Title
Connecting models to the real world : game theory in action

Author
Alexandrova, Anna

Publication Date
2006

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Connecting Models To The Real World: Game Theory in Action

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Philosophy

by

Anna Alexandrova

Committee in charge:

Professor Nancy Cartwright, Chair
Professor Richard Arneson
Professor Craig Callender
Professor Vincent Crawford
Professor Naomi Oreskes

2006

The dissertation of Anna Alexandrova is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2006

DEDICATION

To all my teachers

TABLE OF CONTENTS

VITA

Education:

1994-1995 Kuban State University (Krasnodar, Russia), Linguistics and Philology

1995-1998 Intercollege (Nicosia, Cyprus), BSc Communications, Summa Cum Laude

1998-1999 London School of Economics (UK), MSc Philosophy of Social Sciences

2000-2003 University of California, San Diego, MA Philosophy

2006 University of California, San Diego, PhD Philosophy

Publications:

"Connecting Rational Choice Models to the Real World" forthcoming in Philosophy of The Social Sciences

"Subjective Well-Being and Kahneman's 'Objective Happiness'" forthcoming in Journal of Happiness Studies

"Philosophy of Science: Laws" (with N. Cartwright), forthcoming in The Oxford Handbook of Contemporary Philosophy, eds. M. Smith and F. Jackson, OUP

"Probabilistic Causality, Selection Bias and the Logic of Democratic Peace" (with B. Slantchev and E. Gartzke) forthcoming in American Political Science Review

Review of Max Steuer's 'Scientific Study of Society' (with R. Northcott), Economics and Philosophy 20.2, October 2004, 375-381

Teaching and Research Appointments:

2006 March – June, Philosophy Department Lecturer, UCSD

2002 – 2004, Philosophy Department Teaching Assistant, UCSD

Summer 2002, 2003 and 2004, Research Assistant to Prof. Nancy Cartwright, UCSD/LSE

2000 – 2002, Teaching Assistant for Making of the Modern World (MMW), Writing Program of Eleanor Roosevelt College, UCSD

1999 – 2000, Teaching Assistant Intercollege, Nicosia, Cyprus

1997 – 1998, Research Assistant to the head of Communications Department (Prof. Nayia Roussou), Intercollege, Nicosia, Cyprus.

Selected Awards and Scholarships:

National Science Foundation Dissertation Improvement Grant, 2004-2005

Dissertation Fellowship, Center For Humanities, UCSD, Winter 2005

Philosophy TA of the Year, UCSD, 2003 Inaugural award

Global Supplementary Grant Program Scholarship, Open Society Institute, 2000-2003

ABSTRACT OF THE DISSERTATION

Connecting Models To The Real World: Game Theory in Action

by

Anna Alexandrova

Doctor of Philosophy in Philosophy

University of California, San Diego, 2006

Professor Nancy Cartwright, Chair

How are empirical successes based on idealized models such as those in economics and other special sciences possible? And what does the manner in which those successes are achieved tell us about the models' role and status? Existing answers to these questions, I argue, are unable to make sense of a prominent and typical case of application of game theory – the design of an institution. This dissertation seeks to bring philosophical accounts of models and their application in line with this practice.

In chapter 1 I examine accounts currently on offer. The first account views models as making *ceteris paribus* claims and their application as satisfaction of the relevant assumptions of a model by the target system. I review two ways of identifying the relevant assumptions – by prediction and by robustness analysis – and find them both wanting. The second is the *credibility* account which I claim does not tell us how models become bases for explanation and prediction. On the third account models make tendency claims and they apply to a phenomenon if the model plus additions to it identify the relevant tendencies and disturbing factors underlying the phenomenon.

However, this account assumes knowledge of stable causal relations, which is often unjustified in special sciences.

Chapter 2 presents a study of a design of a social institution on the basis of game theory – the FCC spectrum auctions. I describe the respective roles of theory and experiment in this case, and argue that none of the accounts reviewed in chapter 1 can make sense of this example.

Chapter 3 develops an account of theory application that fits the auction design case and is flexible enough also to capture the use of theory in special sciences in general. On this view, models should be treated not as ceteris paribus or capacity claims but as *open formulae* for causal hypotheses, which we apply by specifying an empirical realization of the causal relation in the hypothesis. This account is in line with the practice of experimental economics, and provides normative constraints for the use of rational choice models for historical explanation

Introduction

Economics is known as a science of models. How are empirical successes based on economic models possible? And what does the manner in which those successes are achieved tell us about the models' role and status? Existing answers to these questions, I argue, are unable to make sense of the most prominent and, arguably, typical case of application of game theory – the design of an institution. This dissertation seeks to bring philosophical accounts of models and their application in line with this practice. Although economics is the main example, the results are generalizable across all social and biological sciences that use idealized deductive models with a causal interpretation.

What is the role of such models in cases when economics licenses successful explanations, predictions and the design of reliable institutions? To answer this question we need an account of the claims that models establish and an explanation of how these claims then figure in explanations, predictions etc. In chapter 1 I examine three such accounts currently on offer. The first account views models as making *ceteris paribus* claims and their application as satisfaction of the relevant assumptions of a model by the target system. What assumptions are relevant? I review two ways of identifying them – by prediction and by robustness analysis – and find them both wanting. The second account of application is the *credibility* account – economic models make claims about causal relations that apply when they pass a credibility test. I argue that this account does not tell us how economic models become bases for

explanation and prediction. The third account I consider is *concretization*: models make tendency or capacity claims and they apply to a phenomenon if the model plus corrections and additions to it identify the relevant capacities and disturbing factors underlying the phenomenon of interest. This account of application cannot function without knowledge of tendencies or capacities. However, as I argue throughout chapters 2 and 3, that knowledge may be hard to come by in special sciences.

Chapter 2 presents a study of perhaps the most successful design of a social institution on the basis of game theory. That institution is the FCC spectrum auctions, and the relevant branch of game theory is auction theory. I describe the respective roles of theory and experiment in this case. Although theory was not treated as making tendency or capacity claims (at least not by some of the designers), and although no robustness analysis was possible, still theory was used to inform the design nevertheless. It provided categories in terms of which auction designers created one material environment which satisfied the demands placed on them by the government. I argue that none of the accounts reviewed in chapter 1 can make sense of this prominent example of theory application. So this leaves the question: how *do* we apply economic theory when robustness analysis fails and when we lack knowledge of tendencies or capacities?

Chapter 3 develops an account of theory application that fits the auction design case and, unlike its competitors, is flexible enough also to capture the use of theory in special sciences in general. On this view, models should be treated not as ready-made ceteris paribus or capacity claims but instead as *open formulae* for causal claims. They are open in that they do not by themselves commit us to any one

situation in which a causal relation holds. Rather, they serve as templates for building causal hypotheses. How do we apply open formulae to generate explanations, for example? I propose that we do so by specifying an empirical realization of the causal relation in the hypothesis, that is a specific environment such that if it obtains then the causal relation holds. I explain how this account makes sense of the case of auction design. It is also in line with the practice of experimental economics, and is able to provide normative constraints for the use of rational choice models for historical explanation.

Chapter 1

Three Accounts of Models and Their Application

I. Introduction

This dissertation focuses on one function of science – that is to license empirical successes such as policy interventions and explanations. In particular, I shall focus on how this is achieved in a field of contemporary economics. Economics is more similar to physics than to, say, psychology, in that it has an intricate and sophisticated body of formal theory, the development of which takes up much of the human resources of the discipline. Rational choice models which trace behavior of ideal agents under different assumptions make up much of microeconomic theory, a branch of economics that focuses on individuals and aggregate units such as firms, or governments whose actions and decision making can be represented as rational. Sometimes microeconomic models are used in policy making, in particular for designing reliable institutions. In cases where institution design is successful, what is the role of models? Further, what can we infer about the status of these models on the basis of how they enter in design and justification of institutions? Can their role here shed light on whether and how rational choice models more generally should enter explanation in social sciences? These are the questions I address in this dissertation. Since the sort of models I am focusing on are similar to models currently used in some areas of biology, my conclusions are generalizable.

In this chapter I present three accounts currently on the market. Each of these accounts answers two questions: what sort of claims do economic models make and

how these claims then come to be applied to empirical phenomena. The two questions are two sides of the same coin; what pieces of knowledge we read the models as containing should determine how we put these models to use. One popular view identifies models with *ceteris paribus* claims and application with satisfaction of models' assumptions by the target system. Another takes models to make causal claims that apply to a situation only if they pass a relevant *credibility* test. On the third view models make claims about separate causal contributions of certain factors, or *capacity* claims, and they apply provided that we have identified the most important causes of the phenomenon in question.

There are two levels at which one can evaluate these accounts. Firstly, they can be judged on general philosophical grounds such as whether they are coherent and complete in the sense of answering all the questions that an account of theory application should answer. Secondly, they can be judged on whether they fit the particular case study of application of economic theory that I will be examining in chapter 2. None of the accounts I review in this chapter pass the second test. However, this, of course, is not a sufficient reason to reject them. First of all, the case I shall examine is only one example of scientific practice. Secondly, there is surely no one way to apply theory, universal across all scientific applications. Some of these accounts may be valid in areas of science other than the one I am discussing. This indeed is my conclusion with respect to the first and the third accounts, respectively those of models as *ceteris paribus* and capacity claims. However, I also give systematic reasons to believe that they treat models by standards that are too high to satisfy in many areas of special sciences. So I argue we need a more flexible account

that fits these practices better. The second account in terms of credibility serves to show a way of evaluating models that is different from what I have in mind when I speak of application. That is, credibility analysis does not secure explanations and policy interventions and thus is not a genuine competitor.

Before that, I shall clarify two issues. First, what exactly does my category of "empirical successes" encompass? In particular, since there are many ways in which scientific knowledge can be used, what scientific practices does it exclude? Second, given that even in economics alone there is a variety of different kinds of models, what sort of models are relevant for this project?

The first clarification is important because the broader one's characterization of empirical success is, the more practices involving models will qualify as conducive to it. So for the purposes of a dissertation it is important to narrow down my focus. The first exclusion are all the achievements that aim directly at development of economic theory. These can take the form of building new models that may be more general or make different kinds of assumptions, etc. So when John Harsanyi developed new tools that extended game theory to deal with situations in which agents have incomplete information this was undoubtedly a scientific success, but not by itself an empirical one. Alternatively, theoretical development in economics can consist in deducing theorems from assumptions common to a number of different models. For example, in auction theory, a branch of game theory, there are theorems proving that many types of auction models that have different assumptions in fact yield the same revenue to the seller. These theorems are regarded as a major

breakthrough in economics, but, once again, it is fair to exclude these from the empirical successes of economics.

So far so uncontroversial. I want to go further and exclude from my analysis some of the practices that many economists and philosophers consider as explanatory. Providing explanations of singular and recurring phenomena *prima facie* qualifies as empirical success, so I owe a justification for this exclusion.

Some practices considered to be explanatory involve *casual* application of models, which occurs when theorists relate their models to empirical phenomena claiming that the causal process identified in a model is actually similar to some empirical causal process. For example, the famous Prisoners' Dilemma game is often invoked in cases of price war between, say, two gasoline stations located opposite each other. In cases like this appeal to vague similarities between the model and the phenomenon is often the beginning and the end of explanation. Economists sometimes call the approach "casual empiricism" and oppose it to the more systematic methods that involve econometric testing and experimentation. Casual application is usually done by theorists not by econometricians or experimentalists, and, as the name suggests, often involves nothing more than drawing a vague and intuitively appealing analogy between the model and the phenomenon.

Casual empiricism has a high status in the profession, probably because it is conducive to theory development. For example, Thomas Schelling, the recipient of the 2005 Nobel Prize in Economics, is famous for developing game theoretical models of arms races, conflict bargaining and deterrence. He took familiar situations, such as the arms race between the USA and the USSR and recast them as non-cooperative games.

Testing these models and applying them for specific predictive and explanatory tasks is more difficult and has not been central to Schelling's career. However, their explanatory power is rarely questioned. Thus the press release of the Nobel prize committee credits these models with "great relevance for conflict resolution and efforts to avoid war", stating that "his analysis of strategic commitments has explained a wide range of phenomena, from the competitive strategies of firms to the delegation of political decision power" (Press Release: The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 2005, 10 October 2005).

It is often said that while economics has few precise predictive successes it has many explanatory ones. Rational choice models, in particular, are said to be better at explanation than prediction. In contemporary political science, for example, wars, rebellions, conflicts or lack thereof are explained using such models (for a prime example, see Schultz 2001). It is thus tempting to base one's account of economic models on these putative explanatory successes. I want to resist this temptation, because it is not clear how to delineate clear explanatory successes in economics and other social sciences that use its methods from mere creation of feelings of understanding. These feelings are known to many who have studied economics. For example, the more game theory one does the more one begins to see social interactions as games, the more models begin to look explanatory.[1] This explains the wide reach of "casual empiricism". These feelings of understanding may be a legitimate topic of research in itself, but I do not think they can give us a secure basis on which to build an account of how models teach us about reality in a way that enables reliable policy

[1] For evidence see Marwell and Ames (1981) and Frank, Gilovich and Regan (1993).

making. So I want to use 'empirical success' as not referring to casual applications. Indeed in the last section of chapter 3 I shall articulate exactly what makes some explanations casual.

This does not mean that the account of models and their application I shall offer has no relevance for explanation. Some current explanations of empirical phenomena in economics shall qualify as genuine by the standard of the account I propose in chapter 3. Other explanations aim not at empirical fit but at compactness and concision, and these will not be illuminated here. Yet more explanations in social sciences do not make an attempt to make use of formal theory such as economic theory. Again my account is not designed to apply to those. However, for those explanations that do aim at empirical fit and do use economic theory but fail to meet the standards of my account, I shall give reasons to doubt that they are genuine explanations. These arguments, however, will have to wait to the end of Chapter 3.

So what sort of empirical successes *are* relevant for constructing a philosophical account such as the one I am aiming it? My view is that these should be the cases in which we <u>know</u>, not just feel, or have casual evidence to think, that economic analysis has been successful in a specific way. One such example is design of reliable institutions on the basis of economic theory. We may have evidence to the effect that such an application is successful from the correctness of our predictions about the performance of this institution. Since, unlike in the case of making an explanation of an already existing event, we can take no advantage of hindsight to correct our claims, designing an institution on the basis of game theory is more risky than putting forward a historical explanation. The most famous applications of

economic theory for policy – such as control of inflation, business cycles and monetary policy by central banks and governments – belong to the field of macroeconomics. However, microeconomic theory has also recently been used in high profile policy decisions. Perhaps its most famous instance is the design of spectrum auctions with the help of game theory, such as the auctions first carried about by the Federal Communications Commission in 1994. When auction theory is used to devise an auction to privatize parts of the electromagnetic spectrum in accordance with the government requirements, as I shall discuss in chapter 2, empirical success means both predictive accuracy (to some extent) and understanding of how parts of this auction fit together in a way that allows some of them to be modified without destroying the rest. Models, which describe what a rational bidder does given the rules of the game, the agent's beliefs and preferences, in this case are judged on their ability to tell us what rules the real auction should have to achieve the results we expect. Models are thus judged on their ability to predict correctly, or approximately correctly, the behavior of real world telecommunications executives under different circumstances. Application of models, in these cases, means their use to discover what makes these executives bid in the way they do and how we could change their bidding patterns in the way that we want. The economists hired to design the FCC auctions needed to secure, strictly speaking, only predictive accuracy within a certain range of precision. However, they did not treat this task as solely one of forecasting. On the contrary they engaged freely in causal explanation of the outcomes of this particular auction, and felt that they had some understanding of why this auction worked in the way it did. This mix of

explanation and prediction is the sense of model application I shall have in mind throughout this dissertation.

Given the nature of auction design this decision takes me away from the sort of economics done in academic journals or books, towards applied economics done in the field. By applied economics I do not mean just modeling outside the classical domain of the equilibrium theory. Often the term 'applied' is used by economists to talk about theoretical work on a topic more specific than the original theory. For example, auction theorists who develop models on the basis of game and price theory often refer to their work as 'applied'. When I use the term 'applied economics' I have in mind application of theory for management of very specific empirical phenomena, such as the Spanish market for tomatoes, the distribution of electromagnetic spectrum in the US, the operation of the Federal Reserve, and so forth. More than pure economics, applied economics is often done outside academic departments in policy think-tanks, business schools, governmental and private research centers. Nevertheless, the Department of Applied Economics in Cambridge University created in 1945 at the instigation of John Maynard Keynes has been particularly influential. So it is fair to cite their definition of the field:

Applied economics is the bringing together of economic theory, measurement and methods of statistical and econometric analysis, and the interpretation of such analytic work to elucidate economic phenomena and to inform economic policy (Begg and Henry 1998, 4).

Note in particular the emphasis on *interpretation* of theory and empirical data. Wynne Godley, a long time Cambridge economist and a committed Keynesian with more

interest in policy than in economic theory, argues that only a murky mixture of intuition, theory, data, and "a store of local knowledge and expertise"(Godley 1998, 261) enables him and his colleagues to forecast and pass judgments on macroeconomic policy. Although this dissertation will concentrate on applied micro-, rather than macroeconomics, I too will be interested in the complex judgments involved in the application of theoretical knowledge.

The second promised clarification concerns the sort of models I wish to discuss. The most straightforward way is to give an example:

A. Private Value Auction with two Bidders:

Two players, Player 1 and Player 2, are competing for a good in a first-price sealed-bid private-value auction. These assumptions mean respectively that the winner pays amount equal to the highest, not the second highest bid, that players observe only their own and not the other player's bid, and that each player is certain about how much they value the good for sale. (The latter assumption of private values is appropriate, for example, when buyers compete for a work of art that they value for its beauty rather than its resale value.) Thus player 1 knows her valuation of the good v_1 and observes her bid b_1, and player 2 knows her valuation of the good v_2 and observes her bid b_2.

Assume that both v_1 and v_2 are uniformly and continuously distributed in the interval between 0 and 1000, that is every value is equally likely and positive probabilities can only be assigned to intervals. This allows us to establish that $\text{Prob}(v_i<x)=x/1000$.

How should the players bid? In game theoretical terms this auction is a game with incomplete and imperfect information. The information is incomplete because the players do not know each other's valuation. Game theorists find it helpful to think of this feature in terms of a move by a third actor Nature. Nature makes the first move in the game and randomly picks players' types, in this case their valuations, from a probability distribution with specific properties. But players do not observe all of Nature's moves: each player knows only her own valuation. The information is also imperfect because each player does not know the full history of the game, i.e. the move made by the other player. Rational players will attempt to bid in a way that maximizes their expected utility given their beliefs about what the other player does (this is the crucial assumption of Bayesian Nash Equilibrium).

Suppose player 1 believes that player 2 will bid $b_2(v_2)=av_2$ where a is some constant. Suppose player 1 considers bidding x, in which case her payoff is given by the following condition: (v_1-x) if she wins and 0 is she loses. Given the further assumption that the bidders are risk neutral, p_1's expected utility (EU_1) can be expressed as follows:

$$EU_1= (v_1-x)Prob(b_2<x)+ 0Prob(b_2>x)$$

$$EU_1= (v_1-x)Prob(b_2<x)$$

Since $b_2= av_2$,

$$EU_1= (v_1-x)Prob(av_2<x)$$

$$EU_1= (v_1-x)Prob(v_2<x/a)$$

Knowing that $Prob(v_i<x) = x/1000$ we can rewrite player 1's payoff function as

$$EU_1=(v_1-x)(x/1000a)$$

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

$\Delta(EU_1)/\Delta x\ [xv_1/1000a - x^2/1000a] = 0$

$v_1/1000a - 2x/1000a = 0$

$x = v_1/2$

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

This model, found in intermediary economic theory textbooks (see, for example, Osborne 2003), applies the techniques of game theory to auctions. Auctions are thought to be particularly well treated by such techniques because they are strategic environments in which bidders' behavior and expectations depend on the behavior and expectations of other bidders and also on the rules of bidding that the auctioneer puts forward. So game theoretical models, into which these features can be built, seem to be well suited for the task of analyzing how different auction designs, information distribution and other features can affect outcomes. Indeed auction theory is precisely a collection of such models and theorems. Theorists develop a typology of auctions (open or sealed-bid, second or first-price, with or without reserve price, etc.), types of information known to bidders (e.g. whether or not they receive it from the same source), assume, along with many other things, that agents play Bayesian Nash

equilibrium, and on this basis solve games. At a further stage of theorizing, relations between models can be explored by theorems.[2]

This model made use of game theoretical solution concepts, some properties of probability distributions and expected utility theory. Although the use of these particular tools is frequent, there is no one set of tools that all models in economics invariably use. Indeed there are many different types of equilibria on offer in contemporary game theory, and there are also models that assume bounded rather than perfect rationality. So it is not by the content of assumptions that these models should be identified.

Is there enough of a family resemblance between these models to justify lumping them together for the purposes of discussion of their application? Frigg and Hartmann (2006) distinguish between several types (not necessarily mutually exclusive) of models across natural and social science. *Iconic* models are "naturalistic replica or a truthful image of the target" (2, manuscript), such as material models of bridges, organisms, molecules etc. *Phenomenological* models are those in which only observable features of their targets are represented. In econometrics, some forecasting models, which state the relationship between measurable indicators without aiming at capturing the mechanism that generates them, are of this kind. *Analogical* models attempt to represent target systems by replicating certain of its properties, as do hydraulic models of economy, the billiard ball model of gas or a computer model of mind. Finally, *idealized* models use simplifications of complex phenomena and

[2] For an introduction to auction theory see Klemperer 1999 or Chapter 2 (section II) of this dissertation for a brief overview.

processes with the objective of making them more tractable or exposing the principal mechanisms characterizing these processes. The latter are the models I wish to focus on. Microeconomic models are, for the most part, of this kind[3], for they are designed with the aim of stripping human action and decision making of all the features that do not pertain directly to a certain kind of rationality. This rationality is supposed to capture an important causal factor in social interactions, which is what justifies examining these models together.

For now, this much explanation is enough. In chapter 3 I shall offer an account of the sort of claims we should interpret models such as above to make. I shall also then be able to state in more general terms what kind of models, other than standard models in microeconomics, this account is applicable to.

Now that we are clear on what sort of empirical successes and what sort of models I have in mind, we can proceed to examining the accounts of these models and their application already on the market.

II. Application as Satisfaction of Assumptions

A. Hausman's version

The first account of model application I shall consider is best articulated by Daniel Hausman in his seminal The Inexact and Separate Science of Economics. Hausman develops a version of what is known as the semantic view of theories, on

[3]Thomas Schelling's famous model of how segregation arises even in the absence of racial prejudice seems to be analogical. One articulation of it uses pennies and dimes to stand for people and a checkerboard for neighborhoods (Schelling 1978). However, it is also an idealized model.

which theories are sets of models and models are structures made up of entities and relations between them (van Fraassen 1980). This view differs from the older syntactic view which regarded theories as sets of statements. But we do not need to get into these details. Hausman regards models such as the one I presented above as constituted by the set of their assumptions. The primary aim of models is to provide definitions of predicates (Hausman 1992, 74-77). For example, in the case above the model should be understood as the set of assumptions required to deduce the equilibrium bids and its goal is to define a predicate "is a private-value first-price auction". So the claim our model makes is as follows: a mechanism described by such and such rules and such and such equilibrium bids is a private-value first-price auction. By that Hausman means that the point of a model is not to make claims about any actual auctions, but rather to define what we mean by a private-value first-price auction. Since such a definition only provides language rules for using this predicate, whether its conclusions, such as the equilibrium bidding function, are true or false is not an appropriate question to ask. Rather models are evaluated on whether its conclusions follow logically. One point of building such models is "conceptual exploration" (Hausman 1992, 77), that is the development of such predicates and concepts on the basis of which more models can be built.

Whether or not models should be understood as predicate definitions is not what is most interesting for us (indeed other proponents of the semantic view of theories do not follow Hausman here, see Giere (1988). How models come to be applied is more pertinent. Economics does not stop at conceptual exploration. When we wish to apply a model to explain an actual economic phenomenon, we make,

according to Hausman, a "general theoretical hypothesis asserting that the assumptions of the model are true of some portion of the world"(1992, 77). A model can come to make claims about the world once it is supplemented with such a hypothesis. If a model defines a classical particle system in which any two bodies attract each other in accordance with the inverse square law, then one possible theoretical hypothesis is that the solar system *is* a classical particle system. For the model that I described above, one theoretical hypothesis is that some particular art auction is a private-value first-price auction. This hypothesis asserts that the assumptions of this model are true of some aspect of reality. It specifies a domain where we think this hypothesis obtains, i.e. one portion of reality in which the assumptions of the model hold. It is at this stage that claims of the model become assertions about the world and hence testable and potentially explanatory (Hausman 1992, 76-77). For example we might explain an outcome of an art auction by saying that the particular set up of rules explains why a painting was sold at a price below the valuation of the highest bidder. When a model is supplemented with a theoretical hypothesis we have a theory which makes empirically testable claims about the world, and thus can be true or false. Depending on how broad is the portion of reality which satisfies a model's assumptions – is it one particular art auction? All art auctions in the New York area? – this theory will be more or less general.

So for Hausman, application of a model means making a claim that "all the assumptions of the model are true of the relevant aspects of reality"(Hausman 1992,

76).[4] One problem Hausman is aware of is that this criterion is much too strict because economic models use assumptions that are not true of any portion of reality.

Indeed, the main feature of rational choice models that has fascinated methodologists since the inception of economics is the use of "unrealistic" assumptions. A number of different types of assumptions have been described as "unrealistic". Some such assumptions are *restrictive* in that they are true of some situations, but not all and hence restrict the range of situations to which a model applies, others are universally *false* in that they do not seem to apply to any empirical cases. An example of a restrictive assumption in the private value auction model is that there are only two players, or that the players know the value of the good for sale, or that it is a sealed bid auction. On the other hand, the assumptions that that actors make decisions faultlessly according to complex calculations of utility as defined by the rules of Bayesian Nash equilibrium is most naturally interpreted as literally false, though many economists treat it as a good enough approximation or a way of describing the central tendency (in the statistical sense) of behaviors for many situations.

So Hausman's criterion needs to be adjusted to accommodate the fact that economic models employ assumptions that are universally false and the fact that economic models get applied nevertheless. With assumptions such as perfect rationality, function differentiability and continuity of probability distributions, how can these models ever become claims about the world on this view? In The Inexact

[4] Other advocates of the semantic view adopt different articulations of what a theoretical hypothesis claims. A theoretical hypothesis can claim isomorphism or partial isomorphism or similarity between the model and the target system.

and Separate Science of Economics, Hausman does not give us much of an answer. To be precise, in his discussion of Milton Friedman's view he notes that "what is relevant in the messy world of economics is not whether the assumptions are perfectly true, but whether they are adequate approximations and whether their falsehood is likely to matter for particular purposes"(Hausman 1992, 167). This is a relaxation of the earlier criterion: now all assumptions have to be true enough relative to the task. However, Hausman does not elaborate on what exactly that means. He merely notes that application of models is an important and thorny problem (Hausman 1992, 80-81). Perhaps Hausman is right not too elaborate on the problem – how closely the assumptions should approximate the features of the real world system to which the model is applied can be decided in a context-specific way on a case by case basis. This may be so, but it does not imply that there is nothing more to be said about how these decisions are made. So I look for an elaboration explaining what Hausman's requirement of assumption satisfaction means.

B. Morgan's version

Mary Morgan, historian and philosopher of economics, has dedicated much of her work to studying how models in economics and other sciences are used in practice. Her work in the 90s in collaboration with Margaret Morrison showed some weaknesses in the standard account of models, i.e. the semantic view (Morgan and Morrison 1999). In particular, Morgan emphasizes that models do not just represent reality passively. Rather we do things with models, we learn by manipulating them.

Morgan documents four activities connected with modeling in science in general (2002, 42):

1. Build a model to represent the world.
2. Ask questions about the model.
3. Manipulate the model to answer questions.
4. Relate answers to the phenomena of the world.

In her most recent writings (2002, 2003, 2005), Morgan chronicles a number of different epistemic practices loosely connected by the term 'experiment'. The most common way to "experiment" in economics is to build a mathematical model (on the basis of a number of assumptions) and perform a comparative statics exercise on it. For example, we could "introduce" an exogenous change in demand for a commodity and "observe" how this affects its price. I use quotation marks because this is *not* what we normally mean by experimental intervention. In 'model experiments', as Morgan calls them, the experimental setup is controlled not by material intervention but by introducing assumptions and results are demonstrated not by measurement but by logical or mathematical deduction. This is in stark contrast to laboratory experiments in economics, in which people may be asked to follow the experimenter's instructions to sell or buy under certain budget constraints in a specifically arranged environment. Here the inputs, intervention and outputs are all material, as opposed to mathematical, i.e. real people, laboratory environments and prices that their interactions generate. In addition to these two types there are also a number of 'hybrid experiments' that

combine features of both laboratory experiments and mathematical models. For example, some inputs may be material or quasi-material (such as real data), while intervention may be performed through a computer simulation.

Morgan's point is that in each of these cases the epistemic warrant attached to the experiment are different. In case of mathematical models, we know all the resources built into the system and hence we know why the result follows in the way that it does. But inferring behavior of the real world entities which the model is supposed to describe is extremely difficult because of the constraints on the situation introduced by the assumptions. In a laboratory experiment, on the other hand, there is a possibility of unknown confounding variables intervening, and issues of external validity are different. There is no one unified epistemology of these practices.

Relevant for our interests, laboratory experimentalists and mathematical modelers, Morgan argues, approach step 4 differently. The former worry about the problem of "parallelism", that is whether the results that hold in the lab will hold also in the world, whether the lab environment is not too idiosyncratic and too constrained to represent the relevant portion of the world. The modelers, on the other hand, worry about realism of assumptions of the model. Although the two are related, they are not the same. For Morgan, the inference gap from the experiment to the world is wider in case of mathematical models. In case of lab experiments we may have doubts about external validity if the lab environment is too "artificial" and about whether we took care of potential confounders, but we at least are dealing with real human beings in material environments. However, in case of mathematical models the confounders and disturbing factors are dealt with by making assumptions. The key absent in

mathematical models is *materiality*. In the lab a scientist has to think through and investigate materially all the factors that may confound the experiments, while a modeler can just make an assumption. This allows for a greater freedom in building different models but the cost is that the target system may not fit the mathematics of the model. I shall return to these issues in chapter 3.

So how can the inference gap between mathematical models and the world be closed? In her other writings Morgan emphasizes one important device by which economists relate models to the world (2001). That device is story telling. Stories or narratives, she insists, are needed not for merely heuristic purposes, but to put the model to life, so to speak. Initially the model is given a dynamic element when we ask a "why" or a "what if" question about some feature of the economic world. For example, what happens to production of a good if there is a sudden change in technology? Then economists typically tell a story that fills in the details into the bare structure of a comparative statics exercise. These stories have a typical narrative sequence in which some change takes place which initiates another change and ends at a new point. Morgan argues that the story is part of the identity of the model, because the model cannot tell us anything about the world without a story attached and, moreover, the bare mathematics do not determine the sort of stories that can be told on the basis of a model. (She does not explain what else is needed to determine stories). By attaching stories to bare models economists are able, Morgan claims, to connect the simple structure embodied in a model to the real world. Knowing the range of specific stories we are able to tell on the basis of a model allows us to understand this model. This account is similar to Hausman's in that it too denies that models make any

claims about the world by themselves, but it differs in that Morgan thinks it is a story, rather than a theoretical hypothesis, that connects the model to the world.

But how exactly does that help with the inference gap? Although in the 2001 article Morgan seems to imply that stories actually make models more realistic (without explaining how), in her 2002 article she explains that story-telling only tells you about the *form* of argument economists use, not about the *grounds* that make the inference from a model to the world reasonable (Morgan 2002, 55). These grounds can only be provided once we can show "how well we have represented the world and whether we have been able to capture to some degree the structure of the problem"(Morgan 2002, 56).

How do we know if we have captured this structure? Satisfaction of assumptions seems, for Morgan, to be the only way. This is why I discuss her views here. Unlike Hausman, she does not require that all assumptions of a model be satisfied:

> Passive irrelevant factors can and should be simplified away, and we can go with idealized accounts of motivations, provided we can give a rather more accurate representation of the behavior, situations or interactions relevant for our purposes. … The realism of assumptions does matter, but not of all of them at the same time (Morgan 2002, 56).

Unfortunately, Morgan does not tell us what makes relevant assumptions relevant and irrelevant ones irrelevant. Her account of application needs such a criterion as much as Hausman's. Nor does she tells us how the compelling problem of

materiality, she raised in connection with economic models, is solved in the process of model application.

C. Which assumptions are irrelevant?

If application of a model amounts to finding or to creating a target system that satisfies the model's relevant assumptions, we need some characterization of relevance. For example, when attempting to implement a private-value first-price auction, we may want to know which assumptions of the model must be satisfied by our institution for the institution to function as the model *predicts*. On the other hand, we may be interested in whether the behavior of some actual auction in which bidders pay first-price is *explained* by the claim made in the private-value first-price model. In this case also we need to know whether it matters for the purposes of explanation that the actual auction does not satisfy all the assumptions of the model. For both possible ways of application, prediction and explanation, such a criterion is necessary.

To make progress on this issue I shall formulate two such criteria, one naturally fitting the predictive goal, and another the explanatory one. The criteria are respectively:

i. Ignore those assumptions whose falsity does not matter for purposes of reliable prediction.

ii. Ignore those assumptions that do not matter for the purposes of derivation of a qualitatively similar result.

One question I ask in evaluating these criteria is whether they give sufficient conditions such that if we satisfy one of these criteria, then the model in question shall apply for respectively prediction or explanation. More specifically, for the first criterion the question is: Are there assumptions whose falsity we can ignore for purposes of reliable prediction on the basis of a model? I argue it depends. As for the second criterion, it is clear that there are assumptions that do not matter for the purposes of derivation. I distinguish between two different formulations of this criterion: de-idealization and robustness analysis. On the de-idealization formulation, if this criterion is satisfied then the model does explain. But the more pertinent question is whether it is reasonable to require that this criterion be satisfied for purposes of explanation. I argue that it isn't. On the second formulation, however, I find no justification for this criterion.

D. Friedman's instrumentalism and its critics.

Perhaps the most influential endorsement of criterion (i) came from economist Milton Friedman. In his 1953 paper he defended economic models with unrealistic assumptions, so long as they can predict well. In fact he took this defense further than my formulation of criterion (i). Making criterion (i) relevant for purposes of prediction suggests that we may sometimes be interested in something other than prediction. For Friedman, there is nothing more, that is apart from theory building and prediction, economists should be interested in. Throughout his paper he uses the term 'explanation' in quotation marks. For example: "Viewed as a body of substantive hypotheses, theory is to be judged by its predictive power for the class of phenomena

which it is intended to "explain"" (Friedman 1953, 184). By explaining Friedman means nothing more than making predictions that turn out to be true. Friedman argues that economists should build models that predict well in the intended range of phenomena, not describe accurately or supply understanding of economic phenomena. Even if economic theory seems to provide understanding, it does so only in virtue of its predictive capacity. Furthermore, Friedman contends that model's "importance" or "significance" is inversely related to the realism of its assumptions. By importance and significance, Friedman means predictive capacity or the amount of true predictions a model makes: "A hypothesis is important if it "explains" much by little, that is, it abstracts from the common and crucial elements from the mass of complex and permits valid predictions on the basis of them alone." (Friedman 1953, 188)

What is the intended range of phenomena? The appropriate domain of economic theory is not one of individual decision making. [5] Hence Friedman argues "of course, businessmen do not actually and literally solve the system of simultaneous equations"(Friedman 1953, 193). Indeed their actions could be "habitual reaction, random chance, or whatnot"(ibid.). The relevant prediction of maximization of returns hypothesis is not how businessmen make decisions but rather it is their actual behavior in the face of external circumstances. So if sociologists or psychologists demonstrate that people do not have the preferences or do not use decision-making rules economists assume they do that is irrelevant for evaluating economic models. Friedman thus claims that it is inappropriate, for the purposes of developing economic theory, to concentrate on examining and improving upon such standard assumptions of

[5] Friedman uses 'model' and 'theory' interchangeably.

neoclassical economics as “perfect competition”, “perfect monopoly” and whether businessmen do indeed reach their decisions by consulting their cost and revenue curves (Friedman 1953,188).

So it is necessary to reformulate our criterion of assumption irrelevance in terms true to Friedman. Only some predictions matter, in particular those that concern the theory’s intended domain. This makes sense if we wish to use the predictive criterion to restrict the number of assumptions that need to be satisfied for a model to apply. Requiring that all predictions matter is tantamount to requiring that all assumptions be satisfied, since a model’s assumptions are included in the set of predictions it makes. So the right version of a predictive criterion is as follows:

> i′. Ignore those assumptions whose falsity does not matter for purposes of reliable prediction in the intended domain of prediction.

On the question of application of models, the difference between Friedman’s position and the views of Hausman and Morgan cannot be overemphasized. Whereas for Hausman and Morgan a model gets applied by finding an environment in which its relevant assumptions are satisfied, Friedman argues that the model’s assumptions are not a useful guide for finding environments in which it applies. He argues: “the entirely valid use of “assumptions” in *specifying* the circumstances for which theory holds is frequently, and erroneously, interpreted to mean that the assumptions can be used to *determine* the circumstances for which a theory holds”(Friedman 1953, 191, original italics). My reading of the difference between specifying and determining the

circumstances for which a theory holds is that the former is tentative while the latter is not. When we use some of the model's assumptions to get an indication about the possible domain of this theory we are not thereby committing ourselves to a view that these are all and only circumstances in which a theory works. So, for Friedman, we look at assumptions at most at the beginning stage of model application.

Philosophers and economists have criticized Friedman's views extensively. I shall concentrate on the arguments that attack two of Friedman's claims: the stronger claim that models necessarily should be viewed as instruments for prediction and the weaker claim that models can be treated in such a manner.

i. Musgrave

Alan Musgrave argues against an instrumentalist reading of economic models thus attacking the first claim (Musgrave 1981). His strategy is to draw distinctions between different kinds of assumptions Friedman talks about and to show that they are not all "unrealistic" in the same way and for the same reasons. Once these distinctions are clear, we can see that models can be accommodated within scientific realism, the view that scientific theories state facts about the true structure of the word. Realism is Musgrave's preferred view, so he wishes to establish that instrumentalism, the view that theories are just instruments for keeping track of observations, is no longer the only option. For example, some of the assumptions Friedman talks about are, what Musgrave calls, *negligibility* assumptions: there is no air resistance, there is perfect competition, etc. Read as descriptive statements they indeed sound false for many situations. However, read as *negligibility* claims, they assert that effects of air

resistance or imperfect competition are negligible for the targeted phenomenon, and these claims may be correct for certain environments. Since many of Friedman's examples of false assumptions are precisely negligibility assumptions, and since these assumptions do not have to be read as making outright false claims, Friedman's assertion that only descriptively false models can "explain much by little" is undermined.

Another two kinds of assumptions Musgrave discusses are *domain* and *heuristic* assumptions, where the former specify the domain to which a theory is meant to apply and the heuristic are just the assumptions that are made for purposes of theory development. If a domain assumption is not satisfied by any situation then the theory is not testable. If we value testability, we should want domain assumptions to be true at least sometimes. It is not clear why Friedman should go along with this claim, since he does not believe that models themselves conclusively establish the domain of their applicability. A model may assume perfect rationality but that does not mean that it only applies to perfectly rational agents. However, looking at the model's assumption is the most natural way of picking out its domain of applicability, and Friedman does not explain another way of doing so.

If a heuristic assumption is false then it is, most probably, because we are at the beginning stage of theory development and plan in the future to replace this assumption with a more realistic one. Musgrave's example is Newton's assumption that the solar system consists of only one star and one planet. When this assumption was later dropped to incorporate other planets, this development was seen as a crucial improvement. For Friedman this improvement can only come from better predictions,

but, for Musgrave, dropping Newton's assumption is crucial for developing a better astronomy even if it would never lead to better predictions.

I think Musgrave has been successful in undermining the stronger one of Friedman's claims. However, it is important not to read Musgrave as doing more than that. He merely shows us that scientific realism about economic models is a viable option and that economic models, in virtue of their assumptions, do not uncontroversially compel us to take an instrumentalist stance. Musgrave does not show that the realist option is actually preferable to the instrumentalist one. To do that would be to attack the second, and the weaker one, of Friedman's claims. I now turn to commentators who have attempted to do that.

ii. Causalists and their critics

The common intuition against Friedman's second claim has been that models should be evaluated on their explanatory power, and that does not reduce to its predictive power. Herbert Simon in his comments on a 1963 session of the American Economic Association dedicated to Friedman's views appeals to our notion of explanation to urge that economists do test assumptions of their models. That businessmen pursue profit maximization is one of the putative causes of the macroscopic market phenomena. If businessmen do not in fact do that, then we fail to provide an explanation of these phenomena (Simon 1994, 215). Friedman treats "perfect competition" on a par with the physicists' "perfect vacuum". But the latter is only a legitimate part of physical explanation to the extent that the real world sometimes approximates it. There is much evidence that many assumptions of

economic models do not approximate any features of real economies. Simon proposes that Friedman's "principle of unreality" be replaced with Simon's "principle of continuity of approximation", which asserts that "if the conditions of the real world approximate sufficiently well the assumptions of an ideal type, the derivations from these assumptions will be approximately correct"(Simon 1994, 216). This, of course, calls for a new kind of economic theory, which Simon later advocated on the basis of bounded rationality.

Simon's line of criticism is a popular one (see also Keen 2001). However, it is unsatisfactory in that it simply begs the question in favor of a realist construal of economic theory. Although Friedman never explicitly disavows the traditional view that science is meant to explain as well as to predict, it seems that his views commit him to something like that. For Friedman, models are meant to predict only. A better challenge would be to show that unrealistic assumptions actually undermine the goal of prediction. This would also be more relevant for my purposes of evaluating criterion (i) or (i′). I formulated them so as to explain how models with false assumptions can nevertheless be applied *for prediction*, thereby implicitly endorsing Simon's intuition that prediction and explanation can come apart. So can some assumptions be ignored if our goal is just prediction?

Many have argued no. Along these lines is the critique put forward by causalists such as John Pemberton (2004). Pemberton claims that to be useful for prediction models must capture causes, or at least the most important causes, present in the environment of application, hence the term 'causalist'. Other causalists include John Stuart Mill, Laplace and members of the early Cowles Commission who worked

on development of forecasting methods in econometrics (see Cartwright 1989, chapter 4). Mill, for example, believed that observable phenomena are a result of many causal tendencies operating at once. Predictive tools based on regular association between occurrent properties (or, to use modern language, phenomenological models) will fail in the long run, because the arrangement of causes tends to change and hence regular association between occurrent properties will likely not last (Mill 1843).

In economics this view was especially pronounced among the members of the Cowles Commission, who founded the discipline of econometrics on the principle that, while some phenomenological models can sometimes generate true predictions, this will most likely be due to chance, and the best and most reliable forecasting must be based on accurate causal models. This, along with a requirement that these causes operate undisturbed, is a sufficient condition for a successful use of these models. Pemberton assumes that 'successful use of models' should not mean just a one-off occasional success, but a relatively sustained consistent one. This is a fair assumption since successful prediction by accident should not count towards the model's successes. The problem with ideal models in economics that leave out many causal factors present in the real world is that they give us no promise of consistently successful performance. To quote Pemberton, "if an ideal model fails to capture material causes then it will on occasion be very wrong"(2004, 2, manuscript).

One real world example of such a failure is the Black-Scholes formula, famously (and disastrously) used for stock option pricing in the 90s. It assumes, among other things, that the expected return on a stock is not a factor in an option's value. Indeed the main equation leaves out the expected return as a possible

determinant of the stock's price and hence the option's value. What happens when a major stake in a prominent company changes hands and there is uncertainty over its future? Expert analysts predict that if the purchaser announced a bid to buy the rest of the company's stocks the stock price would rise and if he or she doesn't it would fall by a certain percentage. The probability of a bid surely influences the option's value. The example is one in which all the experts at option pricing agree that the expected return on a stock matters for the option's value, except for the Black Scholes formula! (Pemberton 2004, 7-8 manuscript). Sometimes violations of a model's assumption can be irrelevant in a context of its application: perhaps if a few businessmen do not obey profit maximization as most economic models assume, it would not matter and the economy could still reach an equilibrium. Here, however we have an example in which it is absurd to deny that a violation of one Black Scholes's assumption is irrelevant. It is as relevant as it can possibly be, as those who got ruined because of the failure of the formula eventually found it.

A similar criticism is put forward by Hausman (1992b). He invites us to go along with Friedman and assume that prediction is the only goal of economic models, and also that not all predictions of a model need to be evaluated, but only those of the phenomena the model is designed to work on (in Friedman's example, these are predictions concerning macrolevel market events). Friedman's mistake is to assume that a model's predictive performance tells us all that we need to know about a model's performance with respect to the relevant phenomenon. If a relevant phenomenon is *reliable driving*, then we do not think that a car's past performance provides us with *all* the information we need to predict its future performance. We do

care about what the mechanic can tell us by looking under the hood. Similarly, the truth, or the approximate truth, of a model's assumptions can give us evidence about whether the model will still predict correctly if we plan to use it in a new environment. Option pricing specialists get precisely such evidence when they examine Black Scholes's assumptions. They find out that violation of one such assumption can on occasion spell doom for Black Scholes's predictive performance.

Hausman argues that we should evaluate more than a model's narrow prediction record. We should also evaluate the truth or approximate truth of its assumptions. Unlike Pemberton he does not argue that models should have assumptions that describe causes that bring about the phenomenon. But this is not to say that Hausman should disagree with Pemberton. If we reject Friedman's claim that models should only be evaluated by predictive success in a narrow range of phenomena, which I think we should on the basis of Hausman and Pemberton's arguments, what exactly is the position we are left with? Whether models should be causal and whether their assumptions should be true are, of course, two different issues. But in this case answering the first question will resolve the second.

iii. New evidence from econometrics and psychology

Pemberton's view that an approximately correct specification of causes is necessary for the successful use of models as predictive tools appears to be contradicted by two areas of recent research, one in econometrics and another in cognitive psychology.

In econometrics I have in mind the work of David Hendry and Grayham Mizon (Hendry and Mizon forthcoming). They distinguish between econometric and statistical models. The former normally identify on the basis of economic theory the Data Generating Process (DGP). A DGP is understood as a specification of the main causes of the phenomenon in question. For example if we wish to make an econometric model of the demand for heaters we specify the price as one of its causes and the season as another: $D= \alpha x+\beta y$ where D is the quantity of heaters demanded, x represents the price and y the season, and α and β are parameters. Other models are statistical, i.e. phenomenological. These are meant for forecasting, rather than analysis of economic policy. What happens to the predictive capacity of these two kinds of models when there is a *structural break*? A structural break is a change in a parameter of an econometric system, for example β turning from 1 to 0. Both econometric and statistical models can fail to forecast properly. However, there are devices developed by econometricians that can "robustify" a forecasting model against a certain kind of structural break. That is, there are cases in which a model's predictive power has little to do with how well it captures the DGP. Hendry and Mizon claim:

A key consequence of these results is that the best available forecasting model need not be based on the 'causal determinants' of the actual economic process, and … may be based on 'non-causal' variables, that is, variables which do not enter the DGP (forthcoming, 5, manuscript).

Thus a purely statistical tool can beat a causally sophisticated model for purposes of prediction. A true causal model would indeed be better at forecasting than a purely statistical model if the causes of the phenomenon stayed stable. However, the true

causal model does not stay true for a long time, whereas a statistical model can be made robust to certain changes in the environment.

Nevertheless, this advantage holds only under very specific conditions. Distinguish a *structural break* from a *policy regime shift*. The latter is a change not just in a parameter, but in some un-modeled variable, for example, an introduction of a completely new policy measure. In our example it could be a new scientific finding about the danger of heaters, which affects demand for heaters but is not part of our model. In that case neither of the models is predictively robust, and the authors argue that we need to use both kinds of model, in particular the econometric model serves as a correcting tool for the statistical model (Hendry and Mizon forthcoming, 17, manuscript). [6]

Evidence from psychology comes from the work on human decision-making by Gerd Gigerenzer and the ABC research group of the Max Planck Institute (Gigerenzer et al 1999). Since Herbert Simon popularized the term 'bounded rationality', economists and psychologists have sought to establish research programs that would study the principles behind the decision-making of real world agents. Just how different are these principles from what standard economics defines as rationality, i.e. maximization of expected utility? For Gigerenzer and his colleagues, they are very different. In particular, they are not general rules such as Bayesian decision theory that calculate the expected payoffs from probabilities of all actions in each environment.

[6] Like me Reiss (2005) concludes on the basis of Hendy and Mizon's results, that knowledge of mechanisms underlying a phenomenon is not necessary for many purposes, including prediction.

Rather they are simple, or in Gigerenzer's words "fast and frugal", environment-specific *heuristics* (Gigerenzer and Selten 2001).

Gigerenzer's examples of fast and frugal heuristics include various specific decision rules: how to catch a baseball, how to pick among known alternatives say when deciding where to hunt (search rules), how to decide when to stop looking (say, for a mate), etc. In many of these examples, Gigerenzer claims that expected utility maximization may not only be an extremely unrealistic way to represent human decision making but also a less predictively successful one. For example, a player trying to catch a baseball in all likelihood does not attempt to calculate the trajectory of the ball given its speed, angle, wind, spin, etc. and optimize his chances of catching it on that basis. Instead he uses the "gaze heuristic" which requires that the player moves in such a way as to maintain a constant angle between his eyes and the ball, and this heuristic should bring him to a spot where he can comfortably catch the ball. Generally, heuristics avoid calculations of utilities and probabilities and instead rely on cues (Gigerenzer et al 1999).

There is also experimental evidence regarding the success of heuristics. One striking case is the *recognition heuristic*. This heuristic is a decision rule that might be used in problems in which one of two objects has a higher value on some criterion and the task is to choose correctly. For example, which city is larger, Munich or Dortmund, which of two foods is safer to eat, etc. The heuristic is as follows: "if one of two objects is recognized and the other is not, then infer that the recognized object has the higher value"(Goldstein and Gigerenzer 1999, 41). So applying this heuristics to the examples above would imply that if Munich is recognized and Dortmund isn't,

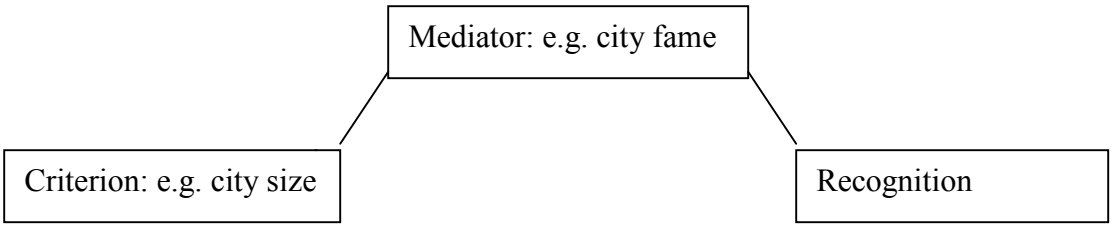
then we infer that Munich is larger. Similarly if one of the foods smells familiar while the other doesn't, the familiar one is safer to eat.

When applied in the right environment the recognition heuristic is at least as successful as other modes of prediction and sometimes more so. German and American undergraduate students were given a task of deciding which city is more populous, San Antonio or San Diego. Remarkably, German students performed better (100% correct answers) than American students at the University of Chicago (62%). Goldstein and Gigerenzer's explanation is that Germans were able to use the recognition heuristic to pick whatever city they have heard of before. Americans, on the other hand, had heard of both cities, and so could not rely on that heuristic and had to rely instead on another rule to make the prediction.

Gigerenzer believes that heuristics are the tools that agents really use and that they can be better both as predictions and explanations of decision making than utility maximizing models of standard economic theory. They are better as explanations, i.e. statements of human cognitive mechanisms, in those cases in which we have evidence to believe that agents are not using utility-maximization to make their decisions (Gigerenzer, forthcoming). Furthermore, heuristics may even be superior when prediction, rather than explanation, is our aim. Gigerenzer is adamantly opposed to Friedman-style "as if" stories even if we only use them for prediction. His main reason seems to be practical: optimization models tend to have many parameters and thus it is

very difficult to evaluate their fit or predictive success. Heuristics, on the other hand, have few parameters and are thus more easily testable (Gigerenzer forthcoming)[7].

But this is not the point relevant for us. Gigerenzer's heuristics qua good predictors often omit full specification of causes of the phenomenon about which agents wish to make predictions. For example, the gaze heuristic does not incorporate the real causes of the movement of the ball and the recognition heuristic says nothing about causal factors influencing city size. In fact its structure is as follows:



This figure adopted from Goldstein and Gigenezer (1999) summarizes the manner in which the recognition heuristic works within its environment. The 'criterion' represents the goal of inference – the point of using the heuristic is to predict its value, in our case, the city size. The value may not be readily available to the agent. This is why we need a 'mediator', which is a variable that is correlated with the criterion and that also happens to be within the agent's accessible environment. In our case, the mediator may be the city's fame, perhaps reflected in the amount of times a city is

[7] That simpler models with fewer parameters can be more successful at prediction has already been pointed out by philosophers of science (Forster and Sober 1994). However, this view is consistent with a requirement that to be successful predictively a model must incorporate the most important, though not all, causes of the phenomenon in question. And it is also consistent with the claim that incorporating its most important causes improves our chances of predictive success.

mentioned in the media. The lines between the boxes indicate statistical correlations. The criterion and the mediator are correlated within a given environment, that is cities that are mentioned in the media are those that tend to be fairly large. Similarly, recognition and city fame are correlated. This correlation is important because it holds only if the agent has little other information about a city except for its fame as reflected in the number of mentions it gets in the media. It is due to these two correlations that recognition is a reliable indicator of the value of the criterion. Note in particular that the structure of this heuristic is entirely phenomenological, that is there is no causal information here.

Another crucial feature of both Hendry's statistical models and Gigerenzer's heuristics is that they come with a specification of the range of environments in which they are likely to be reliable predictors. In the case of statistical models we know that they perform if we have made them robust to a certain kind of structural breaks. To judge whether something is a structural break or a regime shift we need to know which features of the target system the model captures and which ones it doesn't. If a factor is not captured and there is a regime change we think that a model's successful prediction is a fluke.

In the case of heuristics, we are similarly concerned to establish that a given heuristic fits a given environment. Perhaps human decision-makers do that unconsciously or don't do that at all, but that is of no consequence for scientific methodology. If we wish to use heuristics to extract a methodological lesson, we cannot merely look at its performance in its intended domain, we also want to have some evidence about how this heuristic is sensitive to salient environment features.

Under what conditions is the recognition heuristic successful, for instance? Although there are lots of factors that determine a city's size, Germans could take advantage of one fact – if you grow up in a place far from America and don't visit or study it too much, chances are you'd only hear about the relatively populous cities. For American cities, recognition by Germans and city size are correlated. So this is the relationship the heuristic exploits. For this heuristic to work we need to make sure we are in an environment where we have little information about American cities. If we come upon a German who's met someone from San Antonio or who's been to Texas or who's taken a course on American demography (the list of such conditions is likely to be very long), chances are another heuristic will have to be used. In technical terms, the recognition heuristic works when the recognition validity (the proportion of times a recognized object has the higher value on the criterion) is greater than the knowledge validity (the probability of getting the correct answer when both objects are recognized)(Goldstein and Gigerenzer 1999).

We can now take stock. Heuristics and statistical models are phenomenological in the sense that they do not capture any of the causes of the phenomena they are used to predict. They are reasonably good predictive tools for a number of cases, and in some cases they are better tools than causal models. To the extent that we are able to use statistical models and heuristics consistently and successfully, the causalist view is in trouble. In particular, the causalist's strong claim, that causal models are necessary for successful prediction, appears to be false. Moreover, a weaker claim we might impute to the causalist, that a causal model is always an improvement over a phenomenological one, is also in trouble.

What replies might the causalist put forward? One such reply could take form of a dilemma: phenomenological models such as Gigerenzer's heuristics and Hendry's statistical models are not truly successful, or if they are then they are not truly phenomenological. It is, of course, incumbent upon the anti-causalist to demonstrate that heuristics and statistical models are capable of *consistent* successful prediction, since this is the causalist claim that is being attacked. What would such a demonstration involve? One possible answer to this question is that to show that a heuristic or a statistical model is capable of a consistently reliable predictive performance we must show that we understand *why* they yield the right predictions in a given context. What factors in the environment are responsible for this sustained good performance? And how is our model sensitive to these factors? Narrow predictive success, as defined by Friedman, is unable to give us such understanding. Knowing *that* the model predicts is not the same as knowing *why* it does. It is the latter that ensures that we apply this tool to the right environment, the environment of its intended use, which in turn ensures that it predicts correctly. However, this way of defining consistent successful prediction just yields the case to the causalist. Knowing why a model works the way it does with respect to a particular phenomenon is tantamount to knowing the causes of this phenomenon. For example, knowing why aspirin relieves a particular kind of headache involves knowing what factors cause the onset of this headache and how aspirin acts on these factors so as to eliminate it.[8] Similarly, knowing why a model predicts a phenomenon as it does requires knowing

[8] The aspirin example is due to Nancy Cartwright.

why this phenomenon behaves as it does. So being truly successful and truly phenomenological are properties at odds with each other.

However, the anti-causalist can reply as follows. The requirement set up by the causalist to define consistent successful prediction is too strong and question-begging. It is question-begging in that understanding of the reasons behind a tool's predictive success is *not* part of using this tool successfully. To say that it is sets up the problem in a way that makes the causalist bound to win the debate. Sometimes successful use of an instrument involves knowing merely the circumstances in which it will work and those in which it won't. Aspirin is a good case in point. We know *when* to use aspirin for purposes of relieving headaches. For example, we use it for tiredness induced headaches, but not when the headache is due to our skull having split open. However, we don't know exactly *why* aspirin relieves headaches.

The causalist could again object that a *truly* complete list of circumstances under which aspirin has the relevant effect, and hence the truly successful use of aspirin, will require knowledge of the mechanisms by which aspirin relieves headaches. We simply could not specify the full set of circumstances which enable and defeat the effects of aspirin on the headache without understanding the complex chemical reaction that takes place when aspirin is administered. But that would be setting the standards impossibly high. A useful tool does not have to be a perfect tool. If use of aspirin in contemporary medicine does not count as a successful use of a tool, then no use of, say, an economic policy measure would ever qualify as such. Since we wish our account to be relevant in practice, we must settle for a weaker requirement.

So what exactly do heuristics and statistical models teach us about predictive success? More generally, what knowledge do we need to use a phenomenological tool successfully and consistently?

The view that is most warranted by the discussion above is that successful use of a model for prediction requires reliable enough knowledge of conditions under which it will work and conditions under which it won't. I shall call these *conditions of proper use*. The important requirement is that these conditions must be specified independently on the mere circumstances in which the model has been successful. To say that a model can be used for prediction under the conditions in which it is successful for prediction is circular and completely uninformative. We need an independent characterization of the right and wrong circumstances for a model's use. We don't just say that aspirin works whenever it works. Rather we say it is most likely to work when the headache is mild and temporary, when there are no major head injuries, etc. Both the statistical models in econometrics and the heuristics come with precisely such information attached. So while none of the predictive tools examined here attempt to capture causes of the phenomena they predict, there are good reasons why they work as they do and that is that they are *sensitive* to the important causal relations in question. The recognition heuristic is sensitive to the relation between city fame and city size, Hendry's statistical models are sensitive to whether the change in the environment comes from a structural break or a regime change, etc. It is not that we do not need *any* causal knowledge to predict a phenomenon (we probably do, but this is a separate question). Rather we do not need models of causes of this phenomenon.

Where does knowledge of the conditions of proper use come from? In case of models such as those I have been discussing, this information is usually assumed to be contained in the assumptions of a model. Suppose our private value auction model were to be used to predict the amount of bids in some actual auction. One way to identify conditions under which this model will predict correctly is to just follow its assumptions. Since the relation between the assumptions and predictions is deductive, we can be sure that the model will predict correctly under all real world conditions that correspond to the set of assumptions that yield the prediction in question.

However, this way of selecting the conditions of proper use may be too restrictive. First of all, hardly any real world situations replicate exactly the assumptions of such models. Secondly, it is hard to know for sure even if they did. When models such as these are used for prediction, scientists do not treat assumptions as exhaustive of the conditions of proper use of the model (see in particular Guala 2005 and my chapter 3 for a fuller articulation of these ideas). They rely on a variety of experimental methods to figure out just what must be the case for the model to be a good predictive tool. A discussion of these methods is beyond the scope of this dissertation. Here I just wish to emphasize that the assumptions of a model are not a unique guide to the model's conditions of proper use. We should not expect them to be, because assumptions have many other roles to play apart from specifying these conditions. Other reasons such as tractability, elegance, simplicity, etc. also influence whether one or another assumption is used in building a model.

Now we can come back to criterion (i′) – are there assumptions the falsity of which does not matter? On the view defended here truth of assumptions only matters

if these assumptions also correspond to the conditions of the model's proper use. If we do not have any way of learning about these conditions other than by checking the model's assumptions, then the truth of these assumptions is necessary for successful prediction. Indeed the assumptions should be treated as the default statement of the conditions of proper use. However, as I have suggested, assumptions do not have to be treated as containing such information. A model may assume perfect economic rationality. We know that in our intended context of prediction this assumption is false. But we have also run experiments that allowed us to find out conditions *other* than perfect economic rationality under which the relevant prediction is realized. In this case it is the experiments, not only the model, that dictate the conditions of proper use. If our knowledge of these conditions is reliable enough, then we are free to ignore some of the model's assumptions.

However, the criterion (i′) invites us to simply ignore assumptions without making any provisions for learning the conditions of model's proper use in some other way. So it must be rejected. The middle road between Friedman and Pemberton I propose can be formulated as:

> (i″) a model's assumptions can be ignored for purposes of prediction only if they do not correspond to the conditions of this model's proper use.

There is much more to be said on the methodology of forecasting models. Yet, I shall leave this topic here. For the purposes of this dissertation, I shall concentrate on sorting out how models become bases for explanation rather than bases for prediction only. Sometimes having an explanation of a phenomenon also gives us predictive

tools. This was the case in the design of auctions. The next criterion of assumption relevance seeks to show how models become explanatory, so this is what I move on to.

E. Robustness arguments

One way to flesh out what makes assumptions relevant for explanation is to explore the derivational consequences of adopting different assumptions. What happens to the model's predictions when we vary assumptions? For example, would our model of the private-value first-price still yield bids below true valuation if we assumed differently distributed valuations, different rules of the auction or a different kind of rationality? Do the specific assumptions matter (where "matter" means "necessary for derivation")? These are the sort of questions criterion (ii) invites us to ask. In Section C. I suggested that this way of going about selecting relevant assumptions fits better when we have explanatory goals. I had in mind the following intuition: since this method allows us to understand which assumptions of the model are crucial for getting the result we are interested in, it allows us to see *why* the result obtains rather than merely predicting it. This appears to yield an understanding of the important causes or mechanisms or the structure of the phenomenon. Whether and how such an analysis is able to do so in the class of models I am discussing is the subject of this section.

Among philosophers and economic methodologists it is perhaps the most popular method for establishing a connection between the world inside the model and

the real world (Gibbard and Varian 1978, Hausman 1994, Morton 1999[9]). Perhaps the first articulation of this criterion for economics is due to Alan Gibbard and Hal Varian (Gibbard and Varian 1978). Gibbard and Varian disagree with Friedman's call to ignore assumptions of models, because researchers are often interested in extrapolating from a model applied to one situation to what happens in another, and the model's assumptions are presumably an indication of its range of applicability. An *applied* model, for Gibbard and Varian, is a model whose theoretical predicates are interpreted (more or less formally) as representing some elements of real world situations. For such a model to be explanatory its assumptions do not have to be true, however, they should approximate the real features enough for the purposes at hand. (Thus Gibbard and Varian agree with the Hausman/Morgan view of application.) The purposes at hand will inevitably vary: some models are used as *approximations* (in which case complexities of the real world should as far as possible be incorporated into the model) and some as *caricatures*. The latter, purposefully or not, distort the actual phenomena and just like pictorial caricatures can illuminate a feature of reality by exaggerating it. Gibbard and Varian's central claim is that even caricatures are able to provide understanding about what really happens in the economic world and hence explanation provided that the idealized assumptions fulfill their criterion of *robustness*. A model's conclusion is robust if it does not "depend on the details of the

[9] Rebecca Morton, a practicing political scientist (in the mathematical modeling tradition), in her book on empirical evaluation of formal models in political science claims: "Criticizing a model for having restrictive or false assumptions is vacuous unless the resulting theoretical analysis can show that the models' results hinge acutely on the restrictive and false nature of those assumptions"(1999, 144). The proposal is even stronger than our criterion for it shifts the burden of proof: Morton suggests that it is the skeptic of modeling, not the modeler, who must demonstrate that a model's predictions are not robust.

assumptions"(Gibbard and Varian 1978, 674). However, they do not specify what exactly that means.

There are two approaches one might adopt in articulating criterion (ii). On the one hand, one might insist that the false or restrictive assumptions of the model in question be replaced with assumptions that are less false or less restrictive. On this view, if having the less false or the less restrictive assumptions yields, by derivation, a qualitatively similar result then the false or restrictive assumptions do not matter. Hence a model can explain a phenomenon that satisfies only the assumptions that cannot be "relaxed" in this way. Following McMullin (1985) I shall refer to this approach as *de-idealization.* On the other hand, one might require that false or restrictive assumptions be replaced with *different* assumptions that may also be false or restrictive. If these different assumptions also lead to a qualitatively similar prediction, then this might be evidence that these assumptions do not matter. I keeping with common practice I reserve the term '*robustness analysis*' for this sort of exercise. I shall articulate these two options and explain what purposes they might serve. But before that, I would like to discuss an argument against both of the approaches.

The argument comes from Alexander Rosenberg (1992). He argues that the fact that a phenomenon follows from different theories does not do credit to this theory, according to the conventional wisdom in philosophy of science. The requirement of robustness secures the robustness of the explanandum and not the explanans, i.e. that it is the phenomenon in question that is proved to be robust, but not the theoretical result that is supposed to explain it.

This inference, however, results from a confusion between the notions of model and theory. When microeconomists say that the particular assumptions used in a derivation do not matter they normally mean that they could construct other models which describe *similar* situations but use *different* simplifications. For example, a model which makes all the same assumptions about the actors and their circumstances but relaxes the assumption of, say, risk aversion. So these new models will not be new theories (in that they will not postulate a significantly different mechanism or causal process). Rather they will be treated as a family of models that share the relevant features of their structure without sharing all mathematical details. So it is by no means clear that this family of models will demonstrate only robustness of the explanandum (i.e. the result which can be derived from these models) but not robustness of the explanans (i.e. these models).

i. De-idealization

On the de-idealization approach, criterion (ii) reads as follows:

(ii′) for the purposes of explanation ignore those assumptions that can be replaced with more realistic assumptions while preserving the qualitative predictions of the model.

The procedure is known as "relaxing" assumptions. Advocates of this kind of technique assume that a good explanation must be true or approximately true, so when false assumptions are discharged in the way they propose this is evidence that the model is capturing a real process, a process that is approximately true of whatever real world system the model is supposed to represent. If in addition to that this process can

be given a causal interpretation (this is going beyond what advocates of robustness say), then the model is able to provide causal explanation.

Ernan McMullin explains how Bohr's 1913 highly simplified model of the hydrogen atom came to be thought to capture the essential features of the atom (McMullin 1985). In proposing this model Bohr was seeking to account for the patterns of light that heated hydrogen emitted. In the model, the massive proton is assumed to be at rest and the electron to be moving in a circular orbit allowed by the quantum postulate. Bohr claimed that the pattern of light emitted is due to the electron's transition from a higher to the lower energy level, which is predicted by the model. The fact that a lot of the features of light emission were predicted correctly was an amazing confirmation for Bohr's model. However, an even greater piece of confirmation came when the simplifications the original model made were incorporated into a new more complex model (as the process of de-idealization demands). Once the simplifications were accounted for, the predictions about the light emission became even more precise! For example, Bohr's model assumed the nucleus to have infinite mass and thus not move at all. However, in fact it does move around the common center of gravity of the proton and the electron. This motion yields a correction factor for the light emission formula. When this correction was added to the prediction of the original simpler model, the predicted spectrum became even closer to the observed one (McMullin 1985, 260).

De-idealizations need to be properly motivated. For example, we know that it is more realistic to assign a proton a finite rather than an infinite mass. And we know that for reasons independent from the fact that the model predicts better when it

assumes finite rather than infinite mass. McMullin argues that if we simply corrected theoretical laws in the light of empirical evidence than these corrections will be *ad hoc* and will not add to the explanatory value of the theory (1985, 261). McMullin argues that the fact that the technique of de-idealization works is evidence that "the original model idealizes the real structure of the object"(ibid.). The "real structure" must refer to something like a specification of the fundamental or essential properties of the object. If Bohr's model is an explanation of the atom's behavior rather than merely an instrument for prediction of what sort of light it will emit it should tell us what the atom consists in and why it does the things it does.[10] That de-idealization yields progressively better prediction is evidence that the original model and the corrections pick out these essential properties. Leaving aside the question of what exactly the "real structure" of phenomena is, we can still appreciate the intuition that knowledge of this real structure is what enables us to give successful explanations. This then is the rationale for de-idealization.

Note, however, that McMullin adds the requirement that replacing false assumptions with more realistic ones should yield *better* not just qualitatively similar predictions. If it does, this is evidence that de-idealization does not produce arbitrary correction of the original model. We know what feature of the phenomenon we are idealizing, if de-idealization produces better predictions. At the very least we could ask that relaxing false assumptions should not make the model's predictions *less* correct. So the new version of the criterion should look as follows:

---

[10] A different rationale for idealization and de-idealization given by Cartwright in terms of capacities or natures will be discussed in section IV.

(ii″) for the purposes of explanation ignore those assumptions that can be replaced with more realistic assumptions while *at least* preserving the qualitative prediction of the model and possibly improving the prediction's precision.

The criterion (ii″) might be treated as giving a sufficient explanation of how economic models get applied for explanation within the Hausman/Morgan account of application as assumption satisfaction. Should it be treated in this way? The criterion can face two sorts of problems. The first concern is a *practical* one – in practice, is it the case that all the problematic assumptions of a model can be relaxed without qualititatively affecting a model's prediction? If not, then it will not remove the problem of restrictive or false assumptions entirely. So the derivational criterion can be helpful sometimes but insufficient. The second concern is a *theoretical* one – is de-idealization of some assumptions possible in principle?

To probe how practical the derivational criterion is let us go back to our simple private value auction model. Many of its assumptions can certainly be relaxed. We can extend the number of players to three, four or even n. In a first price sealed bid auction with n players an equilibrium bid is $(v_i - x)(x/1000a)^{n-1}$ where x is a bid considered by bidder i and i expects j to bid av_j (a is a parameter). It is also possible to solve this game for a *second*-price sealed-bid auction by changing the payoff function to accommodate the fact that the winner will not be paying her bid but the bid of player 2. We can also introduce common valuations (when a player's valuation depends on the valuations of other players, as it would when the object for sale was an oil well with an uncertain amount of profit potential). To do that we define a player's valuation

by reference to the signals from Nature received by her as well as the other player. Thus $v_i = \alpha t_i + \beta t_j$, where t_i is a signal player *i* receives from Nature and t_j is the signal player *j* receives from Nature. (Signals from Nature can be understood as beliefs about the probability of, say, finding oil in a given tract, or alternatively as types of players – for example, high t is an optimist about the amount of oil in well, while low t is a pessimist). α and β are to be understood as parameters players attach to the signals. If α=1, while β=0, then we have a case of private valuation (player *i* attaches no importance to player *j*'s signal). If α= β, then the player attaches the same importance to her signals as to the signals of her opponent. If t_i is uniformly and continuously distributed in an interval from 0 to 1, and if the signals t are uniformly and continuously distributed from 0 to 1, then the game has a Bayesian Nash equilibrium in which each type of each player bids $(\alpha+\beta)t_i$.

Furthermore, we can relax the assumption of uniformity and prove the existence of an equilibrium under an arbitrary distribution of valuations (Osborne 2003, 308). We can even drop continuity. If valuations are discrete, then we may assume that they can be 'low' or 'high' with a certain probability. There is no longer a pure strategy BNE. To see why suppose that the low type bids b1 and high type bids b2. b2 should not be less than the high valuation, because then there is an incentive to add a small increment and make the bid higher. Nor should b2 be equal to the high valuation because b1 plus a small increment would be better. Nor should b2 be less than b1, because then the higher valuation exceeds b1, so b2 should be higher. So instead the high bidder should randomize her bids. Note also that the low valuation should always bid the amount of her valuation. If she bids more than her valuation,

there is a possibility of a loss and if she bids below her valuation a slightly higher bid would be better.

Auction theory in fact consists of a family of models and theorems where simple games like the private value first-price sealed-bid are progressively modified by introducing new elements (for example, reserve prices) or modifying the original ones (for example, assuming risk aversion instead of risk neutrality). Whether or not the technique of relaxing assumptions to better approximate a real world situation is successful in practice depends on how rich our theoretical resources are, that is how much of the world we are able to formalize appropriately. McMullin famously stated that "mathematical idealization has worked well for the natural sciences" (1985, 254) meaning that formal treatment of the physical world has yielded impressive results both predictively and explanatorily, and that "the extent to which it *is* an idealization has steadily diminished"(ibid., original emphasis). Perhaps by induction we should expect this to happen in social sciences. Whether or not it will is an empirical question which we do not need, nor should want to answer one way or another.

The practical record of de-idealization by derivation in economics is mixed.On the one hand, formal economic theory is now more versatile than ever. Methods have been developed to formalize features as diverse as uncertainty, reputation, freedom, hatred, credibility, learning, etc. Auction theory has progressed precisely because of this. On the other hand, we do not by and large have the evidence that adding up these features improves precision of our predictions. However, testing economic models is extremely difficult and it is thus hard to draw conclusions from this fact.

A bigger problem is that de-idealization arguments are not always available to modelers. Sometimes this is because no theoretical treatment of a phenomenon exists, which means we just don't know how to make certain models more realistic. Alternatively, the very structure of game or utility theory might *prevent* incorporation of some elements into models. Francesco Guala argues that there is a formal constraint in the expected utility theory that prevents incorporation of considerations of reciprocity into games (Guala forthcoming). If he is right then the cases in which we have a reason to expect reciprocal behavior may be unexplainable by means of game theory. A further practical obstacle to "relaxing" assumptions is that once models incorporate some feature of the real world their predictions stop being specific enough to support any de-idealization argument. An example of this in game theory are Folk Theorems that demonstrate that behaviors that were not rational in single shot games can be sustained in repeated games with appropriate discount factors. These theorems are a problem for de-idealization analysis of game theoretical results because once a game is repeated (which for many cases is a more realistic assumption than single shot games), equilibria become numerous and there is no longer a specific prediction about what the agents are going to do. Yet for the sort of de-idealization in question to be sufficient for an account of model application assumption relaxation must be available and it must preserve or improve the precision of the prediction. Rational choice models tend to have a lot of structure. [11] This intricacy can be a practical obstacle to de-idealization. All these are general features of contemporary economic theory. So while de-idealization may be a sensible idea and an appropriate

[11] See Cartwright 1999b and 2002.

approach in some cases, it is safe to conclude that it is not, as things stand in today's social sciences, a sufficient condition for showing that restrictive or false assumptions do not matter.

The practical obstacles to de-idealization are important enough that we should not make application of models hostage to it. In the case I have studied – design of auctions of the basis of auction theory – the material environment of the actual auction presented many factors that were not incorporated into any auction models either because theorists did not know how or because it was too much effort. Nevertheless, the models were applied for both explanation and prediction. For this reason, I believe we need an account of application that does not make de-idealization a necessary step. However, are there factors that thwart this technique in principle rather than just in practice?

This is the more difficult question to which I shall not provide an answer in this dissertation. But here's a brief consideration of what such obstacles might be. The de-idealization approach assumes that we achieve applied scientific knowledge by building more and more complex models, models that subsume more and more reality under the umbrella of whatever theory we are using. For example, in order to provide an explanation of outcomes of a complex auction such as the one I shall describe in chapter 2, the advocate of de-idealization would have to require the building of a giant model incorporating as assumptions all the features of the auction that make a difference. That no such model was feasible in the case I studied (and in general given the state of contemporary economics and limits of computation) is beside the point for the philosophical question. However, a philosopher might argue that it may in

principle be impossible to build such a model, because every empirical situation is distinct in some way. [12] There may never be enough theory to cover the whole world, if the world is "dappled" (Cartwright 1999). But nothing I say here hinges on this.

Another "in principle" obstacle to de-idealization might be thought to stem from the mathematical nature of our theories. It is essential to criterion (ii″) that assumptions thought to be false are replaced with assumptions that are thought to be more realistic, not just some other different assumptions. But one might argue that certain assumptions of economic models are such that "relaxing" them does not make any sense, because they cannot describe any features of empirical reality. For a deduction to go through we often need assumptions that require, say, that utility functions should be twice-differentiable, or that probabilities should be uniformly and continuously distributed in a specific interval, etc. Do these assumptions map onto any empirical features of the real world? If they don't, then what could it mean to relax them? What would it mean to "relax" an assumption if it does not actually make any well-defined claim about what the world is like? Once again, this is not an argument I endorse[13] in this dissertation because it takes a controversial metaphysical stance that I am not ready to argue for. The stance is that the real properties of the world are qualitative, as we experience them, rather than mathematical. This argument also takes a controversial methodological stance – that economics lacks rules for moving from mathematical to qualitative descriptions and back via, say, representation theorems.

---

[12] In her discussion of "material abstraction", Nancy Cartwright suggests that theoretical tools will always run out, so to speak, at some point in the process of de-idealization (Cartwright 1989, chapter 5).
[13] I did endorse it in Alexandrova (forthcoming) but I now realize that I do not have a proper defense of this argument.

To sum up, I defend no 'in principle' objections to the technique of de-idealization. However, the practical obstacles are significant enough to motivate an account of application that does not presuppose de-idealization as spelled out by the criterion (ii″).

ii. Robustness analysis

There is another version of the robustness criterion of assumptions' relevance:

(ii‴) for purposes of explanation, ignore the assumptions that can be replaced with different assumptions while preserving those predictions of the model that correspond to facts that are supposed to be explanatory.

On this criterion we pick out the relevant assumptions of models (i.e. on Hausman/Morgan view, those that need to be satisfied) as whichever assumptions are shared by a series of robust theorems. One example of a robust theorem in theoretical biology is the Volterra principle. It is a result common to a number of different idealized models of predator-prey interaction, and it claims that the introduction of a general pesticide will increase the relative number of prey. Robustness analysis is just the process of discovering through many derivations this common result. Each of the models predicting the result may be idealized in one way or another. Can the robust theorem still be explanatory of the populations that satisfy only some of the assumptions of the theorem?

Jay Odenbaugh uses this criterion to argue that false models in ecology can nevertheless explain (Odenbaugh 2003). He articulates robustness as follows: "A prediction of a model is robust just in case one can (a) replace an assumption(s) A_i

with other assumptions which are logically independent of A_i and (b) show that the behavior of the dynamical system(s) with respect to that prediction remains the same or close to it across independent assumptions." (Odenbaugh 2003, 9). The requirement that the new assumptions be logically independent from the old assumptions seeks to make sure that the prediction is preserved under genuinely different assumptions. But why exactly does the fact that some particular false assumption does not matter indicate that a model can explain? There is no attempt, as in de-idealization, to progressively arrive at more and more realistic assumptions. Thus the rationale that a properly de-idealized model picks out the fundamental properties of the object in question is no longer available.

We must look for the rationale elsewhere. Although the existing philosophical literature does not provide it, we may construct one such rationale on the analogy with the argument Wesley Salmon presents as a reconstruction of the reasoning of a French physical chemist Jean Perrin (Salmon 1984). At the turn of the 20th century, Perrin sought to measure the Avogadro's number and to confirm the existence of atoms. Perrin used thirteen distinct methods to determine the number – by examining the Brownian motion of particles, the Alpha decay, the blackbody radiation etc. It is the fact that all these methods yield approximately the same number that gives such a high confirmation to the existence of a molecule and the Avogadro's number. In this case a large number of defeasible sources of evidence all pointing to the same conclusion establish the hypothesis. Salmon argues that this is an application of the common cause principle – one and the same cause must be responsible for the same result arrived at by such different methods (Salmon 1984, 221). How do we know that the

common cause of Perrin's results is the truth of the atomic hypothesis? Because it is extremely unlikely that independent causes be each responsible for yielding measurements that have such a "remarkable agreement" with each other.

We may extend this argument to robust theorems: the fact that a certain result is a robust theorem in several idealized models makes it likely that the false assumptions in these models do not matter, that the results are "driven" by the common assumptions, and that these common assumptions describe the true laws by which the system in question operates. It just is highly unlikely that the above is false given that the assumptions in many different models line up in just the right ways yielding these results. For example, the argument might be applied to the Volterra principle – the fact that it is a robust theorem makes it likely that it is a correct description of certain laws of real populations even though models that yield the Volterra principle are idealized.

This argument follows a template known as the No Miracle Argument: if some observation O is extremely unlikely if a hypothesis H is not true, then observation of O is evidence in favor of H. This template has been historically used to argue in favor of scientific realism – the view that mature theories (usually those of physics) describe the true fundamental structure of the world, rather than merely serve as instruments for prediction. Given the amazing predictive and technological successes of our best theories it would be miraculous or at least highly improbable that these theories should not get something right about the structure of the world. "The positive argument for realism is that it is the only philosophy that does not make the success of science a miracle", goes the classic statement of the argument by Hilary Putnam (1975, 73).

Magnus and Callender (2004) formalize the argument: let P(Sh) stand for probability of success of a theory h, P(Th) probability that h is true, P(Sh/Th) conditional probability of success of h given its truth and P(Sh/~Th) conditional probability of success of h given that h is not true. The No Miracle Argument is as follows: if h is likely successful (that is P(Sh)>>0), if its success given its truth is likely (that is P(Sh/Th)>>0), and if its success given that it is not true is not likely (that is P(Sh/~Th)<<1), then its truth given its success is likely (that is P(Th/Sh)>>0). Now with sufficiently high P(Sh/Th) and sufficiently low P(Sh/~Th) the value of P(Th/Sh) will be high.

Recently philosophers have noted that this argument may be an instance of a fallacy called *base rate neglect* which amounts to thinking that P(Th/Sh) *must* be high if P(Sh/Th) is high (Howson 2002). Just as we have a psychological tendency to think that testing positive on a fairly reliable test (i.e. a test that yields few false positives) for a disease therefore makes it likely that one has this disease, we think that being a successful scientific theory makes its truth more likely. The problem is that P(Th/Sh) also depends on the sample from which h is drawn, just as the probability that one has the disease given that one tested positive on a fairly reliable test depends on the actual frequency of this disease in a population from which the patient is drawn. We can only infer from success of a theory to its truth if we know the frequency of successful theories. In case of the realism debate, this requirement has the effect of complicating the realist's task enormously. What exactly is the sample? Is it all theories that have ever been advanced? But then P(Th/Sh) is likely to be very low. All theories that have

ever been successful? But then we have a biased sample (for details see Magnus and Callender 2004).

The No Miracle Argument as applied to robust theorems will suffer from the same problem. Probability of truth of h (an arbitrary robust theorem) given that it is a robust theorem depends on the frequency of robust theorems in the population of deductions from models. What might such a population be? Even if there exists a probability distribution over robust theorems (which is a big 'if'), how shall we go about selecting the models and deductions to be included? What would constitute an unbiased sample?

Note that Salmon's argument is immune to the base rate fallacy objection, because we know that *all* the methods Perrin used to test the hypothesis have yielded a similar result. Perrin, as far as we know, did not cherry-pick the evidence, including only those methods that yielded the result that agreed with others. This is why in this case the probability of the atomic theory being false is indeed low given the evidence. The agreement among the different methods is indeed remarkable because we know what the sample is. Demonstrating the same about the robust theorems is a tall order. It is not impossible but there are reasons to be skeptical.

Michael Weisberg in his recent discussion of robustness analysis in biology adopts a different line of justification (Weisberg forthcoming). He emphasizes that robustness analysis *by itself* does not deliver confirmation to explanatory hypotheses. Robustness analysis, Weisberg claims, has the power to confirm causal generalizations if we are able to interpret the mathematical dependencies in the models as causal dependencies. Mathematical dependencies are not readily interpretable as causal

because robust theorems are rife with assumptions that biologists know not to be true of real populations. However, we can gain confidence that these assumptions do not matter if robust theorems are successful at predicting the behavior of real systems and populations. We may not be able to demonstrate deductively that the falsities do not matter – that is we may not be able to de-idealize robust theorems in accordance with the rules presented in the previous section. But the fact that the model predicts correctly where we expect it to, is some evidence towards the fact that the false assumptions do not matter.

It is not clear whether Weisberg wants to claim that robust theorems with false assumptions give an adequate causal explanation of certain empirical phenomena, or whether they are merely "adequate representations" to use his own words (Weisberg forthcoming, 19-20 manuscript). He talks about causal generalizations, but also about "adequate representation" which *prima facie* is different. I think it is important to keep the two separate. Predictively successful robust theorems may indeed prove to be good representations of certain phenomena, but we need to make an extra step to put forward a causal explanation of these phenomena. Part of making a causal claim is specifying the empirical conditions under which the causal relation in this claim holds. If we do not treat all the assumptions of the model as specifying the empirical conditions under which the causal relation the model puts forward holds (because, for example, we know some of the assumptions to be false), then we need some *other* way of learning about these conditions. We can learn about these conditions independently, for example, by looking at the real populations that instantiate a given causal relation. But then the robust theorem is only part of the explanation. Exactly what I mean by

that shall become clear in Chapter 3. For now, I want to agree with Weisberg that robustness analysis as expressed by the criterion (ii″) is not sufficient for a demonstration that a robust theorem explains a particular phenomenon that satisfied some of this theorem's assumptions. Lacking recourse to a No Miracles type of argument, we need some way of demonstrating that false assumptions do not matter. Successful predictive performance is one such piece of evidence. The important point is that it is not enough to demonstrate robustness of a certain result under different assumptions to claim that this result explains some real world phenomenon.

Let us take stock. The Hausman/Morgan view of model application is satisfaction of the relevant assumptions. On this view the model is seen as making a claim of the form 'if all the assumptions are satisfied, then a given result obtains'. Most of the situations to which we want to apply models for purposes of explanation or prediction, fail to satisfy all the assumptions. So we need to show that the truth of some assumptions is irrelevant for these purposes. This is why we looked for a criterion of relevance of assumptions. If prediction is what we are after, we may be entitled to ignore assumptions that do not matter for reliable prediction. In section D. I have tried to show that this is only an option if we have some other reliable means of identifying a model's condition's of proper use. If explanation is our goal, we may be entitled to ignore assumptions that do not matter for purposes of derivation of the relevant result. In section E. I reviewed two ways of fleshing out this criterion and concluded that neither is fully adequate. This is not to say that we never should use these techniques, but rather that they are not enough.

The Hausman/McMullin articulation of the account (i.e. models as *ceteris paribus* claims plus de-idealization) shares a common presupposition that the assumptions of models unless they can be relaxed must be treated as descriptions of the empirical situations to which these models may be applied for various purposes. I think there are reasons to doubt that assumptions must be treated as making such claims and therefore must be either satisfied or relaxed. My main reason comes from the fact that in the case I have studied in depth – design of auctions on the basis of game theory – models were used heavily but not in the way that this account suggests. Scientists did not engage in many derivation exercises as de-idealization and robustness analysis propose. In particular, the fact that de-idealization or robustness analysis was not practically achievable did not mean that the model could not be applied. I consider this a motivation enough to look for another account. So let us move on to a new view of how models are applied.

III. Application by credibility

Economist and philosopher, Robert Sugden, almost entirely rejects Hausman's analysis of how economic models connect with reality (Sugden 2000). The key to his view is to deny that *the derivations from models are identical to the hypotheses economists want to entertain*. It is not the case that Morgan and Hausman explicitly defend this view. However, their insistence on assumption satisfaction as a precondition for model application indicates that they presuppose something like the claim Sugden denies. If all assumptions must be satisfied for a model to apply then the claim this model makes about the world is precisely the prediction deduced from these

assumptions. A further reason to think that Hausman endorses this presupposition is his identification of claims of economic theory with generalizations appended with a *ceteris paribus* clause, where the former and the latter are claims that economists want to entertain. That is, on Hausman's view, economic models tell us how actors would behave if assumptions of economic models were true of them. They are true in the absence of disturbing causes which economic models do not fully spell out (Hausman 1992, chapter 8).

What are the reasons to reject this presupposition? For Sugden the main reason is that it requires that models' assumptions be identified with *non-interference clauses* (i.e. clauses that describe what may disturb the result in the model) and clauses that specify under which conditions the derived generalization holds. As we have learned from the example of the private-value sealed-bid model, such identification threatens to make the claims of the model extremely restricted. Sugden worries that by Hausman's logic "we end up removing almost all empirical content from the implications of the models"(Sugden 2000, 17). Indeed claims of economists start looking like theorems – *if assumption 1, assumption 2, ..., assumption n are true, then such and such result follows*. This, however, is not how economists treat their conclusions. Rather, Sugden suggests, they treat them as claims about *the impact of one or more variables on the putative effect holding constant all other variables*. That this formulation is explicitly causal does not constitute a departure from Hausman's view. For Hausman, models, once supplemented with a theoretical hypothesis, can also be given a causal interpretation (Hausman 1992, 51). The difference is that under

Sugden's interpretation 'holding constant all other variables' does not correspond to the assumptions necessary to derive the impact.

Of course, just the fact that economists treat their conclusions as empirical hypotheses rather than theorems does not mean philosophers should do too. Economists may be wrong and overly optimistic. Sugden knows that and, along with a new view of models, seeks to provide a *justification* for regarding deductive results of models as empirical claims about the impact of one variable on another holding everything else fixed.

The problem of justification cannot, Sugden argues, be solved without going outside the realm of the model and deduction. It is not by investigating the properties of the model that we come to acquire confidence in the empirical hypothesis that it suggests. The inference from a result holding in a model to a result holding in the world must be inductive. If within the model C causes E, then this relationship can be inferred to be in operation in the real world in a number of ways. Both C and E may occur in the world, in which case the model may give us reason to believe that in the world C causes E. Alternatively, E may occur in the world and we may infer that C causes it as an inference to the best explanation. Whichever model of justification gets used the induction from the processes in the model to the processes in the world needs substantive reasons. Where could these reasons come from?

Sugden reviews two existing lines of justification. The first one is a Millian line that takes conclusions of models in economics to be causes that combine just as causes in mechanics do. For Mill phenomena are mechanical if causes are additive in a way similar to vector addition in Newtonian mechanics where we know what each

force contributes on each occasion because each one of them has a stable impact independent of the circumstances. For example, we know that the stable impact of gravity is expressed by Newton's inverse square law. On this view, hypotheses suggested by models are justified as statements of causal forces operating in the world by virtue of the additivity. The models tell us about causal impact of one variable on another in the absence of other causes; but if causes combine additively, then we are warranted in asserting that the causal relationship in the model is still operating in the world, additively combining with other causes. That phenomena in political economy are mechanical in this way seemed to be, for Mill (at least on the first reading of Mill), an obvious truth. Yet neither he, nor contemporary economics, provide any reason for believing that the conclusions of economic theory do behave in this convenient manner. Hence Sugden rightly rejects the Millian justification.

The second line of justification is the aforementioned "robustness" criterion. It requires that the assumptions of models be shown to be non-essential to the result of the model by demonstrating that other assumptions would lead to the same result. Sugden's reasoning here seems to be as follows: even if some assumptions can be disposed of in the manner that the robustness analysis requires, this sort of reasoning makes inferences from one model to another model or a class or models, without allowing us to step outside the model into the world. The hypothesis which we want to entertain on the basis of the model is *stronger* than just the deduction from the model. It is stronger in the sense that the model's assumptions are not exhaustive of

the conditions under which the relation described by the hypothesis holds.[14] So the robustness discovered by derivation is not a good justification either.

Sugden's preferred justification invokes the notion of *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). "Could" here is not meant as a proxy for possibility, but rather as a proxy for similarity. 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 world works, we come to view the world in the model as credible, that is similar to the portions of the real world. I'll quote Sugden at length:

> On this view, the model is not so much an abstraction from reality as a parallel reality. The model world is not constructed by starting with the real world and stripping out complicating factors: although the model world is simpler than the real world, the one is not a *simplification* of the other.
> Credibility in models is, I think, rather like credibility in 'realistic' novels. In a realistic novel, the characters and locations are imaginary, but the author has to convince us that they are credible – that there could be people and places like those in the novel. As events occur in the novel, we should have the sense that these are natural outcomes of the way the characters think and behave, and of the way the world works. We judge the author to have failed if we find a person acting out of character, … these things couldn't have happened. We can praise a novel for being 'true to life' while accepting that every event within it is fictional, as when we recognize aspects of its characters as typical of people we know. When a novel has this form of truth, we can use it to explore 'What would happen if …?' questions, in something like the same way that economists use models. (Sugden 2000, 25, original italics)

[14] This is also a view endorsed by Cartwright 1989, 191-192.

A more precise answer as to what constitutes credibility in economic models is spelled out by *coherence* (Sugden's term). Coherence requires that assumptions "fit naturally together", which means, Sugden explains, that a modeler should not use an "arbitrary mix" of assumptions, just as a novel's characters should not act "out of character"(Sugden 2000, 26). Otherwise, the results "cannot be seen to follow naturally from a clear conception of how the world might be"(ibid.). Presumably we evaluate the model's coherence by tapping into our background knowledge of the sort of causal processes that operate in the world and how they normally operate. A model must respect some aspects of this background knowledge.

One of the examples Sugden works through is Thomas Schelling's model of segregation in housing markets. The model assumes that there are two groups of people (for example, black and white people) who live in a neighborhood represented by a grid, such that one square is one family house. Assume that no family wants to live in a house if the majority of their neighbors are of different colour than the family (assume families to be monoracial). So the families move to different squares on the grid according to this principle. After a few iterations we get concentrations of the same color on the grid, i.e. neighborhoods segregated by color. The empirical hypothesis is that racial segregation may arise even in the absence of racial prejudice – the desire to live among families of your own color. All that is needed is the desire not to be a minority in your own neighborhood.

What makes this a credible model of housing market segregation? For Sugden it is the fact that the elements of the model bear similarity to the real people, neighborhoods and preferences in such a way that enables us to judge that the

processes described in the model *could* be true of our world. It is credible to assume that people would want to avoid being a minority. It may not be credible to assume that neighborhoods are grids, but this particular assumption does not matter (by robustness analysis). What is important is that the overall set of assumptions and the conclusions they underwrite exhibits the sort of credibility that is vindicated by our background knowledge. In a footnote Sugden suggests that one way to recognize credibility is to be able to imagine that the sort of actions people take in a model or in a novel could very well happen to you (n 21, p.30). Once again 'could' here is meant to indicate not just logical or physical possibility, but rather strong plausibility.

For a model's hypotheses to be credible, its assumptions do not have to be satisfied or even approximately satisfied by any particular situation in the real world. Rather they should be "adequately *representative*" (Sugden 2000, original italics). Sugden likens representativeness in a model to a similar notion in literature: a novelist might call a middle-aged Englishman from Ipswich 'Frank', which would be more representative than calling him 'Duck Bill Platypus' (ibid.). Credibility is weaker than truth or approximate truth, but nevertheless carries empirical warrant.

The most important question to ask of Sugden's analysis is what exactly credibility is good for. I chose the focus of this dissertation to be use of models for policy in a way that enables explanation and prediction. Sugden talks about "confidence" (2000,1) in economic models. If Sugden and I are talking about the same thing, then his account is a genuine contender. If not, then credibility analysis accomplishes some other goal. Which is it?

Confidence in models can mean many things, even if we confine ourselves to confidence with respect to its empirical virtues. We might be confident in models as policy tools, or confident that they state some causal claims that can be used to develop explanations. In the former case, we know that a particular model describes a causal relation that will hold to an appropriate degree in an environment in which we might rely on it to devise institutions or plan for the future. In the latter, we know that the model is good enough to be admitted in the pool of models that we can tap to serve as bases for explanation and maybe more. Obviously, the latter case is a case of much stronger confidence, because this condition is easier to satisfy and it does not commit us to much. It merely indicates that we have some reasons to think that this model could, given that it passes appropriate tests, become explanatory or predictive.

Sugden's analysis is best interpreted as giving us the latter, not the former, kind of confidence, and I wish this was more clear from Sugden's writing. What exactly does credibility analysis accomplish and is it a significant accomplishment? It is not clear from Sugden's paper whether having subjected the model to credibility analysis we are entitled to conclude that the quantity X impacts the quantity Y holding constant other variables. This interpretation is supported by the fact that he calls credibility analysis a 'justification' of models, i.e. a way of showing that model's claims are reasonable. But how could we possibly draw such a conclusion just by tapping into our background knowledge? Or maybe the credibility analysis allows us to conclude that it is merely credible to believe that quantity X *could* impact quantity Y in such a way, but this is much weaker and hardly informative. If credibility is significant, it is just one of the first hurdles a model must pass to earn admission into

the toolbox of empirical economist. Many more hurdles lie on the road of becoming an actual explanation or a policy tool.

A further problem is that it does not seem to be very difficult for a model to jump the hurdle of credibility. What models in economics haven't passed this test? Whether they do, seems to depend a lot on who judges. Background knowledge involved in the judgment of credibility is a large and variable resource. In particular, background knowledge of theoretical economists will involve their professional intuitions about the sort of processes that take place in the social world. There is evidence that these intuitions are responsive to just how much economic modeling one does. For example, economic experiments indicate that students who have studied economics conform to predictions of economic theory more than students who haven't (Marwell and Ames (1981), Frank, Gilovich, and Regan (1993)). A plausible interpretation of these data is that serious study of game theory leads its students to internalize its principles and view certain portions of the social world in terms of non-cooperative games (these games are taught and studied more than the cooperative ones). In this case, background knowledge that gets used to pass judgments of credibility will differ among economists and non-economists, and it is not clear whether the judgments of economists are due to their education or to their superior experience of the social world. So we should be careful about drawing conclusions on the basis of judgments of credibility.

Sugden is not, on my reading, offering an account of model application, i.e. an account of how models explain or predict. Nevertheless, his account is valuable because it invites us to investigate an option previously unknown or unarticulated: that

the strict content of the model – its assumptions and deductions – is not identical to the empirical hypothesis that the model suggests. This means that *we may not know what exactly the model claims about the world just by examining the model*. This is a claim about the contents of the model. It also has consequences for the further issue – how do we know whether the model applies, i.e. whether the empirical hypothesis the model suggests about a particular phenomenon is justified. If the empirical hypothesis the model suggests is not identical to the strict content of the model (i.e. its assumptions and derivations), then we cannot settle the issue of justification just by examining the model's assumptions.

This idea is not new. Recall that Friedman did not want us to use assumptions to determine the conditions under which a model applies. Assumptions can be suggestive but not definitive guides on this matter. I think the point is a valuable one. Even better we do not need to accept Friedman's brand of instrumentalism to take advantage of this point. We can believe both that assumptions should be evaluated and also that they do not uniquely specify the material environments to which models apply.

Another source of the same idea is Mary Morgan's discussion of models and stories. For Morgan, the model says nothing about the world until we append it with a story. The story engages the model into an explanation with a beginning, middle and an end. It is the story that fills in the abstract structure of the model in a way that allows for making a causal claim.[15]

---

[15] If so, then perhaps it was wrong to impute to Morgan a pure assumption satisfaction view of model application. A more accurate interpretation of her view is that she takes assumption satisfaction to be

To clarify what exactly the claim is let us distinguish it from a related but a different and well-known idea. Generalizations in special sciences are thought to come amended with a vague *ceteris paribus* clause (Hausman 1992, Fodor 1987 and many others). The vagueness indicates that we cannot fully specify the conditions under which the purported generalization holds. Some even think that in any science "above" physics and chemistry we should not expect to have well-defined properties that correspond to the *ceteris paribus* conditions (Rosenberg 1994). If we could fully specify these properties, we would have universal laws. For Fodor, for example, the distinguishing characteristic of the special sciences is that the "exceptions to the generalizations of a special science are typically inexplicable from the point of view of (that is, in the vocabulary of) that science" (Fodor 1987, p.6). For example, when a macroeconomic generalization fails we may be able to find the cause of this failure in the concepts of macroeconomics. However, often it will fail because of some psychological, or biological or even chemical fact that is outside the domain of macroeconomics. This idea is usually justified by appeal to the notion of *supervenience*: higher level facts such as economic facts are supposed to supervene on or vary along with the more metaphysically basic facts from psychology and all the way down to physics. This is one sense in which theories may fail to specify the conditions under which its generalizations hold.

I have in mind a different sense that does not appeal to supervenience of higher level facts on the lower level ones. Just by examining the model's assumptions we

---

necessary but not a sufficient condition of model application. Her view may be more accurately described as requiring that some assumptions be satisfied (to give some warrant that the story is a correct one), but that does not exhaust justification of the story.

may fail to fully specify the conditions under which the putative causal claim in the model holds in the world, even if we restrict ourselves to conditions that fall within the purview of this theory. Thus an economic model of an auction may fail to specify all the *economic* conditions under which this model applies. Indeed, this is only to be expected if model building does not serve only one goal – that of figuring out the conditions under which certain relationships hold. Models are also judged on their parsimony, mathematical tractability, elegance etc. Because the conclusion in models I focus on follow deductively, they do, as a matter of logic, contain a sufficient set of sentences from which the conclusion follows. I wish to explore the possibility that *not all* of these sentences *must* be treated as describing conditions under which the conclusions of models actually hold in reality. They can be treated in this way, but they don't have to be. Once we accept this fact, we have a more liberal view of how we can apply models for empirical tasks. I introduce these ideas now informally and shall spell them out properly in Chapter 3.

The claim that the models themselves do not exhaust the content of the hypothesis put forward on the basis of the model is consistent with a number of accounts of model application, except for Hausman's account. By itself this idea does not commit us to any particular interpretation of the sort of claims models make and how these claims come to justify policy interventions. My account, which I shall put forward in chapter 3, shall assume and further defend this claim. Now I am moving on to a final view of model application, that also endorses this claim but nevertheless is different from the account I eventually adopt.

IV. Capacities and Concretization

Nancy Cartwright is another philosopher of science who recently revived a Millian approach to economics and also physics (Cartwright 1989, 1999). Unlike Hausman, she interprets Mill as advocating a metaphysics and a science of tendencies, rather than *ceteris paribus* laws. Tendency laws are stable features that produce regular effects in the absence of disturbing factors, that is in idealized circumstances. However, even when other tendencies or disturbing factors are present tendencies are still exercised, which is what allows us to export our knowledge of tendencies from idealized and controlled conditions of the laboratory or a model to the real world. So tendencies are not limited by a *ceteris paribus* clause. Cartwright calls causal tendencies 'capacities' (1989, ch 4). Capacities are characterized by their stability. For example, when we say that negatively charged bodies have the capacity make other negatively charged bodies move away, we do not just mean that they make other move away in certain ideal or near ideal conditions, but rather that the capacity is exerted across a number of environments, even where it fails to manifest itself by actually moving the other body away. However, there are also cases when a capacity fails to even exercise itself. One such case is a case of interaction when another capacity or a disturbing factor neutralizes the original capacity. For example, if I am given a slow-acting poison and then shot to death with bullet the shooting neutralizes the capacity of the poison to kill me.

Modeling, controlled experimentation, statistical methods used in forecasting and other scientific activities pursued in many modern sciences, presuppose, Cartwright argues, an ontology of capacities (1989, ch 5). When an econometrician

makes a demand model which relates quantity demanded to price, the assumption is that price has a stable and measurable effect on the quantity demanded, an effect that is preserved across a number of changes in the environment. Applied to modeling, such as the economic modeling I am discussing, the insight is as follows: we build models to investigate the canonical behavior of a capacity. For example, on the basis of the model I discussed in the beginning of this chapter we may conclude that the first-price auction has the capacity to lower bids under the conditions of private values. In models we idealize away the disturbing factors to allow the capacity of interest to manifest its 'pure' behavior (1999 chapter 4). This is one function of models in many aspects of scientific practice and in many disciplines ranging from physics to econometrics.

These views carry with them an account of model application. The account, on my reading, is as follows: a model applies to a phenomenon, where 'applies' carries a license for explanation and prediction, when the model plus corrections and additions to it identify the relevant capacities and disturbing factors underlying the phenomenon of interest. The process of application, Cartwright argues following Polish philosopher Leszek Nowak, is the process of *concretization*. Law claims, such as agents are perfectly rational or Newton's second law, do not, for Cartwright, describe behavior of real material phenomena. If they are useful in the way many such law claims are useful in science, they express abstract principles which are best viewed as statements of capacities. They are abstract in the sense of describing the behavior of a capacity "out of context" (1989, 187), i.e. without stating the *ceteris paribus* conditions that accompany it in the model. Concretization involves adding back the factors omitted in

the model but present in the concrete situation we wish to explain using our knowledge of capacities. Different theories teach us how to do that by providing us with *rules of composition* of capacities:

> …in the simultaneous equation models of econometrics it is supposed that the equations governing each tendency separately must be satisfied at once; in electronics there are familiar theorems that reduce complex circuits of capacitors, inductors and resistors to basic ones that satisfy known equations. In game theory various concepts of equilibrium tell us how the self-interested tendencies of different agents will combine to produce a series of moves (Cartwright 1995, 179).

In her other writings Cartwright argued that theoretical laws and concepts of physics do not apply to real material without a complex set of approximation and correction techniques (Cartwright 1983). If these approximations and correction techniques describe other factors present in the situation of interest, then it becomes clear why these techniques are not arbitrary. It is the metaphysics of capacities that ensures that the process of concretization is not ad hoc.

What is the difference between the account of application as satisfaction of assumptions and Cartwright's account on in terms of concretization? Cartwright's account does not require that the model's assumptions be satisfied or even approximately satisfied by the target system for the model to be of explanatory value. Instead, it requires that models state capacities. Given that capacities are stable enough to allow us to move from what is true in a model to what is true in a messier real world, models can provide explanatory insight even while their assumptions fail to be satisfied. Of course, the various factors we introduce during concretization must correctly describe the disturbing factors.

Cartwright's account is also able to make sense of the claim that the exact derivation from a model is not identical to the empirical hypotheses researchers entertain on that basis. The following discussion of abstract laws, such as the Coulomb's law, which Cartwright reads as capacity ascriptions makes this point:

> …one looks for what is true in an ideal model in order to establish an abstract law. But there is a difference between what is true in the model and the abstract laws itself. … The first is a kind of *ceteris paribus* law: it tells us what the factor does *if* circumstances are arranged in a particularly ideal way. In the abstract law the ceteris paribus conditions are dropped; and for good reason since this law is meant to bear on what happens in the more frequent cases where conditions are not ideal (1989, 191-192).

We have been unable to find an acceptable articulation of the satisfaction of assumptions account of application. Is Cartwright's account the right alternative? I will argue that it isn't. To be covered by Cartwright's account models such as those in economics must be read as making capacity claims. But this is a tough test to pass. The epistemic standards for capacities are high. There are both philosophical and empirical reasons to give a different reading of economic models, and I shall have much to say about it in chapter 3.

V. Conclusion

In this chapter we examined three accounts of model application. Two of these – Hausman's and Cartwright's – strive to give sufficient conditions for how models can explain. Sugden's, on the other hand, attempts only to give a necessary condition. Both Hausman's and Cartwright's accounts impose demanding conditions on model

application. For Hausman models only explain if some special set of its assumptions are satisfied. I have articulated a rough characterization of such a set for purposes of prediction. As for explanation, I have argued that the two characterizations designed for this purpose – de-idealization and robustness analysis – are not enough. If Cartwright's account is to explain how models apply, then we must be able to interpret models as making tendency or capacity claims. But capacities are supposed to have a special stability which is also a demanding condition.

Can models be applied even if they fail Hausman's and Cartwright's conditions? In the case I shall discuss in the next chapter, they were. This case is often cited among the most successful example of application of microeconomic theory to date and it also appear to be typical. Having explained how models were used in designing an auction, I shall try to put forward a new account of models and their application. One of its features will be that neither Hausman's, nor Cartwright's conditions need to be satisfied for successful use of models.

Another feature of the new account will be based on the lesson we learned from, among others, Sugden. In the heyday of Popper, Lakatos and Milton Friedman, economic theory was thought of as formulating hypotheses about laws governing what actually happens in economic world. The assessment of these hypotheses necessarily had to take place outside economic models, i.e. in the field. Then in the early 90s Hausman and Cartwright urged the return of a Millian view of economics. For some this view implies that models make claims about what would happen if *ceteris paribus* clauses were satisfied. This shifted the assessment away from the field where ceteris paribus clauses are almost never satisfied to examination of models' assumptions.

However, if Sugden, among others, is correct, then models do not contain a full specification of the claims they make about the world, so their assessment cannot proceed by just examining the model's assumptions and their satisfaction. So a desideratum for an account of application is that this account must be sensitive to the fact that models do not contain ready-made claims of the type: F is true under such and such conditions.

These are problems I hope to solve with a new account of models as "open formulae' and their application as "material realization". Some of the evidence in favor of this new account will be philosophical and some empirical. I start with the latter. Chapter 2 follows with an account of spectrum auction design on the basis of game theory.

Chapter 2

How Game Theory Gets Applied for Policy: The Case of Auction Design

In this chapter I tell a story of a famous and perhaps one of the most successful examples of application of microeconomic theory for policy – spectrum auction design. The goal is to not so much to give a detailed story of the process by which these institutions came into being. But rather to see what sort of knowledge underlies the main decisions about auctions' features? How does this knowledge relate to economic theory? The hope is eventually to read an account of model application off this story. Although some of the clues as to what this account might be will become apparent in this chapter, I will postpone explicit theorizing till chapter 3.

I. What are spectrum auctions and why are they significant for philosophy of science?

The radio spectrum is the portion of electromagnetic spectrum between the frequency limits of 9 kilohertz and 300 gigahertz. These bands are used for communication for both civilian and government purposes, and this requires careful regulation. In the USA, for example, the radio spectrum needed for non-governmental purposes – all telecommunications for private and business use including media – is administered by the Federal Communications Commission (FCC). Given guidelines from the government the FCC has to decide on an efficient mechanism for distributing licenses that allow its holders to use certain bands of the spectrum.

Certain licenses, in particular those used by radio and TV broadcasters, go to their users for free. For a long time this was the case for nonbroadcast licenses as well – they were awarded on the basis of a special comparative hearings in which potential users had to demonstrate the "public interest" of the proposed enterprise. (The public interest clause is the original legal requirement of spectrum use from 1920s.) Since comparative hearings tended to become highly politicized, in the 80s the Congress authorized the use of lotteries. But since the 1993 Omnibus Budget Reconciliation Act passed by US Congress, the FCC has the right to use competitive market mechanisms such as auctions to allocate nonbroadcast licenses. (Broadcasters' powerful lobby managed to exclude TV licenses from this requirement, so the networks, among others, still get spectrum for free.) The Balanced Budget act of 1997 even requires FCC to do so unless exceptions apply.

What might such mechanisms be? In academic literature, especially law and economics, the use of auctions for assigning spectrum property rights is an old idea. The consensus seems to be that the it was first suggested in the 50s by a Chicago Law student Leo Herzel and a few years later by economist Ronald Coase (Coase 1993). Although the two wrote independently they were both inspired by A.P Lerner, a Russian-born London School of Economics economist, who explored various ways to improve socialism with market-based institutions (Herzel 1998). The idea was met with much opposition because it was thought that the radio spectrum is a public good which was taken to mean that it should not be treated by a market procedure (Coase 1998). In the 80s this view was challenged as academics proposed auction designs, i.e. a market institution, that could allow government to enforce the "public interest"

standard. Eventually the idea got adopted and implemented thoroughly during the tenure of Reed Hundt, the FCC Chairman appointed by the Clinton administration (Hundt 2000). From 1994 to 1996 the FCC has conducted seven auctions and allocated several thousands of licenses for wireless communications.

Now the auctions are a standard tool. There is much debate about how exactly they should be implemented but not much questioning of the general idea of using them to allocate spectrum rights. Moreover, spectrum auctions became a commonplace subject for academic discussions. Economics journals publish 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. Indeed the FCC website itself hosts records of papers and presentations on the topic by eminent scholars. Arguments appealing to economics became part and parcel of the process by which the spectrum policy is worked out.

This indeed is what makes spectrum auctions interesting from the point of view of philosophy of science – the wide and systematic participation of economic theorists and experimentalists in the design of these institutions. For example, one of the rules in auction implementation is that several months in advance of an auction the FCC should solicit public comments on possible design features. Many academics hired either by prospective bidders or the FCC itself respond with recommendations. The constraints on design are set by the government requirements, or rather by the FCC's interpretation of these requirements. For example, in 1994 these requirements

were: efficient and intensive use of the spectrum, promotion of new technologies, prevention of excessive concentration of licenses and finally insuring that some licenses go to minority-owned and women-owned companies, small businesses and rural telephone companies (McMillan 1994).

In addition to that the auctions have the potential to raise billions of dollars of revenue for the government. With much money and credibility at stake, the precise features of spectrum auctions must be taken very seriously. Exactly what rules of auction will reliably produce the desired outcome is a formidable puzzle for teams of economic theorists, experimentalists, lawyers, software engineers and policy-makers. Auction design takes place at the intersection between theory, informal intuitions, observation, policy debate and, of course, politics. For a philosopher of science, the methods used to solve, or at least to probe, this puzzle shed valuable light on the status of economic models as well as the challenges and process of developing and justifying applied knowledge.

The main theoretical tool for developing spectrum auctions is auction theory – a field at the intersection of microeconomics, general equilibrium theory and game theory which studies aspects of various price-setting mechanisms such as auctions. I thus begin by describing basic elements of auction theory, especially those that get invoked in auction design. Then I move on to describe how the theory was put to use in the USA in 1994, while also making references to the UK and European auctions in the following decade.

II. Auction Theory[16]

Before William Vickrey revolutionized the field by applying game theoretical tools to auctions, operations research literature advised that bids for different auctions be estimated statistically from data on past bids, and then the probability that a particular bid wins should be maximized on this basis (by trading it off against the cost of the bid). This approach obviously fails to incorporate one crucial fact – that auctions are strategic environments. That is bidders' behavior depends not just on their past behavior, but on what each believes the others will do and on the rules governing each interaction. The ability to model this feature explicitly makes game theory a particularly attractive tool. Now bidding is studied as a strategic action affected by the rules of the game and the nature of the bidder's valuations and reasoning.

Auction theory is best viewed as a series of models and an exploration of logical relations between these models in the form of theorems. Theorists start by developing a typology of auctions: the ascending-bid (in which the price is going up), the descending-bid (with price going down), the first-price sealed-bid, the second-price sealed-bid, etc. They then make simplifying assumptions about the number of bidders, units for sale, properties of bidder's valuations, their attitudes to risk and pick an appropriate equilibrium concept. Since a player's access to information about other players' information and intentions is often at stake Bayesian Nash equilibrium is regularly used. In it players pick strategies as a function of their own information and maximize their expected payoff given other players' strategies and their beliefs about

[16] This exposition relies on Klemperer 1999 and on Milgrom 1989.

other players' information. In some auctions, players know their own valuation of the product (private value models), in others they don't know it for certain and can learn something about it from other players' values (common value model), and in yet other models values are partially a function of some exogenous variable (affiliated value model). Some models incorporate the seller's reserve price, some just assume it to be zero. Based on all such features the equilibrium bidding strategies are worked out, and theorists can see how bidders' behavior depends on the types of auction and various elements of its design.

The private value first-price auction I discussed in chapter 1 is a standard example of such a model. In that example, we proved the existence of a unique Bayesian Nash Equilibrium in which bidders bid half of their valuation. The interpretation is that a first-price auction (i.e. an auction in which the winner pays the amount of the highest bid) gives bidders incentives to lower their bids. Now consider a different kind of auction – descending, or Dutch – named so because it is used in Holland to sell flowers. The auctioneer starts by calling a high price and then lowers it gradually until a buyer accepts it. The first buyer to stop the auction wins and pays the price at which the auction stopped. Since in a Dutch auction, the winner pays the price she stopped the auction at and no bidder has an opportunity to come back into an auction after just one bid, the situation is equivalent, as far as bidding strategies are concerned, to the first-price sealed bid auction.

Now let us consider ascending auctions, in which the price starts low and is raised until no bidder wishes to raise it any further (since the bidding increment is thought to be vanishingly small ascending auctions are modeled as second price).

Assuming independent[17] private values (i.e. no motivation to learn from your opponents' bids) what should the bidder bid? Clearly it is rational to stay in the bidding so long as the price is below your valuation and not all other bidders have opted out. When the next to last bidder has dropped out the bidder with the highest valuation will get the object for sale for the amount proposed by the second to last bidder (plus a small bidding increment). This strategy is called truth-telling because it recommends bidding the amount equal to your valuation.

A similar truth-telling strategy is a weakly dominant[18] strategy in case of a second-price sealed-bid auction. To confirm, consider a player i with a valuation v_i. Let the highest bid other than yours equal b. If $b > v_i$, then bidding v_i results in 0 (because the other player wins) and bidding more than v_i results in loss (because you paid more than your valuation). If $b < v_i$ then bidding v_i results in winning the object at $v_i - b$ (which is a surplus), while bidding less yields less. Finally if $b = v_i$, your gain is 0 no matter what. So in each case bidding your true valuation is at least as good as any other option. However, once we introduce common values (in the form of dependencies in the distributions of valuations), the equivalence between second-price sealed-bid and the ascending auctions breaks down since the latter auction allows players to update their valuation by learning from other players' bids. Truth telling now consists in bidding the bidder's expected value conditional on the information that will be revealed if he wins.

---

[17] Values have to be independent in the statistical sense, that if players' valuations are random variables their joint distribution is equal to the product of their marginal distributions.

[18] A strategy is weakly dominant if pursuing it leaves the player at least as well off as any other strategy and in some cases it may leave him better off.

Indeed this is why a possible feature of auctions with common valuations is the so called *winner's curse*. In the environment of pure common valuations, players are thought to have the same valuation but different information about the possible value of the good for sale. If there is uncertainty about the future revenue potential of a given spectrum license, then some players' 'signals' as to what this potential might be is higher than others. Even if all bidders make unbiased estimates of the value of the license (that is estimates that are on average equal to the real value), the highest estimate tends to be biased upward. The winner then could be the one whose estimate of the license value is higher than average. Thus the winner will on average experience negative profits. Fully rational players adjust their estimates accordingly, i.e. downward.

The above analysis showed a strategic equivalence between first-price sealed bid and Dutch auctions, and also some equivalence between second-price sealed bid and ascending auction. What about first-price and second-price auctions? How do they compare? A set of central results in auction theory – Revenue Equivalence Theorems – show that under some conditions these two kinds of auctions are too equivalent with respect to the profit they generate for the bidders and the revenue they bring for the seller. Some intuition for this result is the fact that in both type of auction winners tend to pay below their exact valuation. In the first-price auction this fact is due to the incentive to lower bids, in the second-price auction the winner pays the amount bid by the next to last bidder which is again less. The surprising theoretical result proved in the 80s indicates that these two effects will balance each other out exactly. Here is its statement:

> Assume each of a given number of risk-neutral potential buyers of an object has a privately-known signal independently drawn from a common strictly increasing, atomless distribution. Then any auction mechanism in which (i) the object always goes to the buyer with the highest signal, and (ii) any bidder with the lowest feasible signal expects zero surplus, yields the same expected revenue (and results in each bidder making the same expected payment as a function of her signal). (Klemperer 1999, 232)

All the discussed auction models in equilibrium award the object to the bidder with the highest value, i.e. a bidder's probability of winning is just the probability that her valuation is the highest. Similarly, in all these models the bidder with the lowest valuation has no chance to profit, so, the theorem says, all these mechanisms will yield the same economic result. Note also that the result applies to common as well as private values. Above it is expressed more generally: each bidder receives a signal (not directly her value) from a common distribution. If she can learn something from this signal (for example, by updating her valuation using clues from other bidders' behavior), then we are in a common-values environment, otherwise, values are private. The result can also be extended to cases in which more than one object is for sale.

However, revenue equivalence cannot be proved once we leave the simple world of independent draws from strictly increasing atomless distributions. Independence of signals and specific properties of distributions are necessary assumptions for the revenue equivalence results. It means that if players receive information from non-independent sources – that is their valuations are affiliated – revenue equivalence does not follow. This is one reason why the theorem has not

been influential in empirical work on auctions, and almost entirely irrelevant to issues of auction design (Klemperer 200b). Take, for instance, the empirical fact that ascending auctions are by far more common than sealed-bid auctions, or the fact that real subjects in economic experiments do not treat theoretically equivalent auctions the same, or the fact that FCC spectrum auctioneers ended up adopting features of both in their design. If these decisions are at all reasonable, then there are important empirical differences between types of auctions. Some of these differences may be captured by economic analysis and some may be psychological. Economists have been able to show that ascending auctions are superior, as far as the seller is concerned, if the demand for objects is elastic or if bidders' information is correlated rather than independent. But other, more mundane, factors such as how complex an auction is, how easy it is to ascertain that the seller isn't cheating and bidders aren't colluding may be additional explanations for the fact that real people do not treat different kinds of auctions as equivalent (Milgrom 1989, 17-18).

The more recent and advanced auction theory contains attempts to relax the restrictive assumptions in the simpler models discussed above. One aim of the exercise is analysis of *optimal* auctions. Given a certain set of conditions, how do/should sellers and buyers rank different mechanisms? By *mechanism* auction theorists mean a formal game defined by the set of allowed bids, the allocation rule from bids to probabilities of each bidder winning the auction, and a payment rule (Krishna 2002, 62).

One restrictive assumption is risk neutrality. It is reflected in the way outcomes are weighed in an expected utility function. For example, in the first-price

sealed bid model we wrote down this function [$EU_1=(v_1-x)Prob(b_2<x)+ 0Prob(b_2>x)$ where (v_1- x) is the payoff from winning the auction and 0 is the payoff from losing it] as linear assuming risk neutrality. However, if a bidder is risk averse, that is if she derives more utility from a sure outcome than from a lottery that provides this outcome with a probability between 0 and 1, then this function will be concave (with an extra weight on a surer outcome). In a second price or ascending auction this fact will have no effect on bidding – whether or not one is risk averse it is optimal to bid up to one's true valuation. In a first-price auction, on the other hand, a risk averse bidder has an incentive to bid more aggressively to ensure winning. So a risk-neutral seller faced with risk-averse buyers prefers the first-price mechanism.

Another central assumption is independence of bidder's information. Obviously, this assumption breaks down as soon as we allow bidders to inform their valuations by seeking to find out what other bidders know. This very natural allowance introduces correlations between bidders' valuations, or, to use technical vocabulary, their valuations become *affiliated*.[19] Although, just as in the models above, it is still the case that the bidder with the highest valuation wins and the one with the lowest evaluation gets zero profit, affiliated values allow the auctioneer to extract more marginal profit from high valuations.

To see why, note that when signals are affiliated, a higher private signal indicates to the bidder that her competitors also probably have higher signals. In the first-price auction, this means that she will win less often, which diminishes her

[19] Bidders' valuations are affiliated if a high value of one bidder's signal makes it more likely that other bidders' signals are also high.

expected profit. In the second-price auction there is an additional effect– winning costs more because the higher her own value the higher is the value of the runner-up (which is the price she'll have to pay if she wins). So as far as a profit-seeking seller is concerned affiliation is good, and second-price is superior to first-price. Furthermore, an ascending auction is superior to a second-price sealed-bid in case of more than two bidders and common values, because in this case the price paid depends on even more bidders' signals.

So a seller concerned with revenue ranks second-price sealed bid above first-price and ascending above second-price sealed bid. The intuition behind this ranking is that a buyer's profit depends on her having private information. The more affiliation there exists the less she is able to exploit her private information and the less profit she extracts. This is known as the Linkage principle, according to which affiliation favors sellers at the expense of the buyers. The seller thus has an incentive to reveal as much of his own private information about the good for sale as possible, because that on average will increase seller's surplus by linking bidders' valuations to an exogenous variable. Assuming there is no possibility of systematic misinformation on the seller's part there is no incentive to hide information from bidders.

Furthermore, we can relax the assumption of symmetric bidders, that is that bidders valuations are drawn from a common distribution. Once again, it is natural to want to relax it since we expect real world bidders to have different rather than identical profiles. Theorists have been able to get interesting results: "roughly speaking, the sealed-bid auction generates more revenue than the open [second-price] auction when bidders have distributions with the same shape (but different supports),

whereas the open auction dominates when, across, bidders, distributions have different shapes but approximately the same support"(Maskin and Riley 1985 qtd. in Klemperer 1999, 236).

Note that these results do not discuss any specific psychological or economic asymmetries between bidders, only different mathematical properties of their valuation distributions. And it is this feature that makes auction theory difficult to test in the field, more so than other theoretical domains that use resources of game theory. Auction models almost invariably postulate distributions of valuations – this is how uncertainty about other bidders' beliefs is formalized. Moreover, distributions are postulated to have certain properties, such as shape and continuity, without which probabilities cannot be converted into numbers, which is in turn required for solving the games.

In experimental settings economists are able to control these properties by assigning valuations to experimental subjects in ways that identify the distribution. However, in the field these are unobservable. Even if some observations can be made, the data underdetermine a distribution's properties. This makes it impossible to know what model we are in given a particular empirical case. Usually theoretical models are tested by constructing a structural econometric model which is then estimated from data. However, if more than one distribution is compatible with observable bids, then the model is not parametrically identifiable, i.e. we cannot confirm or reject this model. Of course, distributions are not the only unobservables auction theory models use; asymmetry of bidders and mixtures between common and private valuations also introduce similar complications. As such we do not yet have a way of testing auction

models in the field (Laffont 1997). This implies that when experiments are used to test a particular auction design to be used in the field, the external validity judgment is more complex.

More development of original models is possible and currently being done, but this much of auction theory is enough for us to get on with the story.

III. Background to the FCC auctions

The FCC spectrum auctions began during the FCC chairmanship by Reed Hundt. Hundt was a practicing lawyer who went to high school with Vice-president Al Gore and to law school with President Bill Clinton. When the newly elected administration appointed him as the FCC Chairman in 1992, he began to work closely with the vice-president on communications policy. Gore viewed the communications issues as key to his broader policy objectives. These were no less than to bring about an information revolution: deregulate the communications market, jumpstart innovative technologies capable of bringing about genuine social change and empowerment and make them available to the broader public.

Just how revolutionary Gore's team took these new technologies to be can be gathered from Reed Hundt's memoirs of his time as the FCC Chairman. Firstly, the book called <u>You Say You Want a Revolution</u> recounts a story of successes and failures to reform the FCC and the related industries in the most fundamental way since their inception in the beginning of 20th century (Hundt 2000). Perhaps even more revealing is one episode among the many described by Hundt, when in all seriousness he was lecturing Gerry Adams, the leader of the Irish republican party

Sinn Fein, (who came to the US hoping to have a meeting with Gore, but instead had to make do with Reed Hundt) that internet and cellular phones can bring peace to Northern Ireland.

Gore's team treated the new communication technologies as an opportunity to create a new grass root participation in politics, thereby weakening the effect of the alliance between the big communication firms and the Republican party who relied on them for campaign contributions. Gore's initiative to bring internet to every public school in the country was one example. In meetings with his team he often brought up the dream of giving a schoolgirl in Carthage, Tennessee, access to the full collection of the Library of Congress via what he called the "information superhighway".

Another way to weaken the traditional GOP donors would be to deregulate the existing communications markets such as broadcasting, local and long-distance telephone services and cable television, by letting new entrants to take some of the industry's market share. The 1934 Communications act in effect created a series of communications sector monopolies. These provided a number of economic (in addition to the political) reasons to reform the system. The old monopolies were not conducive to development and implementation of alternative forms of communications technologies. Nor were they seen to provide services at competitive market price. However, just how much deregulation should there be was strongly debated.

Take for instance the television market. In this case the carriers themselves were all in favor of a full free market approach. However, in return for greater freedom, Reed Hundt wanted to extract from the businesses a promise to carry three

hours of educational TV programs every week. He saw this as a way of fulfilling the "public interest" clause of the original 1920's legislation which gave spectrum to broadcasters for free. The large TV networks, on the other hand, argued that by broadcasting television without charging consumers they are already doing more than their share of "public interest" satisfaction.

The political background of many debates between the FCC and the industries was similar. Deregulation was by and large seen as a good thing by both sides, but Hundt as a part of Gore's team wanted to derive benefits in accordance with some traditional Democratic causes. When time came to distribute spectrum for wireless Personal Communications Services the compromise between deregulation and "public interest" amounted to a requirement that along with economic goals auctions also should distribute spectrum to small rural, women- and minority-owned companies.

When the spectrum auctions were finally authorized by the 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 the Congress. The FCC, as well as having to rethink its whole approach to spectrum, would lose its bargaining power to extract concessions from the industries. In addition to the political stakes, there were considerable economic stakes. If the auctions went well they could raise billions of dollars for taxpayers.

IV. Why auction design is important

We already saw from the brief discussion of auction theory that auction outcomes depend on auction features. However, the most vivid way to see why auction design matters is by looking at examples of failures of actual, not just theoretical, auctions.[20]

The New Zealand government adopted a second-price design with no reserve price (i.e. no minimum bid) to auction off some broadcast and non-broadcast licenses in the 90s. The results were deeply embarrassing, because the high value of the licenses was widely known and publicized, and yet the government did not earn any money off it. Perhaps the most illustrative was the case when a student won a license for a small town TV station by bidding $5, but paying nothing since nobody else submitted a bid (no doubt enriching himself afterwards by selling the license).

The Australian case was similarly a failure. The 1993 auction for satellite television licenses was first-price sealed-bid, but with no deposit and no specific payment policy. An unknown bidder outbid the big players such as Rupert Murdoch (which government could give a good spin to), but 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.

More recently, failed auctions for Universal Mobile Telecommunications System (UMTS) licenses, otherwise known as 'third generation' or 3G licenses, were held across Europe. In Netherlands, for example, five licenses were up for bidding in an ascending auction in July 2000. The problem, however, was that there were also

[20] These true stories come from McMillan 1994 and Klemperer 2002a.

exactly five incumbents (i.e. established telecom companies) and each could win no more than one license. But if the government were to raise any revenue from this auction there had to be some competition. When one entrant company, Versatel, tried to challenge one of the incumbents, Telfort, in the fight for a license, the latter threatened to take the former to court. Telfort argued that Versatel did not really believe that it could win a license, so any bidding Versatel would engage in would be clearly aimed at raising the cost for Telfort, for which Versatel would be held responsible in court. Versatel complained to the government but nothing was done, because without Telfort the auction would be even more of a failure. As a result the Dutch government raised over three times less revenue than was predicted. Although the government set such weak goals that even this auction has achieved them, many economists agree that the Dutch auction could have been designed much better (van Damme 2004)

The Swiss auction offered four licenses for sale in November 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 will just pay the very low reserve prices. So the government tried to postpone the auction only to be taken to court by the four companies because it did not specify beforehand the right to cancel an

auction in these circumstances. So valuable licenses in one of the richest countries in Europe were sold for one-fifteenth of what the government hoped for. (Wolfstetter 2003).

So what exactly is required to get an auction right? (By 'getting an auction right' I mean creating a reliable and controllable institution, and since empirical knowledge is my focus, not necessarily a morally justifiable institution.) This question is crucial if we are to address the philosophical issues that motivated us in the first place – what sort of knowledge underwrites applied economics and what is its relationship to economic theory? Since the US auction was considered a success, I shall start with an assumption that whatever US auction designers did to make the auctions run in the way they did is good methodology. But what exactly is this methodology?

It would be convenient if the methodology could be read off uncontroversially from what the actors involved in auction design said and did. This was a hope I entertained up until I started to look seriously into the records left by the protagonists of the story. The hope did not survive because I found not one but a number of different narratives of auction design written from different perspectives. Any historian will, of course, naturally expect that to be the case. After all, the FCC spectrum auctions were a huge politically sensitive phenomenon that put at stake fortunes of the FCC, the Clinton administration and telecom industries and reputations of many economists.

Can my question still be answered in the face of all these different points of view and stakes? I believe it can, partly because my focus is very specific. I am not

trying to get a comprehensive account of what forces molded the FCC spectrum auction policy, or how the auctions looked from the perspectives of all participants. Rather, my focus is strictly methodological – I hope to draw up a sketch of what pieces of knowledge made it possible to run it in the way that it was run.

Given this focus, I shall concentrate on the methodological narratives that can be divided in roughly two categories – those of theorists and those of experimentalists. By 'narrative' I mean an account in which a researcher justifies the decisions about the design of the auctions that ended up being taken. These narratives are reconstructed on the basis of some of the written records I had access to. I must acknowledge at the outset that mine was not a comprehensive historical investigation – I did not interview the main protagonists and relied mostly on the articles they wrote for the audience of their fellow economists. The disadvantage of this approach is that it requires one to treat the protagonist's own polished stories as objective. However, there are virtually no other records (certainly not in the official FCC documents) of the methodological issues apart from these articles. And I have done my best to crosscheck the narratives against each other.

It is by getting clear on how these two groups treated theory and experiments that we can get a grip on the method underlying auction design. Academic commentators on spectrum auctions do not all neatly divide into theorists and experimentalists. In fact, most of the eminent scholars in the field dabble in both theory and experiments and respect both sources of knowledge. The distinction between the two types of narratives is thus somewhat idealized, but helpful nevertheless.

The focus of all these narratives is the precise design of individual auctions. The major pieces of this design – and hence the two main areas of decision-making about the auctions – are the rules of participation and its material environment. Rules are the explicit and public instructions covering entry, bidding and payment all participants received and studied before the auction. The material environment of the auctions includes such features as the software, the venue and timing of the auction, and whatever features of legal and economic environment the auction designers could control. So before we move on with the narratives justifying the decisions to adopt particular rules and a particular environment, it will be helpful to briefly describe what decisions actually ended up being taken.

V. The Simultaneous Multi-Round Auction

Geographically the country was subdivided into 51 Major Trading Areas (MTAs), which in turn were subdivided into Basic Trading Areas (BTAs). BTAs numbered 492, each of which had four spectrum blocks. There were thus more than 2,500 licenses to be won all around the United States.

Also known as the Simultaneous Ascending Auction, the Simultaneous Multi-Round Auction puts all licenses for sale *simultaneously*, as opposed to sequentially, and in an *open* rather than sealed-bid arrangement. Bidders place bids on *individual*, as opposed to packages of, licenses they are interested in and when the round is over they see what other bids have been placed. Then the next round begins in which bidders are free to change the original combinations of licenses but must increase their bid up to the level of the highest previous round bid plus a prescribed increment

if they wish to hold on to a license. The process continues until no more bids on any license are received.

The bidding is regulated by a number of further rules:

-Activity Rules: to remain eligible for bidding the bidder must maintain a certain activity during each round. The extent of required activity increases with each round. For example, a bidder must maintain 50% activity (i.e. bid on 50% of all the licenses she had announced an intention to own before the auction) in round 1, 70% in round 2 etc. If she fails to maintain such activity her eligibility to bid in further rounds is reduced. However, each bidder also gets five waivers that allow 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 specifies the minimum bid increment (between 5 and 20%), which depends on bidders' behavior. The more bidding is going on, the larger the increment. Similarly increments decrease with the bidding activity.

-Designated Entities (that is companies owned by women, minorities and/or small businesses) get 10-40% credits on specific licenses.

-Spectrum Cap: a firm cannot own more than 45MHz of broadband spectrum in any geographical area, hence cannot bid on more than the corresponding number of licenses.

-Payment Rules: the FCC requires an upfront payment (from which penalties are deducted if the bidder withdraws, or which gets refunded if the bidder fails to win the license she was interested in). If a license is won 20% of its price is paid within five days after the auction closes and 80% within five days after the license is officially awarded.

This is a very brief summary of the main rules of the auction. Their full statement takes over 130 pages (FCC 1994b). So how was the adoption of some of these rules and not others justified?

VI. John McMillan and Preston McAfee

McMillan and McAfee are both economic theorists (at the time working respectively at UC San Diego and UT Austin) who became actively involved in the FCC auction design. McMillan was hired by the FCC itself, while McAfee by a potential bidder Airtouch Communications. They are also the authors of the first two articles dedicated wholly to the FCC spectrum auctions that appeared in the *Journal of Economic Perspectives* in summer 1994 and then winter 1996. Most academics who have heard of the auctions heard its story from either McMillan or McAfee, or both.

What is interesting about the story told by McMillan and McAfee is its emphasis of the role theory played in settling the major questions about the features of the auction. Theirs is the story that comes closest to the ideal type of the theorist's narrative.

From McMillan and McAfee we learn that the spectrum auction design is all about application of game theory and that game theory deserves the major credit for its success. The first paragraph of their joint article "Analyzing the Airwaves Auction" reveals just 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". Further on they quote *The Economist*'s witty comment: "When government auctioneers need wordly advice, where can they turn? To mathematical economists, of course…", and *Fortune* which described 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 as saying "Game theory, long an intellectual pastime, came into its own as a business tool" and finally *Wall Street Journal* "Game theory is hot"(all citations from McMillan and McAfee 1996 159).

What exactly was hot about game theory according to McMillan and McAfee is 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"(ibid., my emphasis). In McMillan's earlier article "Selling Spectrum Rights" he tells us "Theory helped to answer key questions of auction design"(McMillan 1994, 147).

This is not to say that McMillan and McAfee accorded *all* the epistemic credit to theory. It is worth quoting McMillan at length:

> Theory has limits. Any nontheorist will readily assent to this proposition: but it is useful to be specific about where those limits lie. First, while theory can identify the relevant variables, it cannot tell us much about their size. Theory sometimes shows that there are effects that work in opposite directions from each other, and data are needed to establish which effect is likely to be dominant. Second, transaction costs arise in carrying out some of the policies that the theory recommends, and these costs may swamp the theoretical benefits of the policies. Third, implementing a recommendation of the theory may require knowledge that is unavailable. In particular, some of what auction theory identifies as optimal seller strategies depend on the distributions of bidders' valuations, which were not known. Fourth, the theory does not specify an unambiguously best form for the spectrum auction, which is so complex that no existing theorem covers it. (McMillan 1994, 151)

Let us see how these limitations played out and what were some of the key questions to which game theory provided answers?

A. Open or Sealed-Bid?

The first issue resolved by theory is whether the auction should be open or sealed-bid. Recall the winner's curse – its effect is strongest when there is a great uncertainty about the value of the good for sale, i.e. the common value environment. Valuations of spectrum licenses should have both private and common elements, because firms have their own unique constraints and circumstances but also cannot be sure about the profitability of the investment they are about to make. So firms who have billions of dollars at stake and have their own private game theorists to tutor them should be expected to bid slowly and cautiously so as to avoid overestimating the true value of a license.

Theory says that open auctions reduce the force of the winner's curse because it allows bidders to observe each other's behavior and on that basis update their

valuations. Also as I discussed in section II, because these valuations are affiliated (i.e. they are informed partly by the same external circumstances such as business potential of cell phones), the buyer has less private information and the seller is able to extract more surplus.

So some models recommend open auction for auctions where revenue generation and efficiency is important. However, there are also models that rank first-price design higher, in particular if bidders are risk-averse (I explained why in section 2). Intuitively it is also clear how an open auction could encourage collusion and thus stifle competition. Two firms which made a secret agreement to reciprocate by keeping the price of certain licenses low, can observe whether their colluder is sticking with her promise in an open auction, but not in a sealed-bid. Collusion requires enforcement which open auction allows. So from the point of view of the seller who wants to avoid collusion this is the second drawback of open design.

McMillan explains that auction designers judged the two drawbacks to be weaker effects that the main advantage of open auction. He emphasizes that this decision required a judgment call rather than a neat theoretical demonstration. This was a case in which theory identified two effects but did not specify which one is the strongest. That, he argues, is an empirical rather than a theoretical matter.

What exactly does McMillan mean here? Presumably, it is possible to construct a model in which bidders are both risk-averse and have common valuations. Economists know how to formalize both of these elements: risk-aversion can be written into the utility function and so can common values. We could “play” with this model assuming first open and then a sealed-bid environment, and thus see whether

open auction would still give higher revenues. That could give a theoretical answer to the question about which effect is the strongest.

This is not an option McMillan considers. Does he have a good reason? On the one hand, it would be helpful to know just how much the effect in question is responsive to the extent of common valuations. On the other, there are reasons to be skeptical that this exercise would answer the question one way or another. We would have hard time knowing if this complex model were applicable, because it would make assumptions about valuation distributions, form of utility functions, the proportion of private and common elements in bidders' valuations, etc. How would we know if these assumptions are satisfied not just in a laboratory but in the field? It is difficult to imagine. Plus we would need to assume that there is such a thing as a typical bidder with such and such utility function and such and such valuation. If not the model becomes even more complex. Even if this model is solvable, we would not be sure how to draw empirical conclusions from it.[21] If this is right, McMillan believes that sometimes theoretical exercises cease to be useful for practical purposes. At this point empirical estimation has to take over. For him, theory was able to support a claim that open auction increases revenue by reducing winner's curse, but not the claim that it increases revenue even when bidders are risk averse and

[21] A different question is whether there is a fact of the matter to ascertain here. Is there an exact proportion of private and common value in a person's preference? The manner in which we write down the utility function certainly assumes that. Recall that in common value games we model players' valuation as $v_i = \alpha t_i + \beta t_j$ where where t_i is a signal player *i* receives from Nature and t_j is the signal player *j* receives from Nature. If $\beta>0$, then we have a case of common values. This formulation assumes that the relation between my own valuation and another person's valuation is additive. Do we have good reasons to think that human desires function that way? One the assumption satisfaction view of application we are committed to believing that unless this question can be answered, auction theory is unusable. This is further motivation to depart from this account.

collusive. McMillan does not say why but we could surmise that the former does not involve as many controversial assumptions as the latter.

Still, however, it is not clear from McMillan's story how exactly theory together with judgment was able to provide a strong enough reason to go with the open design. On what basis was the learning effect judged to be stronger than the effect from risk-aversion and possible collusion? McMillan does not say.

Other auction specialists emphasize that the advantage of the open auction is that it creates a *dynamic*, as opposed to a single shot sealed bid system (Crawford, personal communication). Much experimental evidence shows that one way to move prices and distributions closer to the equilibrium is to give people many chances at learning. Real world bidders are just unlikely to bid rationally the first time. The FCC spectrum auction, however, was a one time affair. So to ensure efficiency it was important to create a dynamic auction that gave bidders plenty of opportunity to receive feedback about their bids and pushed the process towards the equilibrium. However, this would not happen in the absence of activity rules, since bidders then have no motivation to place serious bids until the very end. These are more experimentally driven reasons in favor of a particular kind of open auction. But they do not enter McMillan's story.

B. Simultaneous or Sequential?

The second issue theory helped to resolve was whether the auction should offer all licenses simultaneously, or sequentially, i.e. auctioning them off one by one. 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 case of spectrum licenses aggregation is important. A company wants to have a bunch of licenses that would work efficiently together. So if these licenses are sold one by one the bidder must make guesses about the future prices of other licenses in order to make reasonable decisions about how much to buy in the present. That buyers have to make such often uninformed bets is a disadvantage if the seller intends not simply on raising revenue but also, as in the FCC case, to ensure that licenses get distributed efficiently, i.e. to those who value them most. A second disadvantage of a sequential mechanism is that it eliminates the possibility of backup strategies thus making the auction less flexible. So information revelation and flexibility are the theoretical advantages of a simultaneous mechanism.

However, just imagine what it would be like to have 2,500 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 turns 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. Stopping rules needed to meet three requirements: they "must (1) end the auction in a reasonable time; (2) close the licenses almost simultaneously, to aid license aggregation; and (3) be simple enough to be understandable by bidders."(McMillan 154) Likewise, penalties for bid withdrawal were instituted to ensure that bids be sincere (which is how they reveal

information), but not too crippling so that backup strategies were still available. As for concerns about complexity of the auction, McMillan tells us, auction designers judged the problem manageable and hoped for the best.

It is not clear that in this case theoretical arguments were treated any more seriously than the practical ones. To realize the theoretical advantages of a simultaneous mechanism, amendments motivated by very practical concerns were required.

C. Packages or Individual?

I shall discuss a final area of decision making where theoretical arguments were important according to many, including McMillan. Should bids for packages of licenses as opposed to bids for individual licenses be allowed?

There is a very real property of spectrum licenses that motivated economists to consider package bids – that is the licenses' complementarity. Two or more licenses are complementary if owning one increases the value of another. For example, a company might value a Minneapolis license higher if it already owns Chicago but lower if instead it owns Atlanta. McMillan explains why: it is easier to develop a customer base in adjacent geographical areas, it is easier to manage interferences, to establish roaming, etc (1994, 150). To use economist jargon, the licenses are neither perfect substitutes, nor perfect complements (the Atlanta license is not completely worthless to a firm that owns the Minneapolis license). Complementarities are a real headache for economists because they create a danger of there being no competitive equilibrium at all; that is, there are models on which complementarities lead to cycles

of bidding, losses for bidders, and no decentralized efficient distribution emerges as it should (Plott 1997). Nevertheless complementarities provide a *prima facie* reason in favor of package bids.

McMillan tells us that this argument was in conflict with what theory says. He presents us with a model, due to McAfee, in which package bids (or nationwide bids in this example) lead to an inefficient distribution of licenses:

> Imagine there are three bidders, E, W and N and two licenses, East and West. E values owning East alone at $2billion, West alone at $1 billion, and the nation at $3billion. Firm W values East at $1 billion, West at $2 billion, and the nation at $3 billion. Firm N values East at $1.6 billion, West at $1.6 billion, and the nation at $3.3 billion. Suppose first that nationwide [i.e. package] bidding is not allowed. In the two open auctions, W wins West at a price slightly above $1.6 billion (which is the price at which the second-last bidder, who happens to be N, drops out), and E wins East at slightly above $1.6 billion. This is the ideal outcome, in that licenses go to their highest-value users. The total value of the licenses to their new owners is $4 billion, of which the government gets a little over $3.2 billion. Now suppose there is a national auction in addition to the separate auctions. Anticipating winning the nationwide auction, N has no incentive to bid on individual licenses. Only E and W bid for them, so the price in each stops at slightly more that $1 billion. All three bidders compete in the national auction. Firm N wins the nationwide bid at slightly above $3 billion, which is where both E and W drop out. The nationwide bid of $3 billion exceeds the sum of two separate bids, so the nationwide bidder N wins. This is not the efficient outcome: the total use-value is $3.3 billion and the government revenue slightly above $3 billion. The free-rider point is that E and W would have to raise their nationwide bids by a total of $1 billion to beat N; but neither alone will raise its bid this much, as that would mean bidding more than value.(McMillan 1994, 156, n10)

In fact this argument is similar to the one made by Paul Milgrom and Robert Wilson. They were hired as consultants by Pacific Bell and sent a note to FCC arguing that package bids open the possibility of free-riding and hence inefficiencies.

In the example above the inefficiency is created by the fact that bidder N, whose valuation exhibits some complementarity, is allowed to make a package bid. Theorists such as McMillan, Milgrom and Wilson treat it as a demonstration that there is a potential for inefficient distributions if package bidding is allowed. The model does not establish that this is what will happen in an actual FCC auction, but gives a reason to worry about such a possibility.

However, this argument is inconclusive, for there are also models that give inefficient results precisely because package bids are not allowed. Peter Crampton presents one such example:

> …suppose there are two bidders for two adjacent parking spaces. One bidder with a car and a trailer requires both spaces. She values the two spots together at $100 and a single spot is worth nothing, the spots are perfect complements. The second bidder has a car, but no trailer. Either spot is worth $75, as is the combination; the spots are perfect substitutes. Note that the efficient outcome is for the first bidder to get both spots for a social gain of $100, rather than $75 if the second bidder gets a spot. Yet any attempt by the first bidder to win the spaces is foolhardy. The first bidder would have to pay at least $150 for the spaces, since the second bidder will bid up to $75 for either one. Alternatively, if the first bidder drops out early, she will "win" one license, losing an amount equal to her highest bid. The only equilibrium is for the second bidder to win a single spot by placing the minimum bid. The outcome is inefficient, and fails to generate revenue. In contrast if package bids are allowed, then the outcome is efficient. The first bidder wins both licenses with a bid of $75 for both spots (Crampton 2002, 8 manuscript).

Above is an example that demonstrates precisely the opposite point – that package bids can be the only efficient alternative. So how can this debate be resolved?

Note that neither of the models pretends to be a complete model of the spectrum auction FCC was trying to design. We do not know the exact extent of complementarities in the actual world and so cannot judge which model is closer to

reality. So they are used as arguments merely to the extent that they state plausible possibilities. This means that to be taken seriously these possibilities need to be substantiated. It is not clear that at the time of decision making on the FCC spectrum auction they were substantiated. At the time McMillan seemed unaware of Crampton's model, but instead gave a further reason why package bids should not be allowed. This reason is the complexity of package bids. Just with 492 BTA licenses there are more than 10^{148} possible combinations. This, it was judged, was bound to create dangerous administrative problems both for bidders and the government (McMillan 1994, 157). In this case, complexity was seen as a more serious obstacle than in case of simultaneous vs. sequential debate. Writing eight years later Crampton agrees that despite a theoretical advantage of package bids under strong complementarities (this is the main assumption of his model), its complexity is forbidding (Crampton 2002, 9 manuscript).

Nevertheless package bidding is still officially considered by the FCC as an option (it is listed as one of the two official auction designs on its website), especially when licenses exhibit strong complementarities. The debate continues and it is not clear that the decision to disallow package bids even had a strong justification at the time of design. Theory supports both package and individual bids, depending on the extent of complementarities. The exact extent of complementarities is unknown. So arguments from sources other than theory must account for adoption of one design feature over another.

I conclude this discussion of McMillan and McAfee's narrative with two observations. Their rhetoric emphasizes the role of theory in decision making about

the spectrum auctions. Yet once we look at the details we can see the strength of McMillan's own admission that "theory has limits"(McMillan 1994, 151). It is not clear, even from their own account of the auction design, that theory did more than to suggest possible issues to take into account when considering various design features (more on that in the next chapter). This is no small contribution. However, it is less than one would infer from the opening sentences of McMillan and McAfee's articles, where we are led to believe that at least sometimes, in particularly in the open vs. sealed-bid issue, theory actually has direct policy consequences. Much more on that later.

However, the above is just a rhetorical problem. The real problem is that it is not clear from this narrative that the decisions taken were indeed justified. We learn about theoretical arguments, practical considerations and judgment calls that were at play. But we get no sense of what exactly justified these judgment calls. Now let us move on to a very different narrative.

VII. Charles Plott

Charles Plott is a prominent experimental economist from Caltech interested in design of market institutions. He worked on the FCC auctions from fall 1993 through the fall of 1994, and recounts his experience in a paper "Laboratory Experimental Testbeds: Application to the PCS Auction" in an issue of the *Journal of Economics and Management Strategy* specially dedicated to spectrum auctions (Plott 1997).

Plott approached the task from the perspective of experimental economics, and brought its methods to bear on the problem. He describes experimental economics as a discipline in its own right, not just a tool to test economic theory. The idea is to create simple laboratory environments to study decision-making processes of people acting in these environments. Monetary incentives are used to motivate subjects to participate and make an effort to take the best decisions. Many aspects of these environments, such as the subjects' characteristics, the information they receive, the rules within which they act, etc. are very carefully controlled. The results are compared with theory as well as with many other similar experiments. These simple laboratory environments – or, as Plott calls them, *experimental testbeds* – are treated as prototypes of the more complex real life economic situations. Applied to auctions, the goal is to compare the experimental results with theoretical predictions and also to understand why an auction works the way it does (Plott 1997, 607).

What was the role of experiments according to Plott in designing the auction? Plott identifies three distinct stages at which experiments proved vital (1997, 608). However, these stages are too coarse-grained and do not neatly map on to the *epistemic* roles experiments played. So along with Plott's three stages I shall identify seven epistemic functions. At this stage my goal is just to document the ways in which experiments proved important. Once that is done I shall seek to show how each account of application I reviewed in chapter 1 fails to explain one or another aspect of the methodology used in auction design. But that is a task for Chapter 3. Obviously this strategy assumes that Plott's methodology, or rather my interpretation of it, is a correct methodology. I shall justify this assumption in section 8.

A. Stage 1

At the first stage that Plott identifies experimental testbeds were used to test "broad aspects of the rules"(1997, 608). A number of experiments were run in which subjects with induced preferences participated in prototype auctions with different rules. The questions experimentalists asked included: can the open auction really improve efficiency? Does the option to bid on packages create an advantage for bidders with complementary preferences? Does the opportunity to withdraw a bid improve efficiency? These questions are all explicitly general, in the sense that they do not tie the effects in question to the particular environment of the experiment, and explicitly causal. In the light of Cartwright's metaphysics of capacities, the temptation is to interpret the experiments as testing capacity claims. Although experimentalists undoubtedly would have liked to discover such results, this is not how they treated the experiments at the first stage. Rather the results of the first stage were read as supporting material possibilities, i.e. claims that a given causal relation *can* get realized. That these relations were causal was shown by varying one aspect of these prototype auctions while holding fixed everything else and observing the change. However, the stability of these causal relations was not explored.

Plott describes these experiments as providing "observations" rather than "support" (1997, 614). In this respect a legitimate question arises: is what sense did these experiments 'test' anything? And what was their use?

I think their value consisted in the following. Though Plott's 'observations' did not confirm or disconfirm any theory or capacity claim, they allowed researchers

to see that there exists at least one material environment in which a result originally derived from a mathematical model or supported by a less formal theoretical intuition holds. The piece of knowledge established in this way is a relatively weak "can" claim. But why aren't models such as those described by McMillan and Crampton enough to establish such claims? Why bother with the experiments? My answer is that there is an epistemically significant difference between a 'can' derived from a model and a 'can' claim established by a material experiment. A more precise explanation will have to wait till next chapter. For now I just want to register this as the first[22] epistemic function of experiments, and flag this function as an explanandum for an account of theory application. So function one of experiments was to establish material possibilities.

Another role of experimental testbeds was to enable Plott's team to discover phenomena that were not predicted or explained by theory. More precisely auction outcomes, it is discovered, were very sensitive to features of auctions that, Plott says, have no theoretical explanation. In a sealed-bid auction in which the price is rising continuously and bidders drop out until only one remains, a design known as a Japanese auction, experimentalists discovered a tendency to stay in, rather than drop out, just to drive up the price for the competitor, even though it entails the risk of 'winning' the unwanted item. The interesting fact is that the bubble created in that way is even bigger if bidders have access to information that other bidders are still 'in', than if no such information was available. Plott says that there "seems to be no

[22] I list the epistemic functions of experiments in chronological order reflecting stages of auction design, not in the order of importance.

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, n1). I read Plott as saying that theoretically perfectly rational bidders on average have correct expectations about the behavior of others and extra information should not make systematic difference, but in the experiments the information did make such a difference.

In the above case, experiments taught Plott facts that could not be known just from theory. But importantly these are facts that we *expect* the theory to predict, because these are facts about calculating agents that find themselves in situations which can be accurately enough represented by the concepts of game theory. Whether the agent stays in the game or drops out is just such a fact that game theory is supposed to represent. These are facts at the level at which game theory operates. There were also experimentally discovered extra-theoretical facts that were more concrete than regular game theoretical facts and thus theory should not be expected to yield them. These are facts about implementation of the auctions, for example, that there are a number of ways in which information may be concealed. In game theory concealment of information is a 'success' term, meaning that if the description 'information is concealed' applies, then players have no access to information whatsoever. But once we try to implement an experiment that instantiates a game theoretical model, we discover that there are different ways of concealing information that can have their own effects on real players. Plott explains:

> 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, my emphasis).

Another example of this sort of extra-theoretical knowledge is from Paul Klemperer, an economist who devised the British 3G auctions is 2000. Theory teaches us that if efficiency is a goal then it is important to encourage entry into auctions and to fight collusive behavior. For Klemperer these are the only theoretical facts that matter for purposes of auction design. The 'high power' auction theory such as the equivalence theorems is, on the other hand, quite useless (Klemperer 2002b). But these results only have a policy implication if we know what specific material environment qualifies as entry-encouraging and collusion-defeating. These are not trivial pieces of knowledge. As the Dutch government learned it is not enough to have more bidders than licenses to encourage entry. You also need to make sure that the incumbents do not intimidate the challengers as in the case of Telfort and Versatel. Plott's work shows that such qualifications can get very detailed and specific down to the software.

So the second and third function of experiments are as follows: they enabled Plott's team to discover two kinds of facts that are not explained or predicted by auction theory that existed at the time – facts that are in conflict with theory because they describe features that are explicitly represented in the theory (function two) and

facts at the level more concrete than the theory and that are necessary for implementing theoretical concepts in an operational environment (function three).

The latter facts are emphasized in Cartwright's account of application as conctretization. Concretization, recall, requires that we add the features, that the target system has but that the model abstracts away, back into the model. The task in this case is to find a material environment that qualifies as, say, information-concealing. Exactly what features need to be 'added back' in this way depends on each concrete situation. When engineers construct lasers, one of Cartwright's case studies, they exploit a fundamental principle of quantum physics – inversion in a population of atoms has the capacity to amplify a signal. That inversions could amplify a signal has been known to theoretical physicists for a while. However, exactly how it does so and how we can invert a population of atoms to create a signal is the sort of knowledge that is much less universal than the theoretical principles. It needs to be refined for each context of application. This knowledge goes beyond theory in that theory does not tell us "mundane facts" about the properties of materials and their reactions to each other that are necessary to actually create a laser (Cartwright 1989, 207-209): "A physicist may preach the principles by which a laser should operate; but only the engineers know how to extend, correct, modify, or sidestep those principles to suit the different materials they may weld together to produce an operating laser"(1989, 211).

Plott, as a social engineer, too attached much importance to auction implementation. When one reads theorists' account of the auction design various strategic features of the institution are thought to be its 'key' features. For example, for McMillan the "key questions" of auction design were whether to use open or

sealed-bid, first-price or second-price, package or individual license bids, etc. All these features also happen to figure in auction theory models. So a theorist takes the key questions to be those which get picked out by theoretical models. When Plott describes the auction, its key features are those picked out by models *plus* its operational details. Indeed the first 'key' characteristic of the FCC according to Plott is that "the rules were implemented electronically with decentralized bidders" (1997, 608). Plott's judgment about what's 'key' about an auction is implementation-centric so to speak. This is by no means trivial, since it is by implementing material testbed auctions that he learned that blinks of screens and clicks of keys do matter!

A final role of the experiments at the first stage was to expose the variety of uncertainties and dangers connected to various design features. The gist of the problem was, what Plott calls, "the sensitivity of the behavioral characteristics of the auction process to the environment in which it might be operating"(1997, 621). One interpretation, to use the language of tendencies and capacities, is that the effect of some experiments was to show that certain features of the design that theoretical economists proposed as achieving such and such goal did *not* have the stability we would expect tendencies or capacities to have.

To give an example, recall that spectrum auctions were assumed to involve some complementarities – that is that the bidders value certain licenses more when they are both available and less individually. Recall also that since the business potential of the new technology is not entirely known, values were assumed to be partially common, i.e. correlated with each other and uncertain, rather than private. In environments with common values, winner's curse is known to occur. For a theorist

such as McMillan this fact meant that an open auction must be adopted instead of a sealed-bid. The intuition behind this thought is that when bidders see each other bid, they update their valuations by observing each other and thus bid less cautiously. This prediction is very sound theoretically – rational players dealing with uncertainty should take advantage of the information dissemination that occurs in the open auction. The only factors theorists expected to work against the information disseminating effect were the possibility of collusion and risk-aversion, which were judged not to be strong enough to worry about. There was also experimental evidence that open auction with public information prevents winner's curse better than sealed-bid designs (Kagel and Levin 1986). It is striking that while McMillan treats these results as an argument in favor of an open design, Plott remains entirely agnostic. In commenting on winner's curse he says: "How this might work out when there is a sequence of bids and complementarities is simply unknown. No experiments have been conducted that provide an assessment of what the dimensions of the problem might be"(1997, 626). Here Plott is making two claims: first, that tendency of the open auction to raise bidding activity, if it exists, might be counteracted or even neutralized by sequential bidding and complementary values, and, second, that we do not know how much that might affect the overall auction. The root of the disagreement, on one reading, is that McMillan takes the information-disseminating effect to be a stable tendency while Plott doesn't.

This deep caution about how stable theoretically derived results actually are is highly pronounced in Plott's description of auction design. To quote another striking

expression of his uncertainty, here is his discussion of the dangers that lurk when valuations exhibit complementarity:

> It is well known that nonconvexities and superadditive [complementary] values can destroy the existence of the equilibrium in the competitive model and can also cause instabilities. However, very little is known about what might happen in actual markets with these properties, and during the early stages of rule-making, *nothing* was known about the behavior of the particular rules ultimately adopted by the FCC.(1997, 621, my emphasis)

What are the sources of this radical uncertainty? Plott talks of these uncertainties as generated by "interactions" between different rules and specifications of the auction (1997, 628). One of the examples Plott gives are the waivers. The original rules contained provisions to maintain bidding activity – you must bid on a certain percentage of the licenses you expressed an interest in, or else your eligibility goes down. However, it was thought that bidders should not be coerced into bidding at all times (because losses for bidders are losses for efficiency of the auction), so waivers were introduced, five per each bidder, that allowed bidders to skip a bid. Plott reflects on how this might affect other rules:

> Can one waive and bid at the same time? What happens if you withdraw at the end of the auction: should the auction remain open so the withdrawal can be cleared? How shall a withdrawal be priced? How is eligibility of everyone influenced by withdrawals? Should it go up so anyone can buy the item released to the market? How is eligibility influenced by increments: should eligibility be lost if increments are reduced because of lack of bid? As these interactions become discovered, there is a tendency to change the policy.(1997, 629)

What exactly might Plott be talking about here? The concept of interaction has a statistical definition formalized usually by an interaction term. Suppose we are

studying the effects of two causes x and y on an effect z. Suppose also we believe that this effect is not additive, i.e. z is different from the sum x and y separately, which I shall here suppose to be αx and βy respectively . In this case it might be reasonable to add an interaction term such that $z=\alpha x+\beta y+\gamma xy$, assuming determinism. Here γ represents the coefficient of interaction which might be negative or positive. Unless it is zero the influences of x and y are not additive that is the level of x affects the difference made by y and conversely. Nancy Cartwright talks about interaction in more metaphysical terms: interaction is a case where the contribution of one capacity is changed by the presence or the level of another capacity (1989, 164-166). For example, when we say that the acid and the base interact we do not just mean that they behave differently separately than they do together, or that their relations can be represented by a statistical model such as above. Rather we have physical evidence about exactly how the two capacities interfere with each other (ibid). Is this the sort of interaction Plott is talking about?

Plott uses the term 'interaction' loosely without explaining what exactly he means. He seems to refer to the problem of adding up design features that look reasonable by themselves but once combined with the already existent features of the auction can have consequences reaching further than what anyone expected. For example, waivers are supposed to give bidders flexibility in bidding, withdrawal options are meant to relieve unexpected and undesired winnings, etc. All these factors seem to have well-defined stable effects. However, when combined together their consequences are entirely unexpected, and the effects of their introduction

spread to other design features. I think Plott has in mind something along the lines of Cartwright's definition of interaction.

I want to claim that the above reveals the fourth function of experiments – that is to show that the auction designers did not have the knowledge of many stable causal facts, or in Mill-Cartwright language, 'tendencies' or 'capacities' to work with. Of course, Plott's is not a knock down evidence to this effect. Perhaps, some models do make tendency claims but these tendencies do not have such a broad range of stability. For Cartwright the existence of interactions that neutralize capacities in some environments, is compatible with these capacities being universal outside these environments (1989,179). If so, then the fact that Plott found some facts of interactions is not necessarily evidence for the claim that economists do not know many capacities. Perhaps also our inability to make use of economic theory easily is due to the lack of knowledge about the behavior of some very concrete material facts such as the software. That is, we fail to make or to find material environments in which the tendencies make their stable contribution. These are all possibilities that undermine my claim about the forth function of experiments. Indeed I doubt it is possible to find an *empirical* example that would demonstrate that economic theory does not make capacity claims. This is why I shall discuss *philosophical* arguments against thinking that economic models make tendency claims in chapter 3. I want to treat Plott's doubts as an indication, not a fully fledged argument, in this direction.

i. Aside: discussions of aggregativity in philosophy of biology

Perhaps one way to clarify Plott's epistemic situation is through William Wimsatt's discussion of *aggregativity* (Wimsatt 1997). Wimsatt is interested in *emergent* properties and in understanding how scientists might go about explaining them. An emergent property is "a system property which is dependent upon the mode of organization of the system's parts"(1997, S373). An example of an emergent property is the superior performance and mutual regulation of two pendulum clocks connected by a beam. Starting with different periods, they in time synchronize, and each will keep time more accurately than either one on its own. This example has a clear mechanistic explanation, i.e. an explanation that tells us how different components and their interactions give rise to the emergent property, within the theory of coupled oscillators. But many emergent properties especially in special sciences still await a mechanistic treatment. Wimsatt is interested in developing heuristics that allow scientists to study such properties. He says there are too many different sorts of interactions in emergent systems to create an exhaustive list, and instead proposes to list ways in which a system can *fail* to be emergent. He calls this list "forms of aggregativity", i.e. non-emergence. The list includes four modifications following which the system is meant to stay invariant in a specified way: 1) invariance of the system property following intersubstitution or rearrangement of parts, 2) qualitative similarity following addition or substraction of parts, 3) invariance under decomposition and reaggregation of parts, and 4) no cooperative or inhibitory interaction between parts in production of the system property. Failing any of these conditions makes a system emergent. Wimsatt's example of a system that satisfies all these conditions is a system of sound amplifiers made up of four

connected amplifiers. We can rearrange the amplifiers in a different order without changing the total amplification ratio, which is the emergent property of interest (condition (1)). We can subtract one amplifier from the system without changing the qualitative result (i.e. sound is still being amplified) (condition (2)). We can take it apart and put it back together (assuming we do so in a way that respects the circuit structure) grouping amplifiers differently without affecting the result (condition (3)). Wimsatt says that the fourth condition is not strictly speaking satisfied because the amplification ratio increases geometrically rather than linearly. But we might as well treat it as being satisfied because along a certain dimension, i.e. subjective volume, the growth is linear. Wimsatt acknowledges that these conditions are only satisfied under very strict assumptions, e.g. that amplification ratios of component amplifiers is equal (for full statement of assumptions see Wimsatt 1997 S379). He also says that these are very strict conditions that most systems would fail. Any amplifier slightly more complex than Wimsatt's idealized example would count as having an emergent property. Indeed the only aggregative properties Wimsatt lists as satisfying his four conditions are mass, energy (barring relativistic effects), momentum and net charge.

This indicates that the list may not be very informative for our purposes. I explained Wimsatt's notion of emergent property in order to clarify the way Plott viewed the auction. The auction obviously fails Wimsatt's four conditions if adding a waiver option has such a cascading effect on other components of this system. But given the way in which Wimsatt sets up the conditions, most systems fail to be aggregative. And we do not learn much about the auction by observing that it, along with all but the simplest of the amplifiers, is an emergent system.

Indeed this is what leads Bechtel and Richardson (1993), also interested in the methodology of studying complex systems in biology and psychology, to distinguish between a number of different types of non-aggregative systems. One distinction is between *component* and *integrated* systems (Becthel and Richardson1993, 26), where the former is a system in which the behavior of parts is determined by their intrinsic properties and it is possible to determine properties of parts just by studying what they do without looking at the organized whole. In integrated systems, on the other hand, what constituents do depends critically on the way the system as a whole is organized and it is impossible to understand the activities of each component without studying the whole (Bechtel and Richardson 1993,18). For example, mitochondria was once an isolated independent organism and then became integrated into cell metabolism and "we cannot now understand how they function if we neglect their incorporation in, and their integration into, the complex activities of the cell"(ibid, 26). Integrated systems are, in Herbert Simon's words, "minimally decomposable", that is the function of the system cannot be subdivided into relatively independent functions of the components which add up to produce the overall function (ibid, 27). A minimally decomposable integrated system is akin to how Plott views the auction.

B. Stage 2

Plott refers to this stage as "development and implementation of auction technology"(1997, 627). By technology he does not just mean the software which implements the auction electronically. Rather he means the overall set of the rules, including bidding, activity, stopping etc., *and* the material conditions such as the

software. The experiments run at this stage enabled researchers to find one material environment, i.e. one set of rules and software, in which this technology yields the desired outcome defined in terms of speed, efficiency, income generation and other goals of the FCC. This is the fifth epistemic function of Plott's experiments.

Why was this stage significant? Because it was a way of overcoming the knowledge limitations uncovered in stage 1 (especially as indicated by function 4). If we lack knowledge of stable causal facts about the components of the auction, then we cannot design an auction by combining these components and calculating from our knowledge of tendencies and rules of their composition the result. This is a central beginning stage of the method of concretization but it is unavailable to auction designers. Instead we proceed by trial and error. Or rather by a combination of informed guesses *and* trial and error. We use our theoretical and experimental knowledge to put together one auction technology and see if that produces the right result. If it doesn't, put together a new combination and test its performance, and so on and so forth. The method is not pure guesswork precisely because we have reasons to believe some guesses to be better than others, but it is not Millian composition of causes either.

On this method we test the auction as a whole. It is not the case that all the economically or strategically important decisions were already taken and then the software developed to fit these decisions. Rather the two issues were settled in a parallel manner, because, as experiments at stage two showed, the two were not

independent of each other. For example, it is not until a software was developed and put to use that blinks of screen were discovered to carry economic effects.[23]

Plott's account, more than anyone else's, shows the wholistic nature of auction design process. Different rules do not have a stable effect across different environments, so their performance must be tested as a whole, rather than individually, and anew with almost each change in the environment. Interactive effects between rules and the software that implements them is just one layer of this wholism. Many interactions kick in even before the software implementation is considered. This is why experiments were particularly useful at this stage – an experiment is an actual auction which requires that policy concepts be translated into operational concepts, and so it allows researchers to see how these rules perform *together* (Plott1997, 628-629).

The process of trying out different "wholes" was accomplished by a three-part system of testing (ibid., 630-631). First, Caltech students with experimentally-induced preferences participated in auctions using the actual FCC software. To make sure that the success of these mock auctions was not an accident, researchers hired these 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). The process is called "debugging". This allowed the identification of problems with rules and the software by using the already experienced subject pool of Caltech undergraduates who were paid for keeping diaries about their experience of playing the auction. Finally,

[23] However, it must be noted that Plott's team did not have total freedom here. By the time of the experiments of stage 2 and 3 the broad aspects of the auction (for example that it would be open and disallow package bids) were decided.

researchers implemented "parallel checking" – 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 discovery of one material environment in which the auction rules and the technology that implemented them operated more or less as expected.

C. Stage 3

At the third stage of experimentation, researchers were required to apply their understanding of the processes at work in the experimental auctions of stage two in the field, i.e. at the actual FCC spectrum auction. Plott was a member of the "increment committee" created by the FCC whose purpose was to advise on possible interventions into the first spectrum auction held in July 1994. The FCC reserved the right to intervene by changing minimum increments, speeding up rounds, moving to a new stage of the auction because even after all the preparation uncertainty remained about whether the principles that worked in the lab would apply in a Washington DC hotel where bidders gathered to buy actual spectrum licenses.

That researchers felt confident in their ability to intervene into the actual auction to improve its running is important. It shows that they did know some causal facts stable enough such that if the cause was manipulated then the effect would be known and controllable and would not affect the whole system in an unpredictable manner. For example, Plott's team did know that the duration of the auction was

mainly the function of the number of rounds. They also knew from laboratory auctions that it can take many rounds for the equilibrium price to be reached. But the intervals between rounds did not make much of a difference for efficiency. So the increment committee was free to vary the length of the intervals (ibid., 632-633). Thus the sixth function of the experiments was to teach researchers some of the ways in which it would be safe to perturb the auction.

The final – seventh – function of experiments was to check that the actual auction runs efficiently. Recall that in laboratory auction preferences are induced, that is controlled by the experimenter. Subjects receive a piece of paper informing each of them how much they are willing to pay for a license. This means that in the end of such a mock auction the experimenters are able to check whether the final price reached is an equilibrium price, i.e. whether the winner is the bidder with the highest valuation. In real auctions this is impossible, since bidders' preferences are the bidders' own business! How then can we make sure, or at least get some indication of, whether the auction is efficient or not?

Francesco Guala explains how (Guala 2005). Prior to the real auction a number of laboratory auctions were run that approximated as much as possible the parameters of the upcoming auction. Those parameters included number of items for sale, rules, valuations (or guesses thereof), the extent of complementarities (or guesses thereof) etc. The data from these auctions, such as duration, movement of the price, existence of bubbles etc. were meticulously collected. Similar data were 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 made an argument to the effect that the real auction prices were efficient too. Guala formalizes the argument as follows:

1. If all the directly observable features of the target and the experimental system are similar in structure;
2. If all the indirectly observable features have been adequately controlled in the laboratory;
3. If there is no reason to believe that they differ in the target system;
4. And if the outcome of the two systems at work (the data) is similar;
5. Then, the experimental and target systems are likely to be structurally similar mechanisms (or data-generating processes). (Guala 2005, 180)

VIII. Theorists and Experimentalists: What Justified the Auction?

I have presented McMillan's and Plott's narratives as being different accounts of the methodology of auction design. This way of treating the two narratives is perhaps misleading, for it suggests that they are actually in conflict. But the issue is not whether theory or experiments were more important, whatever that may mean. McMillan does list specific limitations that theoretical arguments faced during the auction design and accords an explicit role to experimental work. Similarly Plott can be read as simply recounting the role of experiments without claiming that therefore they were crucial. It is natural that the two economists with different specializations, one theorist and one experimentalist, would produce accounts emphasizing elements they knew most about.

However, whether or not they are in fact in conflict is not crucial to my argument. The argument I wish to make is different: the two narratives emphasize different aspects of methodology as giving warrant to the particular auction design adopted by the FCC. McMillan's narrative is most naturally read as following 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 reasonable given the state of their knowledge. So the method of concretization, or rather some elements of it, is not the most natural representation of the method of auction design. I shall proceed premise by premise.

McMillan's narrative emphasizes the fact that the success of the FCC auctions consisted in the proper deployment of theoretical resources. Theoretical models allowed us to learn that open auctions reduce winner's curse, raise revenue, encourage flow of information, that package bids can be inefficient, that simultaneous auction favors efficient distribution, etc. We know this from "playing" with models, i.e. modifying their assumptions and observing changes in the conclusions. Sometimes theory made it clear that the same design has both positive and negative effects, as in the case of open auctions reducing the winner's curse while at the same time facilitating collusion. 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. This is particularly clear in the following assertion: "Theory says … that the government can increase its revenue by publicizing any available information that affects the licenses' assessed value" (McMillan 1994, 152). To be precise he does not think that any theoretical claim does that. Recall that while

he treats theory as being capable of demonstrating the result above he thinks it is an empirical question whether the effect of risk-averseness would outweigh the revenue-raising force of the open auction. He also thinks that sometimes theoretical results may not apply if they ignore transaction costs, or if we don't know how to ascertain unobservable factors such as the shape of distributions. Nevertheless, for McMillan some theoretical facts wear their policy implications on their sleeve.

Thus, for McMillan, The FCC auction design was successful *because* it harnessed capacities of different design features properly. This was not achieved by theoretical means, such as de-idealization or robustness analysis. This is because these procedures risk making models too complex to solve and make assumptions about facts that were difficult to ascertain. In addition to these problems, theoretical treatment of features such as multi-object auctions was not at the time available. If a theoretical treatment was available, as in the case of objects with complementary valuations, the results were often discouraging, proving that no competitive equilibrium is possible. In the absence of theoretical means, what justified the decisions, or in McMillan's language 'the judgment calls' of the auction designers? This is the question about which McMillan's narrative is silent.

Since the language of capacities fits McMillan's description of the theoretical arguments, we might surmise that the Cartwright/Nowak method of concretization developed specially to make sense of how the knowledge of capacities gets applied might be used to explain how the decisions McMillan omits to describe were made. One, perhaps the principal, stage of the method of concretization, in the case of auction design, 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 sound very fitting to McMillan's narrative. Simultaneous release of licenses for bidding (which we value for its efficiency) has a capacity to generate chaos through being too complex and to cause the auction to last too long. So we supplement the feature with various bidding and stopping rules that would prevent these effects. Similarly, the open auction reduces winner's curse by exposing information, but just in case our bidders are too risk averse to bid in a way that exposes their knowledge, we add eligibility and activity rules that push the bidders to bid and to expose their knowledge. Peter Crampton, another eminent spectrum auction specialist, seems to have the same idea in minds when he refers to these rules as "a further device for controlling the pace of the auction"(Crampton 2002, 11 manuscript).

Indeed something like this was probably the basis for construction of the experimental testbeds and the mock auctions run by Plott's team. Theoretical arguments were treated as powerful reasons in favor of including particular rules in experiments. This indeed would be the right account of the methodology of auction design if it wasn't for the complications we learn from Plott's narrative. A reason to include some rule in an experiment testing an auction design is not the same as the reason to include this rule in the actual auction. In the first case, we need a defeasible reason, while in the second we need a solid justification. Combination of stable causes is *not* the justification for Plott's team's confidence in their design.

Plott reminds us again and again that experiments demonstrated interactions between different rules that made it hard to draw any conclusions based on models

that naturally exclude features such as software details and small variations in rules. This is precisely what experiments were for – they worked to reduce 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. Given these interactions, Plott treats the material environments as *wholes* in which all elements together are responsible for the result. No single feature of these environments on its own, such as the open design, can be said to produce this result. One interpretation is that this aspect of Plott's methodology reveals skepticism about capacities.

While McMillan credits open design with the revenue raising and winner's curse reducing effect, Plott is agnostic on how winner's curse would play out when complementarities are in place. There were at the time no models that incorporated this feature and no experiments had been carried out in this environment. Plott's skepticism may be more justified. This is because although there is a theoretical result showing that an open auction allows the seller to extract more from affiliated values than a sealed-bid one, more features were added to the auction specifically in order to increase revenue by killing the winner's curse. These features are flexible minimum increments that make it more attractive to bid when the activity is low and activity rules that require bidders to submit bids on pain of losing their eligibility. This does not demonstrate that the decision-makers did not take the theoretical result seriously, but it does show that they did not trust the open design on its own to have the desired effect. There are a number of ways of explaining this decision. One of them is that researchers did not believe that the open design had a stable impact of defeating

winner's curse, i.e. that this effect, though it existed in some environments, was not a capacity.

For Plott theoretical models generate categories in terms of which to start the design process. Theory tells us that we should *care* about whether bidders can see other bids, what information they have about the good for sale, which price they end up paying, etc. However, theory does not establish facts about the stable effect of these features. That it should not tell us the effects of these features in any actual auction even if it shares the model's features is not controversial. Work on the role of models by contributors to Morgan and Morrison 1999, by Cartwright 1983 and 1989, and a number of other philosophers of science have shown us how much extra-theoretical knowledge is required to apply theory properly. Plott's challenge was partly that – to find material environments that realize theoretical possibilities, because, as he repeatedly argued, there is a long way between a policy conception of an auction and its operational realization (Plott 1997, 628-629). However, Plott's methodology may support a stronger conclusion: that there was not much tendency or capacity knowledge at his disposal and that the method of combination of causes was thus inapplicable. I say 'may support' because we do not actually need to decide on this issue for the purposes of the argument in this section. Whether or not auction designers know any tendency or capacity claims, the methods they used to confirm the right set of auction rules and software was *not* that of combination of causes.

On the further assumption that both McMillan and Plott sought to show that the decisions taken by the auction designers were to some degree justified, we may conclude that McMillan and Plott produce different reasons for why these decisions

had rationale. Perhaps, McMillan did not have the aim I impute to his account. Perhaps, he was only trying to show how theory informed auction design, rather than produce a comprehensive account of why the FCC auctions were good policy. In this case he may not endorse the method of combination of causes. Then Plott's account is all we are left with. I think this is a more unnatural reading of the records left by McMillan. However, even if it is the correct one, it is still useful to appreciate the problems of such an account were it to be put forward.

Which narrative gives the more compelling account of the methodology behind the auctions? As should be clear by now, I find Plott's narrative altogether more convincing, simply because it gives more elaborate empirical justification for thinking that the auction thus arranged will perform as expected. Plott's treatment of theoretical and experimental data narrows the gap between the initial intuitions about what auction will work properly and the final operationalization.

An extra consideration in favor of Plott's account is the historical origin of McMillan's 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 telecoms. As we have seen game theory did not give a univocal answer to the question of which auction design will maximize efficiency. For example, whether package bids should be allowed or not depends on the extent to which bidders values are complementary. Since this extent is an empirical fact outside the purview of game theory, it is hard to blame game theory for not providing a clear cut answer. Incidentally, it is through this "hole" that the

telecoms were able to present their arguments as being justified by considerations of efficiency when in fact they wanted particular designs to be adopted out of pure self-interest (for more on that see Nik-Khah 2004).

In an attempt to explain the decisions the FCC ended up taking, McMillan produced a report which 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 was not to show what exactly theory can and cannot do for auction designers, although McMillan ended up dabbling in this subject, but rather 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 be more of a "public relations" exercise, whereas there isn't similar evidence in Plott's case.

IX. Are the FCC auctions successful institutions?

We have come to the end of the story. One last element necessary to bring it to a close is whether the auction designers with their different ways of justifying decisions actually took the right decisions!

The spectrum auctions success is both controversial and not. There is much evidence that it satisfied some of the government's requirements – it attracted many bidders, distributed much spectrum, raised an inordinate amount of money ($20 billion on just the first two years) that surpassed all government and industry expectations (Crampton 1998). Even the first auctions went without a glitch and gave Reed Hundt the photo-opportunity of his lifetime when in front of TV cameras he

presented to Bill Clinton the giant check made out to "American taxpayers" (Hundt 2000). Furthermore, the FCC auctions stand out among the spectrum auctions carried out in other countries as the most predictable and reliable in attaining the goals set by the government. Moreover, economists argue that it is the features of design adopted by the FCC that is responsible for the fact that the FCC auction avoided the pitfalls of, for example, the many European auctions (Klemperer 2002b).

The auctions' efficiency is harder to judge, partly because it is hard to observe bidder's valuations and hence hard to ascertain that the licenses went to those who valued them most. However, it is possible to get some clues about it by examining the prices that the licenses were sold at. If similar licenses sold for similar prices (and approximately they did), then these prices are likely to be market prices. Another clue is that many bidders were able to purchase aggregations of licenses that are consistent with geographic synergies. Finally, there has been little resale in the 2 years following the auctions which suggests that bidders still value the licenses they purchased (Crampton 1998).

Nevertheless there is still much debate about the potential sources of inefficiency in the FCC design. The Caltech economists Mark Bykowsky, Robert Cull and John Ledyard continue to argue in favor of combinatorial auctions that McMillan and Crampton continue to find too complex to realize in practice. Perhaps more importantly there are worries that simple economic efficiency does not guarantee the highest social value. An incumbent might value a license higher than a new entrant, while the entrant might bring down the prices for the consumer (Crampton 1998). Economists advising the FCC focused mostly on simple efficiency whereas the

government authorizing the auction may have been interested social value. These are complex questions, but they are not all relevant for judging whether the original decisions described by Plott and McMillan are justified. If economic efficiency does not always equal social welfare, this is an ethical, rather than an empirical, problem with the auctions. It is an extremely important problem, but not one we expect the auction designers who work with government requirements and under political pressure to address.

What we need to close our story is just the relatively weak requirement – given the way in which the task was formulated and given what was known at the time, the auctions were a remarkable success, a good example of social science informing policy.

X. Conclusion

What should philosopher of science learn from this example? In the next chapter I envisage drawing a number of lessons. Firstly, what claims do models allow us to establish? If following Plott we do not to treat models as making tendency or capacity claims, how do we treat theoretical results? Secondly, what is the process by which these claims come to figure in our explanations? And what does this process have to do with satisfaction of assumptions or credibility, or concretization accounts of application?

Chapter 3

An account of models and their application

This is the chapter in which I pull together the ideas developed in chapters 1 and 2 to formulate and defend a positive account of how models such as the ones I have discussed here should be understood. In section I. I start off by arguing that the two such accounts currently on the market are unable to make sense of the case of auction design on the basis of game theory, or do not do so naturally. Given the prominence of this example of economic policy making, an account explaining this case is worth developing. Section II develops the new account and explores some of its philosophical and methodological consequences.

I. Why look for a new account.

In the conclusion to chapter 1 I suggested that the two main competitors impose requirements that are too high to be met by what is plausibly a typical example of microeconomic theory application. By that I mean that given the manner in which these accounts conceive of theory application, the design of the auction described in chapter 2 does not qualify as an example of one. I start with Hausman's account.

For Hausman, models do not by themselves make empirical claims, but rather supply definitions, for example "a first-price private value auction is such that rational bidders with such and such valuation distribution bid lower their true valuation".

However, once supplemented with a theoretical hypothesis (that some relevant assumptions are true of some piece of reality) we arrive at an empirical claim of the sort "in auction x bids are below true valuation of bidders, *ceteris paribus*" (or more generally, "Φ is true *ceteris paribus"* where Φ refers to a claim about what a particular economic feature does). The assumptions of the model spell out the *ceteris paribus* conditions, or the conditions under which Φ is true (though they may not be exhaustive). The claim Φ takes the form of a conditional "if A, then B", where the antecedent is often interpreted as some economic feature and the consequent as this feature's effect. Note that the CP clause includes more than just the assumptions under which I derived the result Φ in chapter 1. Mine was just one derivation. Other assumptions can be used to derive Φ. For example, we don't need to assume that there are only 2 bidders, that their valuations have a uniform distribution, etc. So the *ceteris paribus* clause refers to the complete set of assumptions which imply the result Φ.

To apply a theoretical hypothesis for explanation it is necessary to identify the set of assumptions that must be satisfied. Which assumptions are these? The first relevant assumption is described by the antecedent A of the claim Φ. In our example, it is the assumption that an auction is a first-price auction. This assumption must be satisfied by a real world auction we are hoping to explain using this model, just because if it is not satisfied, the model is not *about* the sort of auctions we are interested in explaining. But that is not enough, not all real world first-price auctions and their behaviors will be explained by this model. We need a further criterion of assumption relevance. In chapter 1 I reviewed two such criteria. The first one followed the de-idealization approach – ignore those assumptions that can be replaced

with more realistic assumptions while preserving the qualitative predictions of the model. On this approach discussed by McMullin and endorsed by Hausman (1994), to be explained by a model the system has to satisfy all the assumptions of this model that cannot be de-idealized. (The second criterion of relevance was the robustness criterion, but since robust theorems are rare in economics and none were used in auction design, I shall not discuss this criterion here.) Following the de-idealization technique, we relax those assumptions that we have a reason (from theory or from background knowledge) to take to be unrealistic of the phenomenon to be explained and importantly so. Thus, on McMullin's story making the orbit of the electron elliptical rather than circular, and assuming finite rather than infinite mass of the proton are all de-idealizations. To use an economic example, imagine a model of a common value auction which we hope to apply to explain some feature of a real world common value auction (say, an auction for drilling rights in an oil field). If we have a reason to believe that the real bidders are risk averse and that this fact is important for explanation, then we add this feature into our model.[24]

Is this account of application sufficient to make sense of how auction models entered the design of FCC auction? Before I answer this question, I shall flag an important assumption: I treat auction theory as having supplied resources that enabled auction designers to provide a causal explanation of the outcome of the actual spectrum auctions. The overall set of rules, software and the environment in which the

[24] This is not the only function the technique of de-idealization serves. It also can be used to establish that some result derived in a model is not an *artifact* of the particular features of this model but rather a consequence of a particular assumption that may represent the main cause of this result. By changing what we consider to be inessential assumptions of the model and checking the results, we may gain evidence to the effect that the result in question depends on the assumptions we think it depends on and not on the peculiarities of one particular model.

auctions were conducted explain its outcomes. More on the precise role of the model later in this chapter.

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

The important point for now is that the Hausman/McMullin account of models and their application tells a different story from what is going on in auction design. If we want to understand this piece of economic theory application we need a different account. Of course, the fact that the Hausman/McMullin account does not fit the case of auction design, does not speak against the view of models as *ceteris paribus* claims, nor does it discredit the technique of de-idealization. Whether models should be understood as Hausman proposes needs to be evaluated on different grounds. Similarly, McMullin's technique of de-idealization has a place – we do know how to

de-idealize some albeit not all assumptions. I merely want to point out that de-idealization, as discussed in chapter 1, is not sufficient to understand the role of models in auction design.

A. Do models make capacity claims?

What about the capacity account? On this view models make claims about tendencies, that is, claims of the form "A tends to do B", or if the model is explicitly causal the model asserts a capacity claim of the form "A has a capacity to cause B". On this view the model I have been using as an example should be read as claiming "first-price auction rules (in conjunction with private-values) have a capacity to cause rational bidders to bid below their true valuation". I shall concentrate on the capacity rather than tendency claims more generally since economic models most often have an explicitly causal interpretation.

Two important points. Firstly, note that not all assumptions of the model enter into the specification of the capacity claim. Unlike in the previous interpretation, only those assumptions that describe putative causes (the first-price and the private value assumption) and the putative effects (rational bids below true valuation) are present in the conclusion we draw from the model. Hence, we are not supposed to know just from the model what the conditions under which the capacity exercises its contribution are. But specifying these conditions is necessary for establishing a capacity claim. So, it follows that the model serves only to tell us what the capacity is, not to establish that there is such a capacity. Secondly, models usually make claims about capacities by using abstract descriptions which apply only if a concrete

description applies. For example, 'being a sealed bid auction' is an abstract description that applies when certain concrete descriptions apply. A concrete description of a sealed-bid auction can be very detailed: bidders receive no information about each other's bids, and they cannot infer this information from any clicks on the screen, and they cannot communicate by a secret code written into their bids, etc. This is a second reason, why the model by itself may not contain enough information to turn a capacity claim into a usable empirical claim.

The advantage of this account over the Hausman/McMullin view is that it does not require that we treat models as making *ceteris paribus* claims, where "all other things being equal" corresponds to the assumptions needed to derive a particular result. Since models' assumptions are not treated as interference clauses – i.e. clauses that specify the special conditions under which a theoretical claim holds or doesn't hold in the world – we do not need a criterion by which to differentiate the relevant from the irrelevant assumptions. Instead what we do need to apply a model is knowledge of capacities, that is knowledge of the stable causal relations, and of the material conditions under which capacities are manifested in a particular way (these may include corrections, approximations etc.). Some of this knowledge comes from theory – in particular theory gives us abstract statements of capacities and some of the conditions which are relevant to operation of these capacities. But knowledge that allows us to *engineer* systems such as lasers in which capacities are "harnessed", to use Cartwright's term, for particular purposes also includes low-level causal facts that are not part of any theory. We thus do not depend on whether or not it is possible to de-idealize the model by changing its assumptions and tracking derivations. Instead,

on the concretization view, the spectrum auction design is a matter of finding material environments that realize the abstract capacity claims. One thing that contributes to doing so is knowledge of causes and their combination. We take whatever models we have and correct them in accordance with this knowledge. This correction can proceed by ways other than construction of a great big model. Indeed at some point we should expect theoretical tools to run out.

However, the advantages of this view come at a price. The price is to demonstrate that we indeed know some genuine capacity claims. In chapter 2, I suggested that the sort of knowledge the method of concretization presupposes just wasn't there in case of auction design. I gave empirical reasons to think so, i.e. Plott's attitude to models, and also promised to discuss the theoretical ones. This is what I move to now.

One might think that the lack of knowledge of stable causal claims indicates lack of maturity of economics. Cartwright's method is the method of successful natural sciences like physics and chemistry. She claims that statistical methods of behavioral sciences such as econometrics, sociology, epidemiology etc. also assume an ontology of capacities. But it is in physics and chemistry that we have hard evidence that this method is genuinely successful (Cartwright1989, 148-158). However, philosophers of science have also found perfectly successful areas of mature natural sciences which do not use Cartwright's method. Here's Paul Humphreys commenting on the putative claim that capacities are *necessary* for application of abstract theoretical knowledge:

> While it is surely correct that in some cases we need causal knowledge to apply abstract theories, it is certainly not true that we *always* need it. Consider the difference between two commonly used methods in molecular quantum mechanics, the ad initio method and the semi-empirical method. The former begins with, say, Schroedinger's equation and attempts to construct a Hamiltonian that includes the electron interaction terms as well as the electron-nucleus interactions. This is essentially the method described [by Cartwright] …. But it is applicable only with quite small molecules. With even moderately complex molecules, many of the interaction terms must be estimated by semi-empirical methods from the data or by heuristical computational approaches. Neither of these is of course a theoretical process, but neither do they involve knowledge of capacities – in the case of the heuristic methods they are essentially trial and error procedures. It is simply a nonsequitur to infer from the progressive correction on Hamiltonians that these corrections must be causally interpreted (Humphreys 1995, 160, original emphasis)

In response Cartwright clarifies her position: the method of concretization can be used to argue in favor of capacities only where this method is actually successful (Cartwright 1995, 178). Where it isn't, which would include cases described by Humphreys and, according to me, the case of auction design, we need to articulate a different method.

So what are the precise reasons not to treat models such as those we build in economics as making tendency or capacity claims? Making a capacity claim amounts to making a claim about an empirical relation with three characteristics: *causality*, *potentiality* and *stability* (Cartwright 1998). First of all, capacities are causal, that is, to use the example from chapter 1, the connection between first-price arrangement and bids below true valuation is not purely associational. While the causality condition does tell us that a correlation between two or more factors is not spurious, it does not tell us how this relation manifests itself in situations in which confounding

factors are present. To export the knowledge of tendencies into different situations, we need more than causality. Hence the second requirement. The potentiality requirement states that tendencies tend to produce their effects even if in reality those effects are swamped by other tendencies or disturbing causes. Thus although first-price auctions do not always or almost always produce bids below true valuation, they tend to. And thirdly, the stability condition demands that the contribution of a tendency is stable across a certain set of conditions, meaning that there is a set of environments, more or less broad, in which first-price auctions do contribute to lowering bids.[25] That causal claims of this sort describe the aims of many endeavors at least in the special sciences is a widely accepted view among philosophers of science.[26] So one is not adopting a particularly controversial stand by assuming that economists should *aim* at making capacity claims.

The notion of a tendency was first proposed by John Stuart Mill in his System of Logic as a part of the "concrete deductive" method which he thought should be the method of political economy (Mill 1843). It is called concrete because, unlike in geometry where we reason from abstract axioms known a priori, in political economy we start with tendency laws discovered inductively. For example, the law that men

---

[25] In Cartwright 1989, she argues that the range of stability of a capacity is universal calling the condition "contextual unanimity" borrowing John Dupré's expression (143). However, at the same time she allows for cases of interactions in which a capacity's contribution is altered. Many commentators noted a tension: in what sense are capacities universal? (Morrison (1995), Maudlin (1993), Glennan (1997)). The advantage of contextual unanimity is that if it holds, then we are able to infer from statistical regularities to causality and back. However, Stuart Glennan suggested that we may not be able to sustain this high standard in many cases in which we think we know capacities (Glennan 1997). The range of stability, for Glennan, does not have to be universal for a piece of knowledge to qualify as a capacity. Moreover, some tendencies may have a broader range of stability than others. Stability thus should not be thought of as a 'yes or no' condition, but rather as one of degree. I shall follow Glennan's interpretation of capacities. But this is not material to the points I make here.

[26] Jim Woodward's notion of *causal generalization* – a causal relation which holds under a certain set of conditions seems to bear a lot of similarity to the notion of capacity (Woodward 2002).

prefer more wealth to less can, according to Mill, be ascertained introspectively. The method is called deductive because we do not just observe what happens in the world and generalize from that. That is futile because the causal arrangements which bring about social phenomena change all the time. Instead we use knowledge of tendencies and their rules of composition to analyze phenomena into portions caused by different tendencies. To explain we decompose already existing phenomena into relevant parts, to predict we combine tendencies believed to be at work to arrive at a claim about the likely result of their joint operation.

Contemporary economic modelers incorporate many of Mill's methodological insights. The language of tendencies is pervasive. For example, a standard response to criticism about idealization is that economic models successfully track some of the most important and general, though not all, causal trends. Idealized models are often said "capture the logic of issues", "describe the heart of the problem", "delineate the bare structure of social situations" to quote some unpublished sources. Both economists and political scientists who use economic methods assign models the ability "to capture the essence" of social phenomena (Bates et al. 1998, 12) or to "show how people behave in various circumstances" (McAfee and McMillan 1996). I interpret these claims as saying that models are capable of tracking separate causal contributions.

If this element of Mill's views is adopted, his inductivism about discovery of tendencies is not similarly embraced. Mill urged social scientists to seek ready-made tendency claims in an inductive science of *ethology* that was supposed to examine the

fundamental causes of the formation of human character.[27] Contemporary economic theory, on the other hand, derives tendencies from assumptions. In this sense this methodology pushes the inductive element further away than Mill hoped. Nowadays in theoretical microeconomics induction enters only at the level of setting the initial conditions of models and some law-like principles such as the rationality or a particular equilibrium assumptions. Some of these need to be potentially instantiated in the world. From then on, tendencies are established deductively. However, this difference is less significant than the similarities we find between Mill's project and its contemporary counterpart. Mill thought of tendencies as the fundamental laws of psychology that drive all social phenomena. By this he meant that these laws together with laws of composition and initial conditions can be used to derive predictions about macrosocial phenomena. Game theorists use assumptions, some of which resemble psychological claims and others descriptions of initial conditions, to derive tendency claims about individual or group behavior. This exercise, although not strictly Millian, is Millian in spirit.

Modelers in economics may use the Millian framework to describe their activities, but it may not be the right description of the *fruits* of their labor. Models of this sort should be evaluated on all three counts (causality, potentiality and stability) to qualify as making and justifying capacity claims. One way to do so is to take the least controversial method for establishing capacity claims and to check similarities

[27] Curiously, in discussing ethology Mill makes no mention at all of the necessity to study principles of choice or the nature of rationality. Presumably these are as fundamental for the explanation of human behavior as principles of formation of character, which is the primary purpose of ethology. Mill must have taken the nature of choice to be fairly unproblematic.

between model-building and that method. Cartwright adopts this approach in a recent paper called "Vanity of Rigour in Economics". There is a class of controlled experiments in science that aim at establishing the contribution of one factor when all other factors are absent. Cartwright calls them Galilean experiments because Galileo tried to measure the attraction of a falling body by the earth by dropping objects (allegedly?) from the Leaning Tower of Pisa while attempting to minimize the effect of all other causes such as wind, air resistance, etc. This sort of idealization – idealization which creates artificial environments with only one cause acting – can legitimize unrealistic experiments because if an experiment is unrealistic in a way that allows us to observe what a tendency does when no other cause intervenes, the results, Cartwright claims, will be widely applicable. In particular it will be applicable wherever there is no interaction that neutralizes the tendency in question. An explanatory consequence is that in interaction-free environments we can claim to have at least a partial explanation of the outcomes, i.e. an explanation of the portion of the outcome for which a tendency is responsible.

Cartwright calls our attention to one important disanalogy between Galilean idealization and modeling in economics. To qualify as supporting a tendency claim the model has to be unrealistic *in the right kind of way*. In economic models, she argues, the results are more dependent on the model than they are in, say, physics. This is because the theory that goes into construction of the model is "very meager" (1999b, 8), more so than in physics. There just aren't many uncontentious principles about the behavior of people, institutions, etc. available to modelers. On the other hand, we still want to secure deductivity because this is how statements are verified in

economic modeling. Deductivity in models is secured by importing a whole battery of assumptions that constrain the behavior of the concepts in the model. The assumptions have to serve a double duty: to be premises that provide deductivity in the model, and also to interpret the terms that later figure in the predictions of the model. For example, terms in economic models are often homonymous with terms from our everyday life: auctioneer, bidder, sealed-bid rules, reserve price, government, opposition party, threat-making, etc. This is unlike physics in which the theoretical terms are removed from our normal experience of nature. In physics, on the other hand, theories come with *bridge principles* that allow us to connect the two. For example, we know that temperature is mean kinetic energy of molecules. Since in economics for the most part we have no bridge principles, assumptions in models play two roles: giving empirical content and ensuring that the models can be solved logically. Thus in auction theory we talk about 'bidders' but these agents are very precisely defined – they do not change their mind and walk out, do not get distracted or fall asleep, do not …. This is why models in economics produce results that are *overconstrained* and this sets them apart from their original goal, that is to mimic *any*, not just *some*, ideal experiment:

> We now know what would happen – indeed, what must happen – in some very particularly constrained real experimental situation in which the features of interest really occur. But we know it for exactly the wrong reason. We know that the results obtain because we know that they follow deductively given the formal relations *of all the factors that figure in an essential way in the proof*. But the whole point about an experiment designed to establish the tendency of a factor is that the background factors should not matter at all to what happens. We are supposed to be isolating the effects of the feature or process under investigation on its own, not effects that depend in a crucial way on the background (Cartwright 1999b, 10).

The comparison of economic theory with physical theory is, I take it, meant to be illustrative. That economic theory attempts to theorize about commonsensical entities and properties, rather than build high theoretical concepts as in physics, is not necessarily a failure in itself. Rather Cartwright treats it as an indication that economic models might be overconstrained. But the main point is not the comparison with physics. Cartwright's point is that we want idealization, but only the right amount of it. Too much idealization will not yield knowledge of capacities. Instead it will yield knowledge of what a cause does under particular conditions, which will not reveal the precise contribution that this one cause makes.

But what exactly is it about economic models that makes them overconstrained? That economic principles are "meager" or that there are no or almost no bridge principles cannot serve as explanations of the exact way in which economic models are overconstrained. It is unclear how these facts, in virtue of themselves, prevent economic models from measuring the contribution of one cause of interest. To show that, we would need to demonstrate that some particular assumption acts as constraining the operation of the cause in question. To use our example again, to show that our first-price auction model is overconstrained in Cartwright's sense we would need to argue that the combination of the putative causes (first-price rule and private values) is not acting on its own, that their operation is affected by some other assumption in the model. In one sense this is easy to show: all the model's assumptions, including those of Bayesian Nash Equilibrium, are playing a role of defining the contribution of the first-price rule and the private value. But it is not clear whether these assumptions are overconstraining the operation of the putative causes.

They merely define the situation we are talking about, for example, we are talking about rational not just any bidders. Similarly, in Galileo's experiments we learn about the contribution of gravity to acceleration under the condition of being on Earth. Yet we do not think that this condition overconstrains the tendency Galileo is measuring. How do we distinguish the legitimate assumptions that define the situation from those that overconstrain? This question lacks an answer in Cartwright's schema, so we need some other grounds on which to decide whether or not models should be treated as making capacity claims.

Note that the above is an argument meant to show economic models' deficiency as instruments of learning what a capacity of a particular factor is. But we could also ask whether or not the project of modeling succeeds at establishing causality and stability of claims[28], that is to demonstrate by this method of modeling that a particular piece of causal knowledge fulfills the requirements of a capacity claim. Although, as far as I know, no claim of this precise nature has been made in the literature, some of the rhetoric of theoretical economists indicates that the assumption that theoretical modeling is able to establish capacity claims. Recall McMillan's attitude "Theory says … that the government can increase its revenue by publicizing any available information that affects the licenses' assessed value" (1994, 152). He is not saying outright that models establish capacity claims but it is the most natural interpretation of his assertion. So it is useful to evaluate this possibility.

Let us start with causality. Establishing a causal relation between two variables involves, among other things, differentiating between spurious and

[28] Presumably once we establish stability of a cause, its potentiality will follow.

genuinely causal correlations between these two variables. One way to ensure that a spurious correlation between A and B is not mistaken for a causal one is to vary the level of the putative cause A in such a way that a subsequent change in the level of the putative effect B can only be thought to have come about as a result of A causing B. The precise rules for moving from facts of correlation to facts of causation are complex and subject to debate. For our purposes, it is not necessary to get into details. It is enough to note that the sort of manipulation that allows us to gain evidence of causation is analogous to the way in which we manipulate models. Models, like the experiments, are thought to provide answers to "what happens if…" questions. Here's a much quoted claim by eminent economist Robert Lucas expressing this view:

One of the functions of theoretical economics is to provide fully articulated, artificial economic systems that can serve as laboratories in which policies that would be prohibitively expensive to experiment with in actual economies can be tested out at much lower cost. (Lucas 1982, 271)

Lucas is treating models as, or similar to, laboratory experiments. (Indeed it is Lucas's view that Cartwright is criticizing in "The Vanity of Rigour") Just as we may test causal claims in a laboratory experiment, we could, according to Lucas, do that in a model. Indeed intuitively there is an epistemic similarity between the following two situations:

a) an economic experiment in which subjects (these could be undergraduate students, as is often the case in contemporary experimental economics) are given monetary incentives to compete in a first-price sealed-bid auction for a certain good. Their valuations are given to them by the experimenters. As the bidding proceeds their bids

are recorded. Then the experimenters change the rules so that the second highest bid is the price the winner pays. Otherwise the second auction is identical, as far as the experimenters can make sure, to the first one. When bids in two auctions are compared, it is found that in the second-price auction the bids equal valuations whereas in the first-price one they are below it.

b) a paper and pencil model of the first-price private-value auction presented in chapter one. Switching the first-price for the second-price assumption while holding all other assumptions fixed results, by deduction, in "truthful" bids. It is no longer rational to bid below true valuation.

Should we interpret both of these situations as providing evidence for the hypothesis that the first-price as opposed to the second-price rule causes bids below true valuation as opposed to "truthful" bids under conditions of private values? If we treat situation a) as providing such evidence why not treat situation b) as doing so? Indeed in the second case, we may think we are better off, because we know exactly the assumptions that are responsible for the result, whereas in the first case we may have failed to control for confounding factors.

One might think that in fact the two situations afford different epistemic opportunities with respect to the question of causality: in situation (a) we know a relationship of logical entailment that holds between sentences, while in situation (b) we know, albeit imperfectly, a causal relation that holds between concrete events. However, the fact that the conclusions of a model follow logically should not be thought as a reason to either deny or grant models' derivations the status of causal claims. The relation of logical entailment holds between sentences, while causal

relations hold, on most theories of causality, between events. So whether sentences that describe these events stand in a relation of logical entailment is entirely orthogonal to whether or not the relation between these events is causal. To establish that a relation is causal we need to vary the putative cause and see if the putative effect has changed, and this is precisely what we do in the model. Logical entailment here serves as an analogue of the role that the background laws play in ensuring that the cause does produce its effect. I thus see no reason to think that models cannot teach us that some feature is the cause of another in the particular situation defined by assumptions. The situation gets more complex when we attempt to evaluate models' ability to establish stability of these causal relations.

The requirement of stability has to do with the license to extrapolate from the model to the world. This is where there are reasons to believe that models perform worse than material experiments. We have already touched on Morgan's ideas on analogies and disanalogies between models and experiments in chapter 1. Here I am taking up these matters again. Morgan as well as Guala locate the difference between models and experiments at the level of *materiality* (Morgan 2002, 2003, 2005 Guala 2002). Both models and experiments may be able to answer "what happens if…" questions, but this does not mean that it is the same question in each case. What happens if we introduce a new assumption into a model, on the one hand, and what happens if we introduce some change into people's decision making environment, on the other, are two different kinds of interventions when it comes to extrapolation. How exactly are they different? Consider an inference from either a model or an experiment such as those described in situations (a) and (b) to the world, for example

some actual art auction. We may be interested, for instance, in finding out whether a relation we discovered in the models or the experiments explains the outcomes of this art auction. Morgan and Guala claim that we are in a better position to make such an inference from an experiment than from a model because there is an ontological difference between entities in models and entities in experiments. Entities in the experiments are made of "the same stuff" as entities in the target system (Guala 2005 214-215, Morgan 2005 323). From the situation (a) we are able to infer that a causal relation between auction rules and auction outcomes holds with real people of flesh and blood rather than only with the precisely defined agents of situation (b). To quote Morgan again: "the construction of an experiment recreates a part of the real world inside the artificial environment of the laboratory. In contrast, the construction of a model creates an 'artificial' world"(2005, 320). We still need to address the problem of external validity, that is to show that the behavior of the "stuff" in the experiment generalizes to the behavior of the same "stuff" in the wild. But this inference is easier to justify than the inference from the "stuff" of models, which is assumptions and mathematical relations, to the "stuff" in the world. In case of the 'experiment to world' inference we need only solve the problem of external validity, whereas in case of 'model to world' inference we need to solve the problem of realism of assumptions and the problem of external validity. So, it is claimed, we have better reasons to think that the same causes that work in the experiment also work in the target system.

Why shouldn't we just deny that there is an ontological difference between the entities in models and those of material experiments? Why not claim, against Guala and Morgan, that although entities in models are sentences and logical relations

between them, they *refer* to material entities in the world? If so, then it is no more and no less difficult to demonstrate a capacity's stability from the model to the world and its stability from the experiment to the world. I think the reason to be careful here is that, to use Guala's words, "although economic models are often made of "firms", "consumers", "markets", and so on, these are not the firms, consumers and markets of our everyday world"(Guala 2005, 220).[29] The objects in models are different in the sense that their behavior is constrained a lot more than the behavior of the homonymous objects of the everyday life. Having run a material experiment that realizes a model's prediction, we get evidence that these constraints do not matter. But in the absence of such evidence, this difference between experiments and models stands.

Recall what in chapter 2 I referred to as the first function of experiments – that was to establish that possibilities proved in models were also material possibilities. For example, there is theoretical evidence that the option to bid on packages creates an advantage for bidders with complementary preferences. That was the point of Peter Crampton's model about bids on parking lots. What was the point of running an experiment, as Plott's team did, demonstrating this fact? On my view the point was to make sure that there are environments that are made from 'the same stuff' as the environments of the spectrum auction in which package bids have such and such effect. Why is this necessary if we also have a model whose assumptions appear to be satisfied in the situation in question? Because, we cannot be sure that all the assumptions of the model are really satisfied, or satisfied enough, by the target

[29] A parallel point is made by Cartwright 1999b and economist Robin Cubitt (2005).

situation, until we have run an experiment based on this model and made sure that the model's prediction is realized in this experiment.[30] They may sound as if they are satisfied, because objects in models sound similar to the everyday objects we know. But, as suggested above, we are not licensed just on this ground to conclude that assumptions are indeed satisfied. That auction designers insist on demonstrating material possibilities, rather than just building models that seem to satisfy the conditions of the target situation, indicates that they too take the difference raised by Morgan and Guala seriously.

Of course, there is a lot more to the stability of a capacity than just making sure the capacity refers to 'the same stuff' as the stuff in the world. We also need to show that this capacity exercises itself even in the face of other capacities and interferences. How exactly and whether this can be achieved by pure modeling is an important question. Indeed the best evidence that applied economists do not treat theoretical results to have the sort of stability that capacities are supposed to have, is that faced with an empirical challenge such as the FCC spectrum auction design they treated theoretical results cautiously, ascertaining their stability by experiments one step at a time, one environment after another. But the materiality challenge described above is one of the first steps, and hence very basic. Of course, just because experiments may be better than models at meeting this challenge, does not mean that we are unable to establish stability of capacity claims made in models. Still it is clear that this is a challenge for models qua models. Furthermore, auction designers' insistence on establishing material possibilities via experiments indicates that this is

[30] I thank Vince Crawford for this point.

not a challenge that can be met by modeling alone. This then is one piece of evidence that modeling is not enough to establish capacity claims. However, it is not crucial for my argument to accept it. Even if there is no important difference between models and experiments, even if materiality is of no consequence whatsoever, there are still reasons to look for an account that does not make such heavyweight assumptions as does the capacity account. It is clear that in many contexts of extrapolation in special sciences stability of causal relations is precisely what is in question and cannot simply be assumed. If so then we can still apply theoretical knowledge without knowledge of capacities. How then do we do it?

Although there is much more to be said about the relation between modeling and capacity claims, I shall leave this topic. I take to have given reasons to look for an interpretation of models such as those I've been discussing and their application that is different from either the Hausman/McMullin or the Mill/Cartwright views. The main weakness of the first account from the point of view of explaining auction design is that in its weaker version – application is satisfaction of relevant assumptions – the view is not very informative. If we try to strengthen it by providing a criterion for relevance of assumptions – the widely endorsed technique of de-idealization – then it becomes too narrow to be an adequate description of how theoretical knowledge gets applied in practice. Sometimes we may not be able to carry it out and yet can use economic theory nevertheless. Moreover, this account does not have the resources to explain the methodology of combining insights from different models as was done in our case.

The Mill/Cartwright account has great tools for explaining how we move from theoretical claims to concrete environments, and how models can be combined, but it also assumes knowledge of capacities, which in many cases is not justified. There may be ways of dealing with these objections, but my strategy is not to discredit these accounts. Rather it is to suggest a new one and explain its advantages. The new package as a whole, as we shall see, has advantages that speak in its favor. What might a new account look like?

II. An Account of Deductive Causal Models

A. What do models' assumptions tell us?

In chapter 1, I brought up the possibility that we may not know what exactly the model claims about the world just by examining the model. But I did not explain this claim properly. Now is the time to be clearer. I shall discuss a number of ways to interpret this possibility, but before that I need to make two observations vital to the positive account I defend.

The first observation is that in the sort of models I have been discussing we often treat some of the assumptions are more *salient* than others. When we wish to draw a causal hypothesis out of a model, only some of the model's assumptions become part of this hypothesis. Let us use again our auction model as an example. What is the content of this model? Strictly speaking, the content consists only of the model's assumptions and its deductive consequences, which are listed in section I.A of chapter 1. When we wish to use this model for explanation or for designing an institution, we do not just list all its assumptions. Rather we draw a causal hypothesis.

One natural causal hypothesis in the auction model in question is as follows: when values are private and some other conditions obtain, first-price auction rules cause bids below true valuation. This hypothesis has a form of "feature(s) F(s) cause behavior(s) B(s) under condition(s) that include C(s)". Note that not all the model's assumptions go into the causal hypothesis. They could, but they often don't. We fail to list certain assumptions *not* just out of economy of time or space or mental effort. Rather we *mean* to omit them, because we make a salience judgment about which assumptions are more interesting or relevant than others for our purposes. I may be interested in the effect of auction rules on the bids, so with this causal prejudice in mind I formulate such a hypothesis on the basis of the model. I may, on the other hand, be interested in the effect of valuation distribution on the bids, in which case the same model is read as suggesting a different causal hypothesis (say, when there are two bidders, continuity of valuations' distribution causes bids to be discounted by half). There is a distinction between such a causal hypothesis and the model proper. When I say that a model may *suggest* one or other causal hypothesis, this is what I mean.

The second observation concerns general features of the sort of scientific models I am talking about. Up to now I have been talking about 'models that are typically used in microeconomics', but that, of course, is not a satisfactory description. The kind of models I aim my account to cover are used across sciences. They are those models in which some of the assumptions are read as putative causes and their deductive consequences as putative effects. In some areas of theoretical biology there are deductive models which are naturally interpreted as tracing the

effect of a new predator, or a new pesticide or some other ecological property on some feature(s) of populations of organisms. The account I am developing would fit these models naturally. The natural contrast class are phenomenological models which merely attempt to trace certain occurrent properties. An example from neuroscience and electrophysiology are Hodgkin and Huxley equations of action potential (which is the rapid rise and fall in the electrical potential difference across a neuron's membrane). Carl Craver argues that they are best understood as a phenomenal description of the dynamics of the total current that crosses the membrane (Craver, forthcoming), and indeed this is how the equations were treated by its authors. These equations serve many functions, including prediction, but they are distinguished sharply from the mechanism that is thought to produce the action potential. This mechanism involving ionic channels was not discovered till several decades later. My account does not have anything to say on how phenomenological models get employed for prediction, nor how we should read their claims. I return now to the question of what assumptions in models that have causal ambitions can tell us or fail to tell us.

There are a number of ways in which we can treat the assumptions of the model vis-à-vis the causal hypothesis it suggests. Assumptions may fail to describe the conditions under which the causal relationship suggested by the model holds in ways listed below:

1. Assumptions may not give us the full detailed description of all the conditions that need to be in place to realize the causal relation of the hypothesis. For example, no game theoretical model tells us the properties of the software that

need to be in place for bidding to proceed in the way that game theory predicts. This insight is part of Cartwright's account via the distinction between the abstract (often theoretical) and concrete (that is implementation specific) descriptions. This is the least controversial option. No new account of models is necessary to make sense of this possibility.

2. Assumptions may also fail to give an exhaustive list of conditions (even in abstract vocabulary) of all the conditions which realize a causal hypothesis. This is the idea that the causal claim may hold in conditions wider than those described in the assumptions of the original model (or in different assumptions that preserve deduction of the result in question). We may need other theoretical tools to describe the conditions outside the model. For example, we may find that first price rules still lower bids even under the condition of, say, winners having to pay royalties on their winnings. This condition is neither excluded nor taken into account by the model I have been using as an example.
3. Alternatively, assumptions may fail to tell us the conditions under which the causal hypothesis is true when the assumptions are violated. For example, first price rule may still cause bidding below true valuation even when the bidders in question are not perfectly rational or do not play the Bayesian Nash Equilibrium. Unlike in the previous case, such circumstances are explicitly contradicted by the model, but the causal relation (one certainly hopes) may still hold.

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

Which of these possibilities are correct? Certainly the first three. The fourth is controversial. Nevertheless, I suggest that an account of models and their application should allow for all four possibilities. This way we remain on the safe side and neutral on the controversial issue of whether or not all assumptions of models are realizable.

This strategy invites us to challenge the following presupposition common to both Hausman/McMullin and Mill/Cartwright account: whatever else models may tell us they describe at least one situation in which a causal hypothesis the model suggests is true. That is deductive models always provide us with the following claim:

(1) There is a situation S characterized by $\{S_1 \ldots S_n\}$, such that in this situation a certain feature F causes a certain behavior B.

The Fs, Bs and Ss are just the resources we use to construct different causal hypotheses on the basis of the model. The Ss are the assumptions under which the

conditional "if Fs then Bs" can be derived. The task of those who wish to apply the model to some new situation S′ is to ascertain whether in S′ a certain feature F′ also causes a certain behavior B′, where F′ and B′ are real world features and behaviors which bear some resemblance to respectively Fs and Bs in the model. The two accounts I have been examining give us different recipes about how this is ascertained. But they each give such a recipe, either in terms of de-idealization or in terms of a capacity claim. Applying this recipe is sufficient for applying the model to the empirical phenomenon.

B. Models as open formulae

I propose to replace (1) with the following idea. Rather than making a claim about what some factors do in some situations, we may treat models as specifying a *recipe* or a *template* or an *open formula* for such a claim, but not the claim itself. An open formulae takes form of:

(2) In a situation x with some characteristics that may or may not include $\{S_1 \ldots S_n\}$, a certain feature F causes a certain behaviors B, where x is a variable and F and B are property names.

In this open formula there is no commitment to the existence of x and no claim about any empirical phenomenon since the features of x are not specified. X is a free variable, which needs to *be filled in* in order for the open formula to make such a claim. Once x is specified, we get a causal hypothesis of the form "an F causes a B in

a situation S". Without closing the open formula by specifying x, (2) only gives us a template or a schema for a causal claim, rather than a fully fledged causal claim we see in (1). Note that (2) is more general than (1) in the sense that it does not exclude the possibility expressed in (1) that models can come with a specification of the situation in which the causal hypothesis is true. So (2) is more liberal in that it does not commit on this question one way or the other.

Furthermore, (2) is more liberal than (1) in that it is explicit in not restricting the causal claims to be made on the basis of a model to those claims defined by the model's assumptions and in not privileging any assumptions over other conditions under which these relations can hold. Models that I am discussing are often evaluated on their tractability, elegance and, of course, deductive closure. These are all reasons, over and above statement of the conditions under which the result holds, to include certain assumptions in models. So while all assumptions *may* be treated as also defining the situations in which the causal relations the model suggests holds, they do not have to. We thus have the license to go ahead and build many different causal claims on the basis of one model.

There is a cost to this flexibility. Note that on the two existing accounts if the model establishes a causal fact in a particular situation, then if the rules these accounts set up for moving from the model to some real world situation can be used and if we follow them correctly, then we can be sure that this causal fact also holds in this situation. On the Hausman/McMullin account successful de-idealization of a model with respect to some real world situation gives us warrant to assert that if the original model established a causal claim, then the de-idealized one does too. Similarly, on the

Mill/Cartwright account if the model makes a justified capacity claim, then this capacity claim would still be true outside the model provided that we correctly identified the situation in which this capacity is exercised and no factor neutralizes this capacity. In each case, if the original model tells us that an F causes a B in a situation S, then if the model applies to a situation S′ according to the rules of the account, then the warrant to claim that an F′ causes a B′ travels from the model to S′.

On the view proposed here, such preservation of warrant cannot be sustained. Once we treat models as open formulae from which we can build causal hypotheses about the world, we must keep in mind that, unless we can use de-idealization or have evidence for a capacity claim, the causal hypothesis we have constructed has to be confirmed in some other way. There is thus a trade-off – on the open formulae view we get a free reign to pick and choose what assumptions of the model will figure in the causal hypotheses we build on the basis of the model. But by doing so we commit to finding a confirmation of this causal hypothesis. This cost has to be reflected in the account of how open formulae come to form explanations. However, I think this cost is not an objection against the open formulae view of models. This is because the ways in which causal truths are preserved when we move from the model to the world on the two existing accounts (that is, de-idealization or a capacity justification) are unlikely to be available in many cases of theory application in special sciences.

My objections to the two competitors have not been ‘in principle’ conceptual objections. Rather they are more akin to applicability objections. So I do not seek an account that would supersede de-idealization and capacities. Instead I seek a more

general account that would incorporate these justifications, where they are available, but also have the flexibility to deal with situations in which they are not.

C. How open formulae get applied

The view of models I propose is just that – models should be read as partially filled-in open formulae. The next question is how these formulae come to be applied for various tasks. There are many tasks for models in science, and here I concentrate on one – explanation. Sometimes being in a possession of an explanation also provides us with sufficient tools for prediction and intervention (this was the case in the FCC auction example), so my discussion here will apply to those cases. There are also cases in which we are in a position to make predictions and to intervene but not to explain, but I do not intend to cover these. Furthermore, I restrict myself to causal explanations which are by far the most common in applied science. So how do we use partially filled-in open formulae to give causal explanations?

A causal explanation of a phenomenon requires that we make a justified claim about a causal relationship that obtains in this phenomenon. On the two accounts I have discussed this is done by assessing the relationship between the causal relations in the model and those in the phenomenon in question. On the account I am proposing the model is not assumed to supply us with any ready made claim about any causal relationship. So we proceed differently, in three steps:

1. Identify an open formula on the basis of a model by picking from the model's premises and conclusions the Fs and Bs of interest. These

presumably will correspond to the putative causes and the putative effects in the empirical phenomenon in question.[31]

2. Fill in x so as to arrive at a closed formula or a causal hypothesis of the from "Fs cause Bs under conditions S" where Fs and Bs match some aspects of the target situation. For example, if we wish to explain some art auction (these can often be private value) which uses the first-price rule and generates bids lower than expected, we may use the model I explained in chapter 1 to make a hypothesis "the first-price rule causes bids below true valuations under such and such conditions S" where S may or may not be described by the original model's assumptions.
3. Confirm the causal hypothesis. We do that by finding a *material realization* of the causal relation in the hypothesis.[32] A material realization is a material environment such that if it obtains, then an F causes a B. That such a material realization exists tell us that the model which inspired the causal hypothesis in question applies. That we know a material realization of a causal hypothesis tells us this hypothesis is true of whatever situation the material realizer describes. Recall that unless our causal hypothesis is a true capacity claim or we can use de-idealization in order to pick out the assumptions which really matter for the truth of this hypothesis, we need to confirm the causal relation between the F and the B by some other

[31] What if the Bs in the world are not matched by the Bs in the model? In this case, I would argue that we do not have a theory-based explanation of the phenomenon, though we may have some other explanation, which is outside the scope of this dissertation.
[32] The realization has to be material in the sense that the causal relation in question has to be one that actually happens in the world. This does not mean that it may not be represented by a computer simulation or an equation.

> methods. We do that by whatever methods we normally use to infer causal relations. Of these there are a number: randomized controlled trial, natural experiment, Mill's methods and variations on them, mark methods, etc. But I am bracketing the question about how we do that.

So it is by filling in model's open formulae, i.e. by specifying their material realization, that we apply models for explanation. How are material realizations specified? The features of this environment in which the causal hypothesis is true are what allows us to fill-in the causal hypothesis, that is to specify the situation S under which an F causes a B. If in the art auction in question we find that the first-price rule indeed causes bids below true valuation, then we examine this particular material situation to specify S. Part of this specification may match the assumptions of the original model. For example, we may find that our art auction is indeed private rather than common value or approximately so. We may also find that other assumptions of the model are not satisfied (say, the bidders are not perfectly rational), in which case some condition other than rationality as assumed by the model figures in our specification of S.

Obviously, we don't just blindly list every single detail about the material situation in which the causal relation we are interested operates. We would not, for example, typically list the eye color of bidders in the art auction in question as part of S. Not just because explanations are sometimes evaluated on their conciseness. But rather because we seek to include in our specification of S only the factors that are relevant, i.e. the factors the presence of which makes Fs cause Bs in this particular situation. In the FCC auction design case, many more factors turned out to be relevant

than was thought at first. So the safe option is to be very specific in one's characterization of S. This knowledge comes from many sources, theoretical, experimental, observational, etc. Just what must the situation S be like for Fs to cause Bs is typically subject to much debate. Whatever story philosophers of science and methodologists tell about causal inference in various scientific settings will presumably help to clarify the methodology of filling in model's open formulae. But it is not part of an account I seek to defend.

The use of models for explanation presented above is supposed to fit both singular explanation and explanation of recurring phenomena. When the model is used to explain a singular phenomenon, we seek to formulate one material environment which realizes the causal hypothesis the model suggest. Sometimes, however, our aims are bigger. We want to understand a kind of phenomenon, not just its one instantiation. This is when we seek knowledge of material realizers that are *robust* to certain changes. We want to know a set of material environments such that if they are in place then the causal relation we are interested in holds. I call this *empirical robustness* – a property of a causal relation defined by the amount of variation of the situation S, i.e. the material realizer of the causal relation between F(s) and B(s). In case of our private value sealed-bid model and the causal hypothesis "first-price rule causes lower bids under some conditions", empirical robustness of the causal relation between first-price rule and lower bids would be given by a range of real-world material factors such that if they hold, the first-price auction rule causes bids below their valuation.

This sense of robustness is different from the sort of *derivational* robustness investigated by the technique of de-idealization as discussed in Chapter 1. There we study the set of assumptions under which a given prediction can be derived by the rules of deduction licensed by our method. Presumably some of the assumptions required to derive the prediction of lower bids will correspond to the material factors needed to be in place for the bidders to bid as predicted. Thus the assumption of *sealed*-bid auction uncontroversially maps on a possible material factor – an institution in which bidders do not see each other bid, but rather submit their bids in a sealed envelope independently. Similarly, for other assumptions such as that the number of bidders is two and that they know how much they value the good for sale. However, other assumptions, perhaps those describing properties of distributions of valuations and Nash Bayesian decision rules, will not do so. These assumptions may not be easily relaxable, so the derivational robustness is of no help. It may be one thing to show that many assumptions will lead to the same derivation and quite another to prove that simplifications do not matter for empirical results. The sense of empirical robustness I would like to incorporate into my account is more general: sometimes we may not be able to de-idealize away some features of the model by deduction. When we don't, we may still be able to make judgments about robustness of the causal relation in question by other methods.

D. Auction design in the new light

How does this view make sense of the case of auction design I have studied? When we use models for explanation or prediction of the behavior of already existing

systems, application is a matter of *finding* the material realizations of the causal claims suggested by the models (or specifying the already existing ones). When, on the other hand, models get employed for policy design as in the case of the FCC spectrum auctions, application requires *construction* of material environments in which the desirable causal claims suggested by the models are true. Since I made an assumption that in the auction design case the make up of the auction also explains its outcome, I shall describe the use of models in terms of explanation.

Models of auction theory supplied researchers with a number of partially filled-in open formulae: "First-price auction causes bids below true valuations under conditions …", "Open auction defeats winner's curse under conditions …", "Individual rather than package bids do not hinder efficient distribution under conditions …" etc. So, at the beginning, there was a wide range of "F_i cause B_j in x" claims. What researchers ended up with was an explanation of the actual auction in the form:

"F_1, F_2, F_3, … F_n cause B_1, B_2, B_3, …B_n under conditions S_1, S_2, S_3, … S_n"

where Fs stand for features of rules, Bs for aspects of the auction's outcomes (i.e. its revenue generation, license distribution, speed, etc.) and Ss for the material conditions such as the software that auction designers could control. The Bs were partially specified by the government requirements. The trick was to find the combination of Fs and the Ss that would bring about the Bs. Some of the Fs, Bs and Ss figured in the models of auction theory, others came from different sources of knowledge.

In chapter 2 I argued that one function of experiments (the fifth, to be precise) was to find one experimental auction environment, i.e. one combination of rules and software, such that if this environment were instantiated the auction would run as required. This is the method which provides the main inspiration for my account of model application as material realization. Some features of this environment such as the open rules, public information, bids on individual licenses, figured in models of auction theory. For example, there was theoretical evidence that open rules increase revenue when values are affiliated, or that open rules reduce winner's curse.

However, since the models from which these results follow were not treated as making claims about stable causal contributions of these features, it was the aim of the experiments to find such environment in which these features altogether brought about competitive bidding, speed, revenue generation etc. It was not known what the effects of these features are except in this one environment plus whatever environments preserve the outcomes (for example, an environment with longer or shorter intervals between stages of bidding – recall that function six of experiments was to gauge the design's robustness, i.e. to show which perturbations of the auction are safe).

The experiments testing a number of different combinations of rules and software resulted in a very specific piece of knowledge: an open auction for goods such as spectrum licenses leads to efficient, speedy and revenue-generating distributions provided a particular list of conditions holds. These conditions include the detailed rules the FCC developed, the right software, the right legal and economic

environment, and that certain interventions were possible. The conditions are intricate and specific.

When it comes to generalizing this particular design beyond the 1994 auction, researchers are extremely careful (Klemperer 2002b). Nevertheless, auction design is widely regarded as one of the most successful areas of application of economic theory and the FCC auction is its article of pride. The proposed account makes sense of this methodology. Auction theory taught experimenters what features might figure in causal relations characterizing different kinds of auctions. But experiments were needed to, among other things, find one minimally robust material instantiation of these relations.

On this view there is little sense in which auction theory models were *confirmed* or *borne out* by the spectrum auction. Models supplied reasons but not justification to adopt a particular rule. But these reasons should not be viewed as just wild guesses. Instead they were very powerful reasons. Models of auction theory can tell us that given some feature F a behavior B (which we might find desirable in an auction) *must* follow as a matter of logical necessity in some situation S which is described by the model's assumptions. (Remember that this option is not excluded by my formulation of models as open formulae.) That there are such situations gives us reason to believe that we may be able to create situations S′ which have some resemblance to S in which the causal relation between the F and the B could hold. This is a good reason in favor, albeit not a full scientific justification of, the decision to include certain theoretically inspired Fs and Ss into the auction design. However,

the justification of this design was clinched by the experiments that tested the combination of Fs and Ss as a whole.

The challenge comes when we know that there is a limit to how much we can control S. For example, 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 auction models assumed. There were reasons to think that most of these bidders would not be completely unpredictable, since many of them hired their own private game theorists to advise them on bidding strategy. However, allowances had to be made for the less well equipped bidders, for lack of experience, for the fact that the auction is complex and that conformity with predictions of models that assume rational choice is not likely given that the auction is a one shot affair, rather than a repeated game with practiced players. So a variety of other rules were added to push the bidders to behave as the models predict.

A further challenge was lack of knowledge about S. Models make assumptions about the shape of distributions of valuations of bidders, but auction designers had no way to gauge facts about these distributions, since potential bidders would not volunteer such information. These two challenges indicate that we often cannot rely on models' assumptions to tell us what features S should have. If so, then it is necessary to find other knowable material conditions that researchers could control and that could realize models' hypotheses. In these cases theory still plays an important role in the interventions, predictions and explanations we devise. But it does so in ways that are different from what the existing accounts of theory

application tell us. This is why it is hard to make sense of how theory participated in auction design without the flexibility which my account incorporates.

E. Comparing the new account with the existing ones

Which elements of the two existing accounts are preserved and which are rejected by the account I propose?

My account inherits from Hausman the idea that models do not by themselves make any empirical claims. For Hausman it is a theoretical hypothesis that turns a model claim into an empirical claim. For me, it is a formulation of a closed formula. However, there is no place in my view for satisfaction of assumptions. Empirical situations may or may not satisfy some model's assumptions, but if this model applies it does so not in virtue of satisfaction of assumptions. Rather it does so it virtue of the fact that the causal hypothesis it allows us to generate is true of this situation.

What about the Mill/Cartwright account? Unlike Hausman's this account makes causal notions central to the project of modeling and application of its results. The account defended here follows this line: models often are used to make causal hypotheses and their application is a matter of ascertaining causal facts. In the case I examined in chapter 2 this was particularly clear. Note in particular, that on the proposed account, a material realization of a model claim is not just an environment in which features Fs are associated with behaviors Bs. Rather a model claim is realized when an F causes a B in this environment. Nor does my account commit to reading claims in open formulae as claims of association between *occurrent* properties Fs and Bs. They can also be read as claims about *contributions* (a feature F

makes a B-like contribution in a given situation). For example, an open auction does not have to actually cause defeat of the winner's curse in a given material environment; it may only make a contribution in that direction. We then have to find a material realization of this contribution claim.

Where my account departs most strongly from the Mill/Cartwright view is in the interpretation of the stability of model claims. When we move from the 'capacity' to the 'open formula' view of model claims we surrender the idea that causal claims need to have a broad range of stability to be useful for explanation, intervention, etc. Knowledge of capacities need not be assumed for theoretical economics to be usable. When we let go of stability of causes we also must let go of the Millian method of decomposition and composition of causes. Following Humphreys' proposal I am simply elaborating a method that scientists use when this more ambitious schema is unavailable. Perhaps the ontology of tendencies or capacities is still the correct ontology of the economic world. But when contemporary economics aims at empirical success it does not have to assume this kind of knowledge. My contribution is to show how justifiable application of theory is still possible despite the fact that this framework is inapplicable.

F. Consequences for experimental economics

In Methodology of Experimental Economics (2005) Guala makes a number of observations about the relation between economic models and economic experiments which fit neatly with the account proposed here. What exactly is under test when we

find or construct empirical environments on the basis of models? How is economic theory tested?

There is an assumption often unarticulated that an experimental setup testing a theory or a model must as far as possible approximate this model's assumptions. Otherwise we would not be testing this model. According to Guala, this is the wrong attitude (2005, 217-222). Experimental situations are in many ways independent from theoretical models and to be good tests of theory, experiments need not match models perfectly. This is what Guala means when he says that experiments mediate between models and real world situations in the wild (Guala 2005, 209-213). I shall explain using his own example of experiments inspired by the Prisoner's Dilemma (PD) game.

The model is often thought to explain or fail to explain a number of familiar phenomena in economics and beyond. What exactly does that this mean? From a one-shot PD model we know that given a certain payoff and choice structure there will not be, as a matter of logic, any cooperative behavior on behalf of the rational players. We may try to build an experiment that instantiates exactly all the assumptions about payoffs, impossibility of communication, utility maximization, etc. Quite apart from the fact that this is practically impossible, we would not be learning anything new from such an experiment. After all we already know what must happen if the assumptions are exactly replicated! What is it, then, that we learn from experiments that give people monetary payoffs identical to payoffs of the PD game? According to Guala, we learn whether or not the particular application of the PD game is successful. My reading of this claim is that such an experiment is an opportunity to

test a causal hypothesis about a concrete situation constructed on the basis of the PD model. The model only partially specifies this hypothesis, it is thus natural that the experiment should not replicate exactly the model's assumptions.

In what sense do the model's assumptions fail to match the conditions of the experiment? Consider an experiment in which real people are invited to play the one-shot PD game. The causal hypothesis under test is that a certain structure of payoffs causes subjects to defect. Note that in giving experimental subjects monetary payoffs identical to those of the PD game (for example, (10,10) if both cooperate, (5,5) if both defect, etc.) we are assuming that this is all that utility amounts to in this situation, i.e. that these subjects in this situation are motivated only by money. If this is not the case, then the subjects are playing some game other than the PD. Importantly, that subjects are only motivated by money is not an assumption of the original model. Rather it is an extra-theoretical judgment made specifically to define the conditions of this experiment. And yet it is a vital specification of the conditions under which the causal relation between payoff structure and behavior holds (if indeed it does). Furthermore, the PD model assumes subjects to be faultless utility maximizers. This assumption, may, of course, fail to hold in an experimental set up with real people. But this is not taken as a reason to reject the hypothesis. If there are reasons to think subjects are not faultless utility maximizers and yet we see defection as the model predicts, it may be the case that the structure of incentives similar to that of the PD game causes defection under some conditions other than faultless utility maximization. We then study this experimental set up to figure out the exact factors that are responsible for this relation holding. These could include anonymity, the level

of payoff, assignment of social roles etc. For each application of the PD game we need to fill in the causal hypothesis with a variety of assumptions that may or may not match the assumptions of the model. This is the hypothesis that we've been testing all along, not the model itself. For models more complex than the PD game, there shall be further assumptions that fail to be satisfied in the experimental set up devised on the basis of these models. Assumptions such as continuity of the utility function, perfect foresight, completeness of preferences, infinity of traders, absence of transaction costs, etc. often do not have counterparts in experiments. Hence, Guala argues that "…the idea that certain key assumptions must be instantiated in the laboratory in order for it to account as a "proper" test of a model is misguided. A model is an entity made of many components each of which may or may not be a good counterpart to what goes on in a real situation" (Guala 2005, 221).

An account of models that explicitly separates the assumptions and deductions, on the one hand, and the causal hypotheses these models may inspire, on the other, naturally makes sense of these examples. Experimental economists treat models as open formulae which can be used to design many different kinds of experiments and make a number of different hypotheses. It is a virtue of the proposed account that it fits so seamlessly with the practice of experimental economics.

G. Improving analytic narratives

In the introduction to chapter 1 I claimed that the true cases of explanation on the basis of economic theory may not include some cases in which this theory is currently thought to be explanatory. I drew a distinction between casual and real

explanations. Now we have the resources to make this distinction clear. I shall do so via a discussion of a methodology that is gaining popularity among the modeling-minded social scientists.

Rational choice modeling is sweeping across political science and some areas of sociology. A methodology usually adopted in such approaches is sometimes referred to as "analytic narratives" (Bates et al 1998). The proposal is to build rational choice models to explain various historical phenomena, such as the stability of peace between clans of 12th century Genoa, the institutional foundation of US federalism, the rise and fall of the International Coffee organization, etc. The idea is to learn enough about a phenomenon so as to be able to mentally isolate a rational strategic interaction between actors, which can then be formalized in a rational choice model with some degree of similarity to the choices and trade offs the actors faced in the phenomenon in question. The model can be supplemented with a narrative, or a Geertzian "thick description" – i.e. an explanation of the meaning actors attach to their actions, circumstances and surroundings, that is their significance within the local culture. The rational choice model may predict an outcome that fails to occur or not predict a specific outcome at all (in case of multiple equilibria), in which case the narrative will be engaged to explain what other factors, which the model leaves out, account for the outcome. The methodology purports to unify nomothetic and ideographic approaches in the sense that the narrative captures the uniqueness of the situation under investigation while the model captures the general features of the type of phenomena this situation falls under.

The account developed here can be applied to partially articulate desiderata for a good analytic narrative. My account does not have anything to say on what makes a good thick description, but it does have normative implications for the use of models for purposes of formulating causal explanations. In an analytic narrative the model is supposed to play some explanatory role. What conditions must be met for a claim derived from a model to play this role well?

I shall use the example of how a game theoretical model called the Bargaining Game with Strategic Opposition developed by political scientist Kenneth Schultz (2001) can be used to explain the effect of domestic politics in democracies on their ability to bargain with other states. The game has three actors: government, its internal opposition and a target state. The government is in competition with the target state over some good, and in competition with the opposition for public office. The government has full control over foreign policy decisions, and the opposition has access to information about the expected value of going to war with the target. Because of the relative openness of public debate in a democracy, the target state has access to the opposition's official opinion on whether or not the government should go to war. The good in dispute is divisible and the actors are either risk-neutral or risk-averse. Government and opposition act so as to maximize their probability of (re)election. The target is assumed to value the good for what it is. The electorate in the challenging country values the government's ability to secure more rather than less of the good and at lower rather than higher price. The opposition's performance is judged on its support of high-return low-cost wars or criticism of high-cost low-return wars.

The government's strategy is a function of the expected value of war. Depending on this, the government will issue a genuine challenge or bluff. Opposition will seek to support use of force by the government if the expected payoff of war is high and hence will only support challenges whose probable outcome it considers attractive. Since the opposition has no motivation to collude with the government in a bluff, the target will be able to know the real costs of the potential conflict to the government and thus to judge the seriousness of the government intentions. If the opposition is rational and office seeking, a supported threat is more credible, as far as the target is concerned, than an opposed one. A credible threat increases the probability of the target conceding without a militarised conflict. A threat with which the opposition would be likely to disagree would carry less credibility and is more likely to be interpreted as a bluff by the target. Thus, democracies have fewer opportunities to exploit private information. The presence of the opposition party has a dual effect on the use of coercive diplomacy by democracies: *democracies threaten more selectively than non-democracies but when they do the threats are more effective* (for inducing concessions on the part of the target without fighting a war).

On the basis of models such as these, political scientists seek to explain contemporary or historical phenomena. For example, Schultz's derivation is supposed to account for the outcome of the Fashoda crisis on 1898, in which the British threat to challenge the French incursion into colonial Egypt was so effective. It also allegedly explains why Hitler was confident enough to remilitarize the Rhineland shortly before WWII. All these situations, Schultz argues, share the sort of

trade-offs that the participants in an international bargaining situation come up against when they are deliberating on a course of action, and these trade-offs are represented by the model.

What exactly must we demonstrate to maintain that the claim derived from the model does indeed explain an outcome such as the Fashoda crisis? On the view developed here we must show that a causal relation the model suggests (presence of opposition causes more credible or fewer threats) was indeed instantiated in the Fashoda crisis – i.e. that the behavior of the Liberal (opposition) party made the Tory Prime Minister's threat to France more credible. Such evidence can come from the records of what actors said and did at the time, how they understood and interpreted their actions and circumstances. Kenneth Schultz indeed provides much documentary evidence to this effect. His is, I believe, the correct use of a rational choice model for explanation. We really see that in the Fashoda crisis, the French gave up because they were convinced that if they stood firm the British would carry out their threat and send troops to retain that part of their colony. And we see that the British government knew that the opposition party would support their threat and went ahead with the threat for this reason. So indeed this is an example of a threat being more credible because it comes from a democracy. Historical records show that the Fashoda crisis is an instance of this causal relation. And we can state the conditions which made this relation hold in this particular case. To use the language of my account, we can give details of the material realization of the causal claim suggested by the Bargaining Game explained above.

This sort of explanation differs from what I called 'casual' explanation in the following way. When models are used casually, they are built so as to reflect certain features of the real world situations, features that the modeller considers key, or features that she knows how to incorporate into a model. The fact that there is a rough correspondence between some of the model's assumptions and the situation of interest is often the beginning and the end of the demonstration that the model is explanatory. On the view developed here, it is not enough to show that the model has similarities with the target system. What is missing in these cases is a clear articulation of a causal hypothesis and a demonstration that this causal relation indeed obtains. To return to the discussion of Sugden's view in chapter 1, casual use of models for explanation may be viewed in the light of what Sugden called 'credibility judgment'. We may be able to see that the model captures some processes we know to take place in the world. I argued that this is a reason to admit the model into the initial pool of explanatory resources, not to grant it the status of actually being explanatory. One of Sugden's examples is George Akerlof's Lemons Principle – imagine a used car market with certain idealized properties, it follows that when sellers know more than buyers about their cars, no sale takes place even though cars are worth more to buyers than to sellers. Akerlof remarks that something like the same principle is in operation in insurance markets, employment of minorities, credit markets in underdeveloped countries, etc. On my reading, he is not here committing to any specific causal claim or any explanation. Rather in the spirit of casual explanation he puts forward a rough schema that can generate hypotheses and explanations given further knowledge. The lesson for the methodology of analytic

narratives is that credibility analysis is only one stage of building an explanation of a phenomenon on the basis of a model. It is important not to stop at this stage.

F. Instrumentalism or realism?

Philosophical accounts of scientific theories tend to take a stand on what is perhaps the best known debate in philosophy of science – realism vs. instrumentalism. Realists believe that the contents of scientific theories, or at least some of them (usually the well-established theories in physical sciences) should be taken as stating true facts about the world. Realists are thus committed to existence of whatever unobservable entities and processes that are postulated by theories. Instrumentalists, on the other hand, treat scientific theories as doing something other than describing how the world really is, for example, organizing our experiences into useful classification schemes or working to improve our predictions. Instrumentalists are thus only committed to existence of entities and processes that we can directly perceive.

When this debate is applied to contemporary physical sciences, one challenge realists issue to instrumentalists is to defeat what is known the "no miracle argument" – given the amazing predictive and technological successes of our best theories it would be miraculous or at least highly improbable that these theories do not get something right about the structure of the world. When it comes to economics, the debate takes a different form, simply because successes of economics are much less obvious than those of physics. The predictive record of economics does not make

realism an obviously compelling position. So philosophers look for different evidence to decide among the alternatives.

For example, Alexander Rosenberg argues for an instrumentalist view of economics on the basis of the alleged metaphysical properties of what he sees as the principal components of economic theory – beliefs and desires (Rosenberg 1992). They are *intentional* categories: they are supposed to represent definite states of affairs but our best guesses as to how they represent the world imply that there is much *indeterminacy* in exactly what we believe and what we want. In addition to that, mental states are *intensional* in that substituting synonymous or co-referring terms risks changing their truth value. Brain states, on the other hand, do not have this property. Intensionality implies that the identity conditions for a belief will depend on other beliefs one maintains, which, Rosenberg argues, will prevent any attempt to divide beliefs into discreet units that the expected utility theory operates with. Given all this, the central categories of economics are doomed to have only as much empirical success as our commonsensical reasoning about people's choices and behaviors. However, for other philosophers, the fact that economic theory operates with such "commonsensibles" as beliefs and desires, of which humans have direct experience, is an argument in favor of realism about economics (Maki 1992). Can beliefs and desires be properly scientific categories? This approach takes us into the depths of the most difficult questions in epistemology and philosophy of mind.

I attempted to take a different approach, an approach that could potentially deliver answers that are not hostage to the difficult issues raised by Rosenberg. Rather than examining economic theory directly for its realist or instrumentalist credentials, I

sought to find out how it functions in cases of clear (or more or less clear) empirical successes. My reasoning has been as follows: take the best example of economics in action and infer from it to the status of economic theory. Where do the conclusions I reached put me on the realism vs. instrumentalism question?

The account of models I presented in this chapter is instrumentalist view. The most important instrumentalist claim I make is that economic models are partially filled-in formulae that serve as raw material for constructing empirical hypotheses. By themselves open formulae do not make any claims, explanations or even predictions. They are bases for generating causal hypotheses and their existence can give us reasons to treat a certain causal hypothesis as credible, to use Sugden's term, but credibility of a causal hypothesis is not the same as its confirmation. In this sense models are heuristical tools for exploring such hypotheses that may or may not rise to the status of assertions that are capable of explanation. This is the instrumentalist element of my account. Does that mean that this is an instrumentalist account of economics? Only if one thinks that models are all there is to economics.

So are there grounds for realism about economics knowledge? Economic models operate with categories that sound like categories we know to exist in the world. For example, that an auction is sealed-bid or open is an assumption that is clearly satisfied by some real world properties, for example the E-bay online auction is manifestly an open auction. That economic models make use of such commonsensibles is not in itself a good enough reason to adopt a realist attitude towards them. The objects in models are different from the homonymous objects of the everyday life in the sense that their behavior is more constrained. That by itself

does not exclude a realist interpretation. In particular, if the Mill/Cartwright view was correct about economics, we would have warrant to be realist about those parts of economic theory that yield knowledge of tendencies. Tendency claims describe action of isolated causes and thus describe at least parts of the real structure of the social world. Indeed this is the most thoroughly articulated view of modeling disciplines such as economics that warrants realism. However, applied to models in economics this view may be too good to be true – it is just not clear that models yield tendency or capacity claims. If so, then the burden is on the realist to give some other interpretation of these models that both makes sense of the practice of applied economics, as I have tried to, and makes realism a compelling position.

However, I see no reason not to be realist about causal claims inspired by models and confirmed by means of an investigation such as the one I described in chapter 2. The combination of rules, software and economic circumstances gives a good explanation of the outcomes of the FCC spectrum auction and we are reasonably warranted to be realist about the particular causal claim that underwrites this explanation. It also seems reasonable to be realist about the claim that endorsement of the government's threat by the opposition party is a cause of the Fashoda crisis unfolding as it did. There are many such local achievements across social sciences – achievements in which we have every reason to think that we understand the causal structure of the phenomenon in question. Although these successes do not warrant realism about economic theory, they do warrant realism about many knowledge claims in the applied areas of social sciences.

References

Alexandrova, A. (forthcoming) "Connecting Rational Choice Models to the Real World" *Philosophy of the Social Sciences.*

Bates, R.H., Grief, A., Levi, M., Rosenthal, J.L., Weingast, B.R. (1998) Analytic Narratives, Princeton NJ: Princeton UP.

Bechtel, W. and Richardson, R. C. (1993). Discovering complexity: Decomposition and localization as strategies in scientific research. Princeton: Princeton University Press.

Begg, I. and S.G.B. Henry (1998) Applied Economics and Public Policy Cambridge: Cambridge University 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. (1995) "Reply to Eels, Humphreys and Morrison" *Philosophy and Phenomenological Research* 55/1: 177-187.

Cartwright, N. (1998) "Capacities" in The Handbook of Economic Methodology, JB Davis, DW Hands and U Mäki (eds), Edward Elgar Publishing, 45-48.

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

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

Cartwright, N. (2002) "The limits of causal order, from economics to physics" in Fact and Fiction in Economics, Mäki, U. (ed), Cambridge: Cambridge University Press, 137-151.

Coase, R.H. (1993) "Law and Economics at Chicago" *Journal of Law & Economics*. 36/239: 248-49.

Coase, R.H. (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.

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

Crampton, P. (2002)"Spectrum Auctions" in Handbook of Telecommunications Economics (eds.) Cave, M., Majumdar, S. and I. Vogelsang, Amsterdam: Elsevier Science B.V., Chapter 14, 605-639.

Craver, C. (forthcoming) "When mechanistic models explain" *Synthese*

Cubitt, R.P. (2005) "Experiments and the Domain of Economic Theory" *Journal of Economic Methodology*, 12/2: 197-210.

Edward Nik-Khah. 2004. Designs on the Mechanism. PhD thesis University of Notre Dame.

Eric van Damme. (2004) "The Dutch UMTS auction" in Spectrum Auctions and Competition in Telecommunications, Gerhard Illing and Ulrich Klüh (eds.), MIT Press, 263-294.

FCC (1994a) "Second Report and Order". 94-61, FCC Docket No.

FCC (1994b) "Fifth Report and Order" 94-178, Washington, D.C., July 15.

Fodor, J.A. (1987) Psychosemantics. Cambridge Mass., MIT Press.

Forster and Sober 1994"How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions." *The British Journal for the Philosophy of Science* 45: 1 – 35.

Frank, R., T. Gilovich, and D. Regan (1993) "Does Studying Economics Inhibit Cooperation?" *Journal of Economic Perspectives* 7:159-71.

Friedman, M. (1953) "Methodology of Positive Economics" from The Philosophy of Economics (ed. D.M. Hausman, 1994, Cambridge: Cambridge University Press) reprinted from Essays in Positive Economics Friedman, M. (Chicago: University of Chicago Press 1953).

Frigg, R. and S. Hartmann (2006) "Scientific Models" in The Philosophy of Science Encyclopedia, eds. S.Sarkar and J.Pfeiffer, New York, Routledge.

Gibbard, A. and H.R. Varian (1978) "Economic Models" *Journal of Philosophy* 75: 664-77.

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

Gigerenzer, G. (forthcoming) “Striking a Blow for Sanity in Theories of Rationality” to appear in Augier, B., and March, J.G. (eds), Essays in honor of Herbert Simon, Cambridge, MA: MIT Press.

Gigerenzer, G., and R. Selten, (eds) (2001) Bounded Rationality: The Adaptive Toolbox. Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P.M., and the ABC Research Group 1999. Simple Heuristics that Make Us Smart. New York: Oxford University Press.

Glennan, S. (1997) Capacities, Universality and Singularity *Philosophy of Science* 64:4, 605-626.

Godley, W. (1998) "Using Figures to Guide Macroeconomic Policy" in Begg and Henry 1998, 258-263.

Goldstein, D. and Gigerenzer, G. (1999) “The Recognition Heuristic: How Ignorance Makes Us Smart” in Gigenerzer at al. 1999, 37-58.

Guala, F. (2002) "Models, Simulations, and Experiments", in Model-Based Reasoning: Science, Technology, Values, L. Magnani and N.J. Nersessian (eds). New York, Kluwer, 59-74.

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

Guala, F.(forthcoming) “Has Game Theory Been Refuted?” *Journal of Philosophy.*

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

Hausman, D.M. (1992b) “Why look under the hood?” in Essays in Philosophy and Economic Methodology. Cambridge: Cambridge University Press, 70-73.

Hausman, D.M. (1994) "Paul Samuelson as Dr. Frankenstein: When Idealizations Escape and Run Amuck" *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *Idealization in Economics*. B. Hamminga and N. de Marchi, (eds). Amsterdam: Rodopi, 229-43.

Hendry, D.F. and G.E. Mizon (forthcoming), “Forecasting in the Presence of Structural Breaks and Policy Regime Shifts” in Identification and Inference in Econometric Models: Festschrift in Honor of T.J. Rothenberg, D.W. Andrews, J.L. Powell, P.A. Ruud and J. Stock (eds). Cambridge: Cambridge University Press.

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

Howson, C. (2000) Hume's Problem: Induction and the Justification of Belief. Oxford: Clarendon Press.

Humphreys, P. (1995) "Abstract and Concrete" *Philosophy and Phenomenological Research*, 55/1:157-161.

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

Kagel, J.H. and D. Levin (1986) "The Winner's Curse Phenomenon and Public Information in Common Value Auctions", *American Economic Review* 76:894-920.

Keen, S. (2001) Debunking Economics, Pluto Press: Australia.

Klemperer, P. 1999 "Auction Theory: A Guide to the Literature", *Journal of Economic Surveys* 13 (3): 227-286.

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.

Krishna, V. (2002) Auction Theory, Elsevier Science.

Laffont, J-J. (1997) "Game Theory and Empirical Economics: The Case of Auctions Data" *European Economic Review*, 41: 1-35.

Lucas, R. (1981) Studies in Business Cycle Theory, Oxford: Blackwell.

Magnus, PD. and C. Callender (2004) "Realist Ennui and the Base Rate Fallacy" *Philosophy of Science* 71:320-338.

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

Marwell, G. and R.E.Ames (1981) "Economists Free Ride, Does Anyone Else?" *Journal of Public Economics* 15:295-310.

Maskin, E.S. and Riley, J.G. (1985) "Auction Theory with Private Values" *American Economic Review*, 75, 150-55, qtd. in Klemperer 1999, p.236.

Maudlin, T. (1993) “Review of Nancy Cartwright’s Nature’s Capacities and Their Measurement” *Journal of Philosophy* XC: 599-603.

McAfee, R.P. and J. McMillan (1996) “Analyzing the Airwaves Auction” *Journal of Economic Perspectives*, 10/1: 159-175.

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

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

McMullin, E. (1985) “Galilean Idealization” *Studies in History and Philosophy of Science* 16/3: 247-273.

Milgrom, P. (1989) “Auctions and Bidding: A Primer” *Journal of Economic Perspectives* 3/3: 3-22.

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

Morgan, M.S. (2001) “Models, Stories, and the Economic World” Journal of Economic Methodology 8/3: 361-84.

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

Morgan, M.S. (2003) “Experiments Without Material Intervention” in The Philosophy of Scientific Experimentation, ed. Hans Radder, Pittsburgh, University of Pittsburgh Press.

Morgan, M.S. (2005) “Experiments versus models: New phenomena, inference and surprise” *Journal of Economic Methodology*, 12/2: 317-329.

Morgan, M.S. and M. Morrison (1999) Models as Mediators, Cambridge: Cambridge University Press.

Morrison, M. (1995) "Capacities and the Problem of Singular Causes" *Philosophy and Phenomenological Research* 55/1: 163-68.

Morton, R. (1999) Methods and Models, Cambridge: Cambridge University Press.

Musgrave, A. 1981 “Unreal Assumptions in Economic Theory: The F-Twist Untwisted” *Kyklos* 34: 377-387.

Odenbaugh, Jay (2003) "True Lies: Robustness and Idealization in Ecological Explanation", Lewis and Clark College, manuscript.

Osbourne, M. (2003) An Introduction to Game Theory, M.J.Osborne, New York: Oxford University Press.

Pemberton, J. (2004) "Why Idealized Models in Economics have Limited Use" *Poznan Studies in the Philosophy of the Sciences and the Humanities, Idealization XII: Correcting the Model*. M. R. Jones and N. Cartwright (eds), 35-46.

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

Press Release: The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 2005, 10 October 2005.

Putnam, H. (1975) Philosophical Papers, Volume 1: Mathematics, Matter and Method, Cambridge, Cambridge University Press.

Reiss, J. (2005) "Do We Need Mechanisms in Social Science?" manuscript, Complutense University, Madrid.

Rosenberg, A. (1992) Economics: Mathematical Politics or Science of Diminishing Returns, Chicago: University of Chicago Press.

Rosenberg, A. (1994) Instrumental Biology or the Disunity of Science, Chicago: University of Chicago Press.

Salmon, W. (1984) Scientific Explanation and the Causal Structure of the World, Princeton, NJ: Princeton University Press.

Schelling, T. (1978) Micromotives and Macrobehavior, W. W. Norton & Company.

Schultz, K. (2001) Democracy and Coercive Diplomacy Cambridge: Cambridge University Press.

Simon, H. (1994) "Testability and Approximation" in Philosophy of Economics, ed. Hausman, D.M. Cambridge: Cambridge University Press, 214-216.

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.

Weisberg, M. (forthcoming) "Robustness Analysis" *Philosophy of Science.*

Wimsatt, W. (1997). "Aggregativity: Reductive heuristics for finding emergence" in L.Darden, ed. *PSA-1996*, v.2. *Philosophy of Science* 66: S372-S384.

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

Woodward, J. (2002) "There is no such thing as a Ceteris Paribus Law" *Erkenntnis* 57/3: 302-328.