4. KINDS OF THEORY

ERIC MASKIN*

INTRODUCTION

Usually economic theory is conveniently divided into just two categories: the positive and the normative. But in my view, this is far from a complete taxonomy. This is not to say that "positive" and "normative" are useless labels. It is simply that, by themselves, they leave too much out.

In this paper I will argue that there are at least four other sorts of theory that do not readily fall into the traditional categories, viz., theory as parable, axiomatic theory, as benchmark, and theory as tool kit. To make this argument, I will draw on examples from my own work. I use these particular illustrations not because they best exemplify the particular categories that I propose but because only in the case of my own work can I be sure of the motives behind the research inquiry. And, as we will see, intentions play an important role in the question of how theories are classified.

THE CONTRAST WITH NATURAL SCIENCE

Let me begin by contrasting the nature of theory in economics with that in the natural sciences. In science, by and large, there is only one kind of theory: positive theory. That is, given some observations to be explained, the purpose of theory is to explain them as simply, generally, and deeply as possible (of

*This paper is a revised version of a paper that was delivered in the Harvard Graduate Economics Forum in February 1997. I would like to thank the NSF for research support. It is dedicated to Ryuzo Sato, who, through the Technical Symposia that he oversees at the Japan-U.S. Business and Economic Center, has fostered many kinds of theory.

course, the objectives of simplicity, generality, and depth may not always be in harmony).

Why should economics be different? One important reason, in my view, is that economics is about *human* phenomena, i.e. phenomena that are, in some sense, under our control (or at least give the impression of being under our control). So, we are inclined to ask not only how things are but also how they *could* be. And once we are in the realm of the "could be," we have entered into normative economics.

But there is at least one other important explanation for the contrast, one that has often been remarked on—and lamented—by economic empiricists, viz., the fact that economic data tend to be far poorer than those in the hard sciences, particularly those hard sciences where laboratory experiments are possible. For this reason, economic theories must typically do a good deal more than simply "fit the facts," since that minimum criterion is often all too easy to achieve.

POSITIVE THEORY

Partly to provide a point of reference, let me first outline a more-or-less standard piece of positive theory. The particular issue is income and wage inequality.

It is a well-documented fact that income and wage inequality rose in the United States (and a number of other industrial countries) over a period of about twenty years (from the mid-1970's to the mid-1990's). An open problem is to identify the mechanism behind this trend.

Michael Kremer and I (Kremer-Maskin (1996)) suspected that job-matching might have something to do with rising inequality. We had been studying asymmetric assortative matching models and noticed something interesting. Suppose that a firm consists of two "tasks," one managerial, the other productive. Assume furthermore, that the firm's output depends on the skill levels of the employees carrying out those tasks. Specifically, suppose that, if s_m and s_p are the skill levels of the "manager" and "producer" respectively, then output is $s_p s_m^b$, where s_p and s_m are positive numbers. That is, the managerial task is more sensitive to skill than is the productive task.

Now, suppose that there are many firms and many potential employees in the economy. Assume that the population of employees divides into those of low skill level L and high skill level H, where L and H are positive numbers. In competitive equilibrium, employees will be assigned to firms *efficiently*, i.e. in a way that maximizes total output (let us suppose that this is a one-good economy).

An interesting question to ask about equilibrium is whether, within firms, L-employees are paired with other L-employees and H-employees with other H-employees or whether there is "cross-matching": L-employees are matched with H-employees. This turns out to depend on the relative values of L and H. Total output from two cross-matches (two matches between an L-employee and an H-employee) is

$$2LH^3 \tag{1}$$

(it is clearly efficient to put the H-employee in the managerial task, since that is more sensitive to skill). If instead there is no cross-matching, these four employees generate output

$$L^3 + H^3. \tag{2}$$

Now, if

$$H < \left(\frac{1+\sqrt{5}}{2}\right)L, \tag{3}$$

then one can verify that (1) exceeds (2). That is, when the dispersion of the skill distribution is relatively small, equilibrium implies that high and low skilled workers are found together in the same firm. By contrast, if

$$H > \left(\frac{1+\sqrt{5}}{2}\right)L, \tag{4}$$

then (2) exceeds (1), and so firms will be skill-segregated in equilibrium.

What does this have to do with inequality? Well, let us suppose that, over time, the value of H rises, i.e. the dispersion, mean, and upper tail of the skill distribution all increase, whereas the lower tail remains about the same (these changes in the distribution correspond to historical reality in the U.S. in the period under consideration). As long as (3) holds, the wage going to an L-employee will be

$$LH^2 - \frac{1}{2}H^3 \tag{5}$$

(to see this, note that the wage going to an L-employee is what is left of total product LH^2 after the wage $\frac{1}{2}H^3$ to his partner is subtracted.) Now, formula (5) is increasing in H for H sufficiently near $L(H < \frac{4}{3}L)$. But for $H > \frac{4}{3}L$, the L-employee's wage is actually declining in H. This implies that in addition to the direct increase in the wage dispersion induced by the increase in the skill dispersion, there is an additional increment to wage dispersion brought about by the decline in L-employee wages. A modest increase in skill dispersion is magnified into a substantial rise in wage inequality.

But there is an unexpected prediction of the model, a common occurrence in the realm of positive theory. Once H rises enough so that (4) holds, cross matching ceases. In other words, the typical firm's employees become more homogeneous (from the standpoint of skill) over time. This implication seems to be confirmed by the data we have assembled for the U.S., U.K., and France.

Thus we have an example of the classic phenomenon in which the attempt to explain a puzzling stylized fact—in this case, an increase in skill dispersion accompanied by a considerably larger increase in wage inequality—

the empirical validation of this prediction lends further credence to the theory.

THEORY AS PARABLE

Often in economics we build models not so much to predict, but to formalize a loose intuition we have for why things work as they do. There is certainly a positive aspect to this sort of theorizing. But the objective is to build a model that, rather than forming the basis of an econometric study, tells a story.

Here is an example of one of these intuitions, in this case having to do with money. Money has long been something of an embarrassment to economic theory—everyone agrees that it is important, but in much of our analysis it is simply left out. For example, it plays no role in the most highly developed representation of a competitive economy, the Arrow-Debreu model. One exception to the rule that money plays no role in our economic constructs is the overlapping generations consumption-loan model, where money serves as a store of value from one period to another. But even there, once we introduce other assets that survive over time—land, for example—money loses its central value-storing purpose.

From every day experience, of course, we know money acts not only as a store of value, but also as a medium of exchange. In a world of barter, if I have apples but want bananas, I must find someone wishing to exchange bananas for apples in order to trade. That is, there must be a *double coincidence of wants*. By contrast, in a monetary world, I can first sell the apples for money and then use the money to buy bananas, eliminating the need for the double coincidence.

This elimination is supposedly one of the great advantages of money. But the apples-and-bananas story I've told does not completely clinch the argument. After all, even if there is no money, a banana seller who is completely uninterested in apples per se might still be willing to accept my apples as payment provided she could expect to retrade them later. In other words, money is not needed if apples can serve as the medium of exchange.

Still, we retain the intuition that apples may prove an awkward medium. Indeed, the banana seller may be reluctant to accept apples because she cannot properly evaluate them; she may simply not know very much about apples. If she believed that I knew as little about apples as she, this imperfect information might not matter very much: the exchange rate between apples and bananas would be set by using the apples' expected value. However, the very fact that I am selling apples may suggest to the banana seller that I know more about them than she, in which case she will be suspicious. This suspicion will make it difficult for me to get a reasonable price for any apples other than those of low quality. That is, trade may be significantly constrained by adverse selection.

As traders, we are reasonably familiar with the goods we buy or sell on a regular basis. But there is a vast array of goods with which we have little

experience. Furthermore, different traders have different sets of goods that they know well. So, if someone tries to sell us something that we don't know much about, we become wary of being exploited.

From this perspective, the essential nature of money becomes clear: it is simply a good that is well enough known to every trader to overcome the adverse selection problem.

Abhijit Banerjee and I decided to try to embody this story in a formal but very simple model, in order to bring out the intuition more clearly. The idea was to adhere as closely as possible to the standard competitive model. The model has the following elements. There are three types of traders, A, B, and C and three goods, also labeled A, B, and C. For each X = A, B, C a trader of type X can produce good X (and only good X). A-traders consume good B (and only good B), B-traders consume C, and C-traders consume A. Each good can be of two qualities, high and low. The high quality good generates more utility but is more costly to produce (producers can divide their fixed input endowments in any way they choose between high- and low-quality goods). A consumer or producer of good X can distinguish between the two qualities, but nobody else can. Trade unfolds in a finite number of trading periods and is always bilateral (between one buyer and one seller).

One conclusion that emerges directly from this model is that there has to be a medium of exchange (since traders are not self-sufficient) and that any medium of exchange must be a *low-quality* good (since no two traders have two goods in common for which both can distinguish the qualities). Thus the model delivers a formalization of the classic law, due to Gresham, that bad money drives out good. Furthermore, if the ratio of the marginal benefit to marginal cost exceeds the ratio for low-quality goods, then producing low-quality goods is economically inefficient. Thus the fact that such goods *have* to be produced in order for trade to be viable constitutes the *true* cost of barter: the distortion of production toward low-quality goods. From this standpoint, the advantage of fiat or paper money is that it eliminates the production distortion by providing an alternative to low-quality goods as a medium of exchange.

This model very much plays the role of "parable." It is far too simple and stylized to constitute the basis of any sort of empirical work, but it serves to make precise our informal understanding of the drawbacks of barter and the advantages of fiat money.

AXIOMATIC THEORY

Often in the social sciences we are confronted with the task of comparing different institutions. Consider, for example, the various electoral methods used in practice. There is majority rule (candidate a is "socially" preferred to candidate b if a majority of voters prefers a to b), plurality rule (a is "socially" preferred to b if more voters rank a first than rank b first), rank-order voting (a is "socially" preferred to b if a receives more total points than

b, when each voter assigns m points to his favorite candidate, m-1 to his next favorite, and so on, and m is the number of candidates on the ballot), and a good many others. How do we go about comparing these different methods' strengths and weaknesses?

One systematic way of proceeding is axiomatically. That is, we might first consider what are the desirable or reasonable properties that we would want an ideal electoral method to satisfy. We can then investigate the extent to which particular methods satisfy these axioms.

This is the approach that Partha Dasgupta and I (Dasgupta-Maskin (2000)) have taken recently. An obvious axiom to begin with is the *Pareto* principle, the natural and fairly weak requirement that if all voters prefer candidate a to candidate b, then a should be socially preferred to b. A second axiom—reflecting the fundamental democratic idea that everyone's vote should count equally—is the principle of *anonymity*: suppose that for some particular preferences on the part of voters, a is socially preferred to b. Now permute voters' preferences, so that, say, voter i now has what used to be voter j's preferences and voter k now has what used to be voter i's preferences, etc. Anonymity requires that it should remain the case that a is socially preferred to b.

Just as anonymity in effect, means symmetry across voters, so the next axiom, *neutrality*, requires symmetry across candidates: suppose that, for some particular preferences of voters, a is socially preferred to b. If we now consider some other voter preferences with the properties that (1) if a voter preferred a to b under the original preferences, he prefers c to d under the new preferences, and (2) if a voter preferred b to a under the original preferences, he prefers d to c under the new preferences. Then c should be socially preferred to d.

Finally, *transitivity* requires that social preference be consistent: if a is socially preferred to b and b is socially preferred to c, then a should be socially preferred to c.

If voters preferences can be *anything*—that is, they are completely unrestricted—then the Arrow impossibility theorem (Arrow (1951)) implies that *no* electoral method can satisfy all the above principles. For example, majority rule satisfies all the principles except transitivity. However, matters are different if we can restrict voters' preferences to some limited class. Majority rule, for instance, satisfies even transitivity when preferences are "single-peaked" (see Black (1948)). What Dasgupta and I have shown, in fact, is that if some electoral method (different from majority rule) satisfies these principles for a restricted class of preferences, then majority rule also satisfies the principles on that class. Furthermore, there exists some other class for which majority rule satisfies the principles and the other electoral method does not. This is the sense in which majority rule "works" more often than any other electoral method.

We would submit that this is an important virtue of majority rule, possibly one that accounts for its longevity as a practical method. But such a virtue would not be apparent without an axiomatic analysis.

THEORY AS BENCHMARK

Sometimes theorists formulate a model not because it describes the world as it is, or even as we might hope it to be, but rather because it serves as a useful benchmark or idealization against which we can compare failures of the ideal.

The Arrow-Debreu model of a competitive economy (Arrow-Debreu (1954)) provides an example. Actually, this model serves two functions: it is often used as a positive theory. That is, there are situations in reality that are considered close enough to the Arrow-Debreu ideal so that the model applies. But more often there is, in practice, some failure—of perfect competition, perfect information, or complete markets—and, by comparing the outcome of the idealized model with the actual situation, the theorist may learn the implications of the hypotheses being violated.

Another prominent example of a benchmark is the Coase Theorem, which asserts that, in the absence of transaction costs, bargaining should lead to an efficient allocation. Of course, in practice transaction costs are rarely zero, and so comparisons of actual with Coasian outcomes are illuminating.

Recently Jean Tirole and I (Maskin-Tirole (1999)) have performed a benchmark analysis as applied to contractual arrangements. In the incomplete-contract literature (see Hart (1995)) it is typically assumed that when parties write contracts, there are transactions costs associated with specifying or even foreseeing most contingencies. Thus, contracts will necessarily be far less highly contingent than the parties might wish them to be.

Tirole and I point out, however, that there is a conflict between two important hypotheses of this literature: the assumption of significant transaction costs and the assumption that agents can foresee the future well enough to perform dynamic programming. More specifically, we demonstrate that even if transaction costs prevent agents from describing physical contingencies ex ante, they do not constrain the set of payoffs that can be achieved through contractual arrangements.

The idea behind this demonstration is straightforward. If parties have difficulty foreseeing physical contingencies, they can write contracts that ex ante specify only the possible *payoff* contingencies. Then, later on when the state of the world is realized, they can "fill in" the physical details. The only serious potential obstacle to such an arrangement is incentive-compatibility: will it be in each agent's interest to specify these details accurately for the state in question? But here the techniques of the implementation literature (see Palfrey (1998)) can be invoked to ensure that agents have the right incentives.

Now, one might ask the question, "Do we observe such contracts in practice?" The answer is, "Occasionally, but not very often." But that response does not constitute a relevant criticism of our theory. After all, the theory is there to provide a benchmark. If experience does not conform to theoretical prediction, then it behooves us to investigate why not, rather than simply dismissing the theory.

THEORY AS TOOL KIT

It is commonplace for dynamic game-theoretic models to generate multiple equilibria. Indeed, in the best-known dynamic game of all—the repeated prisoners' dilemma—the "Folk Theorem" (see Fudenberg and Maskin (1986)) establishes that there is a continuum of equilibria. Although such a multiplicity may be of theoretical interest, it creates a real headache for the applied economist who wishes to use the model to make predictions. Sometimes the predictive power of a model can be improved if the analyst is able to eliminate certain equilibria as theoretically "implausible." A classic example of such an elimination is the use of subgame perfect equilibrium in games such as that of Figure 1.

The "story" behind the Figure 1 game is that a potential entrant (Player 1) is contemplating entering a market in which there is already a monopolistic incumbent firm (Player 2). If Player 1 stays "Out," Player 2 gets all the profit (a total of 2) to itself. But if Player 1 comes "In," then Player 2 has the choice either to "Accommodate"—in which case the two firms split the market equally—or to "Fight," causing both firms to be harmed.

There are two sorts of equilibria in this model. First, there is the equilibrium (In, Accommodate). Second, there is a class of equilibria in which Player 1 chooses Out and Player 2 randomizes between Accommodate and Fight (with sufficient weight on the latter to make Player 1's choice a best-reply). There is in a real sense, however, in which this latter class is implausible. Specifically, if it happens that Player 1 actually chooses In, then Player 2 should obviously choose Accommodate—to choose Fight would be self-destructive. Thus, in any equilibrium where Player 2 chooses Fight with positive probability, Fight serves only as a threat to induce Player 1 to choose Out—and an empty threat at that, since it would not be carried out were Player 2 put to the test.

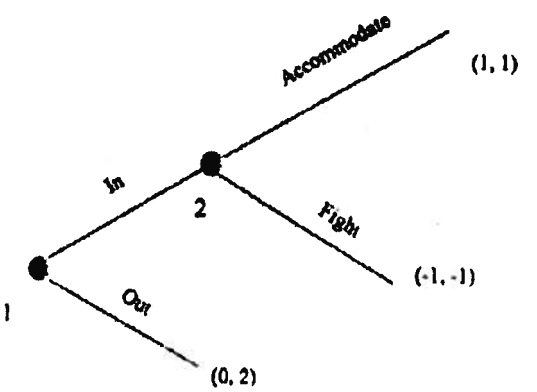

Figure 1. A game of Potential Entry

Reinhard Selten (Selten (1965)) formalized the implausibility of the (O Fight) equilibria in which Player 1 chooses Out. Specifically, this equili rium fails to be *subgame-perfect*, i.e. its strategies do not constitute an eq librium at all points in the game (indeed, they fail to form an equilibrium the point where Player 2 moves). The concept of subgame-perfect equili rium has, in effect, become an important tool for game-theoretic analysts that for a large class of games—including the example of Figure 1—it eli inates equilibria based on empty threats.

In recent work Jean Tirole and I (Maskin and Tirole (forthcoming)) ha attempted to provide another such tool in the form of Markov equilibriu Roughly speaking, a Markov equilibrium is one in which strategies depe only on those past variables that remain payoff-relevant in the present. F example, consider a dynamic game in which, in every period t, each play i's payoff π_t^i depends on the actions a_t chosen in period t and the state of t system θ_t: $\pi_t^i = \pi_t^i(a_t, \theta_t)$. Suppose furthermore that the state θ_t is a functi of last periods actions and state: $\theta_t = f_t(a_{t-1}, \theta_{t-1})$. Finally, suppose that Play i's overall payoff is a discounted sum of his period payoffs:

$$\sum_{t=1}^{\infty} \delta^{t-1} \pi_t^i,$$

where δ is the discount factor. Now, in principle, Player i's choice of acti in period t could depend on all previous actions $a_1, \ldots, a_{t-1}$ and all sta $\theta_1, \ldots, \theta_t$. But in fact, none of these variables except θ_t actually affects a of the players' payoffs from period t onwards. Hence a Markov equilibriu will make player i's period t behavior depend only on θ_t. Of course, in t example, it is easy to say what a Markov equilibrium is because the payo relevant state θ_t is defined exogenously. In a more general game, identifyi the payoff-relevant variables need not be so straightforward, which accou for Tirole's and my effort.

But, in any case, work on the concept of Markov equilibrium is clearly exercise that is neither normative nor predictive. Rather it is the attempt build a piece of machinery that will aid in the positive enterprise.