar. On the contrary, some are much more successful than others. And typically it is the teams with the biggest development and testing budgets that end up producing the winning cars. That is, within any one season, the quality of racing cars correlates with teams' development budgets rather than with their levels of theoretical knowledge.

Of course, theoretical knowledge is essential, too. No matter what the testing budget, presumably no team would be competitive without knowledge of Newtonian mechanics, gas laws, the chemistry of fuel combustion, materials science knowledge of rubber and lightweight composites, and so on. But it does suggest that, just as the experimental test beds were essential to producing a successful spectrum auction design, so are the wind tunnels and practice laps essential to producing a successful racing car design. In both cases, abstract theory is necessary but not sufficient.

What does this tell us about progress? Within any one season, when regulations are constant, there exists progress in racing car design in that the cars become faster and more reliable. Within that time span, the theoretical knowledge being employed presumably does not change significantly. What does improve, rather, is the ability to apply it effectively. That is, the development of better racing cars through the season is due to the accumulation of new context-specific engineering know-how. It is *not* due to any global nearing to the truth of our underlying theoretical picture of the world.

The context-specificity of such progress is reflected by its limited *exportability* to new contexts. The know-how that a racing-car team accumulates through a season, just like the practical know-how that went into the spectrum auction design, is only exportable to a limited extent, or at least is not as exportable as the relevant theoretical knowledge. The success of one season's car does not guarantee success under new regulations in the next season, as the sporting record has demonstrated many times. Similarly, the successful design of a spectrum auction in the conditions and regulations of one country does not necessarily carry over to those of a new country (Klemperer 2002b). In both cases, the context has changed, and, thus, a new round of practical know-how must be acquired. For this reason, just as the winning racing-car team often changes from season to season, so a successful auction design in one country will not necessarily be successful in a different country. Thus, the spectrum auction in Switzerland was a comparative failure, despite being run six years after the successful FCC one. Of course, these are matters of degree, and there is exportability to some extent—thus racing-car teams learn from previous designs, just as U.K. spectrum-auction designers in 2000 learned from their U.S. predecessors. But the point is the asymmetry between on one hand context-specific practical know-how with limited exportability, and on the other hand context-general theoretical knowledge with great exportability.

Progress and Theory Revisited

In sharp contrast to the kinds of theoretical progress discussed earlier, this context-specific engineering progress is amply endorsed empirically. Thus, racing cars demonstrably do run faster and more reliably through the course of a season, just as the U.S. spectrum auction was demonstrably more successful than its predecessors. Nevertheless, as repeatedly emphasized, a necessary condition for engineering success is relevant theory, namely, in the case of the spectrum auctions, the new game-theory auction models. So we have an apparent paradox: Although there is no warrant for claiming theoretical progress in the sense of empirical success or increased closeness to the truth, still theoretical progress in *some* sense is necessary for the context-specific engineering progress for which there *is* warrant.

So what might this other, different kind of theoretical progress amount to? We must be careful to avoid yet another, final, red herring. In particular, one common thought is to view this different kind of theoretical progress as akin to progress in pure mathematics, a kind of internal progress of new concepts and categories, proofs and refinements, and so forth. For example, modern theory has seen the development of such concepts as multiple equilibria and asymmetric information effects, and in turn how these effects interact with other factors such as different auction procedures. No doubt, this is progress in some sense—but not progress for which the spectrum auction provides any empirical warrant. And without a clear connection to empirical progress, it is not clear why, as scientists (as opposed to pure mathematicians), we should value it. How is it to be measured, for instance? Was asymmetric information a more important conceptual breakthrough than new notions of equilibrium? What criterion could we use to decide – an aesthetic one? Any criterion encompassing empirical success would seem inevitably to run into the same difficulties that we discussed earlier, such as that the worth of these breakthroughs would be highly application-specific and dependent in any case on practical know-how, independent of theory.

We think the underlying problem here is again the attempt to define progress with respect to theory in isolation. To repeat: If the example of the spectrum auction is typical, then the only notion of progress in economics that is empirically warranted is of the context-specific engineering variety. Thus, the criterion for judging a new piece of theory should be: How much does it help economic engineers achieve empirical success? An expanding and dazzling theoretical repertoire of new methods, heuristics, and categories is, alas, of no empirical value *in itself*. Rather, it can only be valuable instrumentally. Nevertheless, this does at least finally yield us an indirect kind of theoretical progress, namely, that new theoretical development is progressive insofar as it generates useful new aids to the engineers.

The hard-nosed corollary is that theoretical developments that offer little prospect of being useable by economic engineers are of correspondingly little scientific value.[6] Of course, no one can ever know for sure in advance exactly which pieces of theory may eventually prove useful in this way. Readers will form their own judgment of the prospects for any particular example. Notice, however, that the issue is *not* the familiar one of pure versus applied research, wherein the former is damned by philistines as being perhaps noble in its pursuit of scientific truth but, alas, of no practical use. Rather, in the case of economics, the philistine argument is altogether stronger, namely, that the pure research cannot even be said to be pursuing scientific truth.

The design of the FCC spectrum auction is worthy of celebration. It was a tremendously sophisticated (and lucrative) piece of technology—on a par with, say, a laser or a jet airplane. Economic theory can be correspondingly proud of its contribution to this success. But the sting in the tail from the story is this: that it is only through such contributions that theory partakes of scientific progress. Its only glory is via its utility to those working elsewhere.

NOTES

Authors' names listed in alphabetical order. The authors are equally and jointly responsible.

1. For an opposing view, see Nik-Khah (2008). He argues that in fact the auctions' only success was large revenues. Other objectives, such as decentralization of ownership and promotion of wireless technology in rural areas, were not met. Cramton 1998 and 2006 mostly disagrees. For our purposes, it is enough to say that the auctions were a more successful piece of technology than previous alternatives (such as the auction designs used in other countries, or nonmarket mechanisms).

2. Notice also the historical origin of McMillan's protheorist paper. McMillan was originally hired by the FCC as an independent advisor—independent in the sense that he was not at the same time employed by a potential bidder. His task was to help the

FCC arbitrate between the conflicting advice received from game theorists employed by the different telecoms firms. As we shall see, game theory did not give a univocal answer to the question of which auction design would maximize efficiency. (It was through this hole that the telecoms companies were able to present their suggestions as being justified by considerations of efficiency when, in fact, the motivation was self-interest—Nik-Khah 2008.) In an attempt to explain the FCC's eventual decisions, McMillan produced a report that synthesized the debates that were taking place between the theorists (FCC 1994a). This report eventually turned into his 1994 paper. The primary purpose of the report might, thus, not have been to show what exactly theory can and cannot do for auction designers, but rather only to show that the FCC was aware of the conflict between different theoretical considerations and took that into account. So to this extent, McMillan's paper may, by political necessity, have been in part a public relations exercise.

3. Strictly speaking, we should not speak of *F*s causing *B*s "in the model." The causal dependence of *B*s on *F*s cannot simply be read off the deductive relations between the assumptions describing *F*s and *B*s. Rather, it is an interpretation given to the model using some background knowledge.

4. It may be disputed whether scientific progress is best viewed as progress toward truth or, instead, as progress only toward some lesser goal, such as better predictions, interventions, or empirical descriptions. None of our arguments turn on this general issue. We shall be emphasizing empirical success, but presumably such success would, in any case, be the best reason to believe in the metaphysically stronger claims of a model's truth or approximate truth.

5. Ken Binmore, chief designer of the extremely successful UK spectrum auctions of 1999 and 2000, sees his achievement as being akin to engineering in exactly this way (personal communication, one of the authors).

6. As mentioned in footnote 4, this remains true whether we understand *scientific value* to mean "closeness to the truth" or mere "empirical success". We are committed though to *scientific value*, implying something beyond mere mathematics-type, purely abstract theoretical development. Similar remarks apply to *scientific truth* and *scientific progress* below.

REFERENCES

Alexandrova, A. (forthcoming). "Making Models Count". *Philosophy of Science*.

Boyd, R. (1990). "Realism, Approximate Truth, and Philosophical Method." In C. Savage, Ed., *Scientific Theories,* 355–391. Minneapolis: University of Minnesota Press.

Cartwright, N. (1983). *How The Laws of Physics Lie*. Oxford: Oxford University Press.

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

Cartwright, N. (1998). "Capacities." In J.B. Davis, D.W. Hands, & U. Mäki, Eds., *The Handbook of Economic Methodology,* 45–48. Cheltenham: Edward Elgar Publishing.

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

Coase, R. (1998). "Comment on Thomas W. Hazlett: Assigning Property Rights to Radio Spectrum Users: Why did FCC license auctions take 67 years?" *Journal of Law and Economics* XLI: 577–580.

Cramton, P.(1998). "The Efficiency of the FCC Spectrum Auctions." *Journal of Law and Economics* XLI: 727–736.

Cramton, P. (2006). "Simultaneous Ascending Auctions," In P. Cramton, Y. Shoham & R. Steinberg, Eds., *Combinatorial Auctions* 99–114. Cambridge, MA: MIT Press.

Da Costa, N. & French,S. (2003). *Science and Partial Truth: A Unitary Approach to Models and Reasoning in Science*. New York: Oxford University Press.

FCC (1994a). "Second Report and Order," 94–61, Washington, D.C.

FCC (1994b). "Fifth Report and Order," 94–178, Washington, D.C.

Frigg, R. (2006) "Scientific Representation and the Semantic View of Theories." *Theoria* 55: 49–65.

Frigg, R. & Hartmann, S. (2006). "Models in Science." *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/models-science/

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

Gruene-Yanoff, T. (2007). "Learning from Economic Models." Manuscript, Royal Institute of Technology, Sweden.

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

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

Herzel, L. (1998). "My 1951 Color Television Article." *Journal of Law and Economics* XLI: 523–527.

Hundt, R. (2000). *You Say You Want a Revolution*. New Haven: Yale University Press.

Klemperer, P. (2002a). "How (Not) to Run Auctions: The European 3G Telecom Auctions." *European Economic Review* 46: 829–845.

Klemperer, P. (2002b). "What Really Matters in Auction Design." *Journal of Economic Literature* 16 (1): 169–189.

Laudan, L. (1984). *Science and Values*. Berkeley: University of California Press.

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

McAfee, P. & McMillan, J. (1996). "Analyzing the Airwaves Auction." *Journal of Economic Perspectives* 10 (1): 159–175.

McMillan, J. (1994). "Selling Spectrum Rights." *Journal of Economic Perspectives* 8 (3): 145–162.

McMullin, E. (1985). "Galilean Idealization." *Studies in History and Philosophy of Science* 16 (3): 247–273.

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

Miller, D. (1974). "Popper's Qualitative Theory of Verisimilitude." *British Journal for the Philosophy of Science* 25: 166–177.

Miller, D. (1975). "The Accuracy of Predictions." *Synthese* 30: 159–191.

Miller, R. (1987). *Fact and Method*. Princeton: Princeton University Press.

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

Morgan, M, & Morrison, M, Eds. (1999). *Models as Mediators*. Cambridge, England: Cambridge University Press.

Newton-Smith, B. (1981). *The Rationality of Science*. Boston, MA & London: Routledge & Kegan Paul.

Nik-Khah, E. (2008). "A Tale of Two Auctions." *Journal of Institutional Economics* 4 (1): 73–97.

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

Psillos, S. (1999). *Scientific Realism*. London: Routledge.

Putnam, H. (1975). "How Not to Talk about Meaning." In *Philosophical Papers, Vol. 1, Mathematics, Matter and Method*. Cambridge, England: Cambridge University Press. 250–269.

Reiss, J. (forthcoming). "Social Capacities." In Stephan Hartmann & Luc Bovens, Eds., *Nancy Cartwright's Philosophy of Science*. London: Routledge.

Roth, A. (2002). "The Economist as Engineer: Game Theory, Experimental Economics and Computation as Tools of Design Economics." *Econometrica* 70 (4): 1341–1378.

Suárez, M. (2004), 'An Inferential Conception of Scientific Representation.' *Philosophy of Science* 71 S: 767–777.

Sugden, R.(2000). "Credible Worlds: The Status of Theoretical Models in Economics." *Journal of Economic Methodology* 7 (1): 1–31.

Van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.

Wolfstetter, E. (2003). "The Swiss UMTS Spectrum Auction Flop: Bad Luck or Bad Design?" In H-G. Nutzinger, Ed., *Regulation, Competition, and the Market Economy. Festschrift for C.C.v.Weizsäcker,* 281–294. Göttingen: Vandenhoeck & Ruprecht.