Bounded Rationality, Indeterminacy, and the Theory of the Firm

Roy Radner

BOUNDED RATIONALITY, INDETERMINACY, AND THE THEORY OF THE FIRM*

Roy Radner

Two issues have become increasingly apparent in attempts to apply the prevailing notions of 'economic rationality' to the theory of the organisation of business firms. The first goes under the rubric of *bounded rationality*. This is hardly a new idea, and has been forcefully brought to our attention in the work of Simon and others, although awareness of the problem (in economics) goes back at least to Clark (1918). I shall try to provide a more detailed taxonomy of bounded rationality than is usually done. In particular, it is important to distinguish between (1) costly rationality, like the costs of observation, communication, and even computation, that require only an extension of the standard 'Savage Paradigm', and (2) truly bounded rationality, like not knowing the implications of everything that one knows, which – as far as I know – goes far beyond the Savage paradigm.

The second issue, which I shall call *indeterminacy*, arises in attempts to apply the theory of strategic games to models of organisations, namely, one often faces a very large multiplicity of solutions, which significantly weakens or even destroys the predictive power of the theory. By 'solution' I mean here the so-called non-cooperative equilibrium, usually associated with the names of Auguste Cournot and John Nash (and extended and refined by John Harsanyi, Reinhard Selten and others to cover games in which the players have incomplete information). Although indeterminacy can arise even in static games with complete information, it seems to be especially prevalent in dynamic games and/or games with incomplete information.

Both of these issues have profound implications for the organisation of the firm. In discussions of this topic, it has become commonplace to set up the straw figure of the firm as a black box, mysteriously choosing production and investment plans to maximise profits. It is true that there are still undergraduate price theory textbooks that present this picture, but economists have long been familiar with the phenomenon of the separation of ownership and management, at least since the publication of the book by Berle and Means (1932), if not before; in fact, Adam Smith had some trenchant comments on it (1776, pp. 264–5). This phenomenon is also referred to as *the separation of ownership and control* (see Fama and Jensen, 1983). Today, one is likely to see this analysed in terms of the 'principal-agent' model, with the owner(s) cast as the principal, and the manager(s) as the agent.

Going beyond this model, it is also recognised that there are many owners, managers, and workers. One approach has been to visualise the firm as a 'nexus of contracts' (Jensen and Meckling, 1976). However, it is clear that many, if not most, of the relationships and activities within the firm, or even between the owners and the managers, are not closely determined by contracts in the usual sense of the word, but by expectations about the roles of the various actors. This leads to the idea of 'incomplete contracts', in which many contingencies are not explicitly covered by the terms of the contract, and yet the parties to it have expectations about the probable outcomes in those contingencies. If these expectations are mutually consistent and self-fulfilling, then we might model an 'incomplete contract' as a two-part strategic game: (1) the parties first bargain about the explicit terms in the contract, to which they are committed, and then (2) the parties' post-contract actions are determined by strategies that form an 'equilibrium' of the post-contract strategic game.

(Of course, all contracts are to some extent incomplete (see, e.g., Kornhauser, 1983), but intra-firm contracts may be even more incomplete than those between firms. For more on incomplete contracts and related material see Hart (1989; 1991) and Williamson and Winter (1991).)

Finally, the *rules of the game* are themselves subject to change, by evolution, by mutual consent of the interested parties, or by the direct intervention of an outside party. The rational choice of the rules of the game is sometimes called the problem of *mechanism design*.

The paradigm that is implicit in the above account is open to two serious criticisms. First, the cognitive requirements placed upon the individual decision-makers, not to mention the mechanism designer, in such a model far exceed the capacities of present-day humans and computers. This is the phenomenon of 'bounded rationality'. Secondly, these models typically do not yield sharp predictions. This is the phenomenon of 'indeterminacy'. Notice that bounded rationality is a crucial attribute of human *behaviour* that is not adequately accommodated in our theories. On the other hand, indeterminacy is an attribute of our *theories* that prevents us from making sharp predictions about behaviour.

In this essay I shall try to provide a taxonomy of bounded rationality that is more detailed than one usually finds in discussions of the topic by economic theorists. I shall also illustrate the phenomenon of indeterminacy in a simple model of mechanism design, namely, a model of bargaining about transfer prices within a firm. I hope that this discussion will lend some plausibility to my judgement that economists will not make further progress on the theory of the organisation of firms until we can deal more successfully with both of these phenomena.

I. COSTLY AND BOUNDED RATIONALITY; DECENTRALISATION OF INFORMATION AND DECISION

I.I. *The Savage Paradigm of Decision-Making Under Uncertainty*

Although the themes of this section are bounded rationality and multi-person decision-making, I start with a brief account of the paradigm of individual rational decision-making under uncertainty that is dominant in economics and game theory today. This paradigm is primarily associated with the name of L. J. Savage, although Savage himself traced it to Bernoulli, Laplace, de Finetti, von Neumann and Morgenstern, and others. In any case, Savage's book, *The Foundations of Statistics* (1954), remains the deepest and most coherent account of what I shall call here the 'Savage paradigm'.

Savage developed his theory primarily to illuminate and rationalise statistical practice, and to resolve some of the controversies that arose about statistical methodology following the appearance of the revolutionary ideas introduced by R. A. Fisher and J. Neyman. Ironically, his theory had little effect on statistical practice, but it was embraced by economists and game theorists as a powerful model of how decision-makers should respond rationally to uncertainty and information.

The theory had three central themes. First, it laid down – as axioms – certain minimal principles of consistency of decision in the face of uncertainty, which, in Savage's view, embodied all of the notions of rationality that could reasonably be imposed.

Secondly, it showed that these principles of rationality implied that a decision-maker's choices among alternative actions could be calculated as a function of two 'psychological' scales: (1) a numerical scale of probabilities of events, and (2) a numerical scale of the utilities of outcomes. More precisely, the rational decision-maker would prefer an action that yielded the highest mathematical expectation of the utility of outcome, the expectation being defined relative to the decision-maker's scale of probabilities.

Thirdly, the theory implied how a decision-maker should rationally modify his decisions in the light of new information, or more generally, how the choice of actions should evolve in a sequential decision problem as new information is accumulated. Thus, although at the beginning of their lives as rational decision-makers, persons might reasonably differ in their beliefs about the likelihood of various events, they must all follow the same *rules* (formulas) for updating those beliefs in the light of new information. (Essentially, these are the rules of the calculus of conditional probability, as exemplified by Bayes's Rule.) Since most realistic models of economic decision-making involve sequential decision, this is a crucial part of the theory in terms of its implications for rational behaviour, and will also figure prominently in the discussion of bounded rationality.

As sketched above, the Savage paradigm appears not to take account of the resources used in the process of decision-making itself. Savage was, of course, aware of the fact that decision-making requires resources, and provided a

lengthy discussion of this matter. Nevertheless, he chose not to incorporate the costs of decision-making explicitly into his model. It is convenient to classify the costly (resource-using) activities of decision-making into three groups:

1. *observation*, or the gathering of information;
2. *memory*, or the storage of information;
3. *computation*, or the manipulation of information.

In addition, when we consider groups of decision-makers, we shall have to take account of the costs of:

4. *communication*, or the transmission of information.

Of these activities, and their related costs, the first, second, and fourth can be accommodated by the Savage paradigm with relatively little strain, although they do have interesting implications. On the other hand, as we shall see below, a serious consideration of the costs of computation leads to serious doubt about the realism – or even approximate realism – of the paradigm as a theory of human decision-making.

I.2. *Costly Observation and the Wald Paradigm*

In his book, *Statistical Decision Functions* (1950), Abraham Wald codified the idea, already implicit in the Neyman–Pearson theory of hypothesis testing, that statistical procedures could – or even should – be regarded as methods for making decisions in the face of uncertainty. Perhaps stimulated by his own pioneering studies of sequential statistical procedures, he also explicitly introduced the cost of observations into his general model of statistical decision-making. The classic example that is said to have led Wald to develop his theory of sequential analysis is that of destructive testing in acceptance sampling. Typically, in order to fully test an artillery shell one must fire it, and thus destroy it (unless, of course, it fails to fire, in which case it is a dud, and also useless). Before deciding to accept a lot of, say, 1,000 shells from the manufacturer, the U.S. Army would typically test some fixed number of shells from the lot, and accept the lot only if the number of defectives in the sample were less than some specified level. Since each observation (test of one shell) was costly, it was important to minimise the number of shells needed to achieve a given standard of reliability of the acceptance procedure. Wald showed how this could be done by a *sequential* procedure; instead of fixing the sample size in advance, the decision-maker should decide after each successive test whether to (1) continue testing, (2) stop testing and accept the remaining shells in the lot, or (3) stop testing and reject the lot. In fact, under certain conditions, the optimal sequential testing procedure produced significant savings as compared with any fixed-sample-size procedure. (Wald (1947) was the first to develop a systematic theory of sequential statistical procedures; see the Introduction of this last reference for a history of previous research on the topic.)

   Other interesting cases of 'costly' observation arose in the study of clinical trials. Suppose that several new drugs to treat disease X have been developed

to the point where they can be tested on human patients. In addition to the out-of-pocket costs of administering the drugs (which may be different), there is also the 'opportunity cost' of administering an inferior drug to a patient who might otherwise have had a greater probability of a successful treatment. Again, the optimal procedure is a sequential one, in which a decision about whether to continue testing a new drug (and whether or not to adopt it) is made after observing the results for successive patients or groups of patients. (See Basu *et al.*, 1990.)

In both of the above examples, there is no difficulty in principle in fitting the decision problem into the Savage paradigm, although taking account of the costs of observation will typically complicate the analysis. In fact, it is interesting to note that problems of clinical trials are notoriously difficult to solve analytically, and the numerical computations needed to approximate an optimal policy fairly well are typically very demanding.

In these examples, it is also necessary for the decision-maker to remember the results of previous observations. If the observations are complicated, the memory required to do this might be costly, or even infeasible in very large problems. The costs of memory are in some sense analogous to the cost of observation, but not identical. (For a discussion of costly memory in strategic games, see Osborne and Rubinstein, (1994, ch. 9).)

I.3. *Computation, Communication, and Decentralisation*

The decision-maker who follows the Savage paradigm will typically have to perform some computations in order to determine what actions should follow from his particular observations. It will be useful in what follows to distinguish decisions that are relatively routine, or periodic, from those that are relatively unique, or 'one time only', and start our discussion with the former. For example, manufacturers periodically update their production plans on the basis of observed sales, market research, and observations of the economy. Firms periodically process new financial data to determine dividends, short-run borrowing, etc. In a large firm, such decision cycles involve the collection and processing of vast amounts of data, and the calculation of hundreds or thousands of individual decisions. This computational activity is far beyond the capability of any single human decision-maker, even when armed with the most powerful computer. The computational task involved in corporate decision-making is, therefore, divided among many humans and machines. *Inevitably, the activities of information processing for decision making are decentralised.* (In this sense, the activities of decision-making in a firm are not different from the activities of physical production.)

The activities of information processing clearly uses resources – humans, machines, materials – and are thus costly. In US firms, more than a third, or even half, of all employees are engaged in information-processing activities, or in jobs that support such activities (Radner, 1992).

But there is an additional, more subtle, 'cost' of information processing, namely the *cost of delay*. This is not an out-of-pocket cost, like salaries, machine maintenance, etc., but a loss of profit due to the degradation of the resulting

decision. For example, if it takes a year to process the statistics on recent sales, the resulting production plan will probably be obsolete long before it is completed. Up to a point, such delays can be reduced by the use of more information processors – human and machine. This is analogous to replacing a single 'serial' computer with a 'parallel machine' made up of many serial ones. However, it is a striking 'law' of computer science that, *even with an unlimited degree of parallelisation, the computational delay must increase unboundedly with the size of the problem*, given the technological capabilities of the component processors. (It would be more accurate to characterise this statement as a *folk theorem* rather than a general mathematical result; see Schwartz (1980); and Radner (1993).)

For example, suppose that every T units of time a cohort of N numbers arrives, whose sum must be calculated. Suppose further that one has available P identical 'addition processors', and that the units of time are chosen (e.g., one microsecond) so that each processor takes one unit of time to do an addition. It can be shown that, no matter how large P is, and how the task of addition is divided up among the individual processors, *it will take at least*

$$1 + \log_2 N$$

units of time to add the N numbers. This illustrates what I shall call *the iron law of delay*. Furthermore, the number of processors needed to attain this minimum delay will be at least

$$(1 + \log_2 N)\,(N/2\,T),$$

a smaller number of processors will result in an even longer delay. This illustrates what I shall call *the iron law of size*. (Of course, to obtain this precise result, one needs a precise model of computation; see Radner (1993) and Radner and Van Zandt (1992) for details. In fact, this result holds for any associative operation, not just addition.) The delay can be further reduced only by using faster (and hence more costly) individual processors, which also has its limits, given the state of computing technology.

Information processing also typically requires that incoming data and intermediate results be stored in some memory. This adds further to the cost of information processing, and to the delays, as well, since it takes time to read data into a memory and to access the information that is stored there. Armchair empiricism suggests that machine memory is in some sense relatively cheap compared to computation, at least in the current state of technology.

Finally, the decentralisation of information processing (parallel processing) requires that the individual processors communicate with each other. Such communication requires additional resources and causes additional delays. Again, in the present state of technology, it seems that information *transmission* is relatively cheap, compared to computation in the narrow sense; witness the flood of 'junk' mail, telephone calls, and E-mail. It appears that today it is much easier and cheaper to send and receive memos and papers than it is to 'process' them.

The costs and delays attributable to information processing have an obvious but important implication for organisation: it will not be efficient (except in the

tiniest organisations) for every decision to use all of the information available to the organisation as a whole. In fact, in a large organisation, only a small fraction of the available information will be brought to bear on any single decision; the efficient choice of information will be influenced by its cost and its relevance to the decision in question. Combining this observation with the fact that individual decision-makers (processors) are limited in their capacities for information processing, one is led to the *inevitability of decentralised decision-making*, in which different decisions – or groups of decisions – are made by different decision-makers on the basis of different information.

I.4. *Deciding How to Decide: Truly Bounded Rationality*

The considerations sketched in the last two sections make the typical problem of rational decision-making under uncertainty appear much more complicated than the simple examples one encounters in textbooks on microeconomic theory and management science. Nevertheless, one can, *in principle*, stretch the Savage paradigm to accommodate the various costs, individual capacities, and delays, associated with observation, memory, computation, and communication in 'routine' decision-making. I shall call this suitably 'stretched' theory the *Extended Savage Paradigm*.

In fact, I chose to talk about 'routine' decision-making only to fix the ideas. Nothing about the theoretical framework necessarily prevents it from being applied to one-of-a-kind decisions. In particular, one example of a one-of-a-kind decision is the problem of devising decision-rules for a given set of routine decisions. In other words, it would appear that the extended Savage paradigm can be applied to the study of rational decision-making in the firm at all levels. This, in fact, was the research programme set forth in *The Economic Theory of Teams* (Marschak and Radner, 1972).

It may be instructive to inquire why progress on that research programme has been so slow in the past two decades. In the theory of teams, it seemed convenient to distinguish between (1) the organisation proper, including the members of the team, their tasks, their statistical decision rules, and the structure of observation and communication in the team, and (2) the 'organiser' or 'designer' who determines everything described in (1). It soon became evident that even some simple problems in team theory could quickly become analytically – and even computationally – intractable, *from the point of view of the organiser*. If this were true of the simplified theoretical models of the theory, how much more so would it be true of the decision problems of real firms! (For references to literature on team theory see Van Zandt (1995) and Kim and Roush (1987).)

Here we come face to face with the hard core of 'bounded rationality'. It is not that, in themselves, the costs of observation, communication, memory storage, and routine computation, necessarily prevent a team of decision makers from conforming to the Savage paradigm of rationality. Rather, it is that the task of designing decision rules that satisfy Savage's consistency requirements is beyond the intellectual capabilities of any organiser or team of organisers.

Savage understood this problem, although he did not have a satisfactory formal solution to it. He contrasted the two proverbs, 'Look before you leap', and 'You can cross that bridge when you come to it'.

> Carried to its logical extreme, the 'Look before you leap' principle demands that one envisage every conceivable policy for the government of his whole life (at least from now on) in its most minute details, in the light of the vast number of unknown states of the world, and decide here and now on one policy. This is utterly ridiculous, not – as some might think – because there might later be cause for regret, if things did not turn out as had been anticipated, but because the task implied in making such a decision is not even remotely resembled by human possibility. It is even beyond our power to plan a picnic or play a game of chess in accordance with the principle, even when the world of states and the set of available acts to be envisaged are artificially reduced to the narrowest reasonable limits. (Savage, 1954, p. 16.)

To put it another way, decision makers are not merely uncertain about 'empirical' events such as, 'It will rain tomorrow in New York City', or 'General Motors has developed an efficient electric automobile'. They are also uncertain about logical inferences, such as 'Fermat's Last Theorem follows from the axioms of arithmetic', or 'The 123rd digit in the decimal expansion of pi is 3'. Or, if these seem far removed from the business world, try: (1) 'Given all I know about the old and new drugs for treating a particular disease, what is the optimal policy for conducting clinical trials on the new ones?', or (2) 'Given all that AT & T's management knows about the telecommunications industry, and about AT & T's capabilities, what should be its business strategy for the next five years?', or especially, (3) 'Given all that is known, theoretically and empirically, about business organisations in general, and about telecommunications and AT & T in particular, should AT & T re-organise itself internally, and if so, how?'

Essentially, what we are dealing with here is the decision-maker's *uncertainty about the logical implications of what he knows*. Although I cannot justify it scientifically, I have a feeling that this kind of uncertainty is at least as important in business (and other) decision-making as uncertainty about empirical events – and perhaps even more so. In any case, I am not aware of any generally acceptable theory of what it means to be rational in the face of such uncertainty. (See, however, Lipman (1995) for a start in this direction.)

This confronts the economist with a dilemma. On the one hand, he or she can continue to investigate models of 'rational' decision-making that are simple enough to be tractable for the economist, but are hopelessly unrealistic (this is the current mainstream approach). On the other hand, at the risk of being branded a 'behavioural economist', he or she can abandon the attempt to explain observed behaviour as 'rational', and simply record various empirical regularities, e.g., 'rules of thumb'. As a compromise, the economist can try to show that competition will weed out 'irrational' or 'non-optimal' behaviour in the long run, even though individual decision-makers and

organisations are not themselves capable of deliberately determining what is optimal; in some sense, this is the best of all possible worlds. (See, e.g., Winter (1991) and the references cited there. I should mention, however, that Winter is generally critical of the *competitive selection* hypothesis, at least in its starkest forms.)

II. CONFLICT AND THE DECENTRALISATION OF AUTHORITY

II.1. *Conflicting Goals and the Power of Decentralised Information*

Thus far I have discussed decision-making in the firm as if it were simply an extension of a one-person decision problem to a group of persons and machines, this extension being made necessary by the limitations on the capacities of individuals to perform the various activities required in the process of making decisions. However – as economists and other social scientists are fond of emphasising – different persons in an organisation are likely to have conflicting goals, or at least partly so, and these conflicts are likely to impose further 'costs' on the organisation.

Of course, if the firm has a sole owner, and if the persons hired by the owner are told exactly what to do in all circumstances, and are perfectly monitored and disciplined, then the fact that they have conflicting underlying goals will be harmless (to the owner). But if the persons are hired to make decisions, and the information and decision-making is decentralised in sense of the preceding sections, then the individual decision-makers will in general have some freedom to take some actions that are more in their own interest – and less in the interest of the owner – than they would if the owner were in complete control. Thus *the decentralisation of information and decision confers power on the individual decision-makers* (subordinates?), which they may be able to use to further their own interests at the owner's expense. The situation is even more complex if there is more than one owner, and especially if some or all of the owners are also decision-makers in the firm.

II.2. *Incentives and Mechanism Design*

To study the effects of conflicting goals in the presence of decentralised information and decision-making, economic theorists – and experimentalists, too – have primarily utilised the general model of 'games of strategy'. In any particular instance, the model specifies: (1) who are the 'players' in the game, i.e., the decision-makers; (2) what strategies are available to each player; and (3) the 'payoff function', which specifies the expected utility for each player corresponding to each profile of strategies chosen by the players. Thus the 'Games Paradigm' extends the Savage paradigm by making a decision-maker's utility depend both on his actions *and* on the actions of other decision-makers, as well as on the 'state of nature' or natural environment, which is beyond any decision-maker's control. The most widely used predictive concepts are some variation on the idea of a 'Nash equilibrium', namely, a profile of strategies such that no player can increase his expected utility by *unilaterally* changing his own strategy.

In the application of game theory to the design of organisations, it is imagined that the 'organiser' (owner?) can choose the rules of the game – and hence the number of players, their strategy sets, etc. – subject only to some constraints that reflect the available technology (including the technology of information processing), the availability of employees, the outside opportunities of potential employees, etc. It is understood that the organiser may be one of the players in the game, or may represent 'society'. It is also imagined that the organiser will make this (constrained) choice to maximise his own expected utility; this is called the problem of *mechanism design*. Here the 'mechanism' refers to the particular rules of the game chosen by the organiser.

In order to predict what expected utilities for the organiser and other players will result from a particular mechanism, theorists typically predict that the players will adopt a Nash profile of strategies. According to this theory, if the organiser chooses a mechanism so as to maximise his own utility, then he will predict that for any choice of mechanism he will receive the utility corresponding to a Nash profile for that mechanism. A mechanism, with an associated Nash profile, is called *second-best* (given the technology and preferences of the players), if there is no Nash profile of any mechanism that makes every player (including the organiser) as well off, and some player strictly better off.

It is interesting that the general theory of mechanism design has thus far made little headway in illuminating the choice among *alternative* architectures of incentives and decision-making in the firm. The examination of a specific design problem for the firm may shed some light on the difficulties. I take as my example the problem of transfer pricing, and specifically, the design of bargaining mechanisms.

Divisions of large firms often 'buy' goods and services from other divisions of the same firm, and there has been much discussion of the question of how to organise such interdivisional transfers, and the corresponding 'transfer prices'. Although various formal accounting procedures have been used for this (e.g., 'cost-plus pricing'), it is also common to allow the divisions to bargain about whether the transfer should be accomplished, and at what price. In order to bring analysis to bear on the relative merits of different mechanisms, e.g., whether cost-plus is better than bargaining, we need models of how the participants would behave under these different mechanisms. I shall illustrate this with a standard game-theoretic model of bargaining in an extremely simple situation.

Suppose that there are two divisions, say S and B, and that division B (the 'buyer') is interested in obtaining a lot of 50,000 custom parts from division S (the 'supplier'). The lot will have a value of V to B, if B obtains it; that is, V is the maximum that B would pay to obtain the lot from S. (This is sometimes called B's 'reservation price', it may reflect the price at which B could obtain a comparable lot from an outside source.) Correspondingly, it will cost S an amount C to provide the lot to B. (This reservation price may reflect S's opportunities for using its productive capacities in some other way.) From the point of view of the firm, the transfer should take place if V is at least as large

as C, in which case the transfer price should divide the 'surplus' $(V-C)$ in some 'equitable' way between the two divisions. The problem is that B does not know S's cost, C, and S does not know B's value, V; furthermore, the firm's management knows neither. The firm wants to design a mechanism that will result – in situations like this – in as efficient an outcome as possible, i.e., a second-best combination of mechanism and Nash profile.

A familiar example of a mechanism is the 'sealed-bid' mechanism. The rules of the game are as follows: (1) simultaneously, the buyer picks an offer bid, say v, and the seller picks an asking bid, say c (we may think of each bid as being placed in a sealed envelope); (2) the sealed bids are opened, and if v is at least as large as c, then S provides the lot to B; otherwise no transfer takes place; (3) if the transfer takes place, the transfer price, p, is the arithmetic mean of the two bids, i.e., $p = (v+c)/2$. If a transfer takes place, the buyer's profit (on this transaction) is $(V-p)$ and the seller's profit is $(p-C)$; if no transfer takes place, then each player's profit is zero. (I assume that if no transfer takes place, then the seller incurs no cost.)

Recall, however, that B does not know C, and S does not know V. Following the theory of Harsanyi, the game-theorist will typically model the situation as follows. V and C are random variables with a given joint probability distribution, which is known to both players, B and S. At the beginning of the game, 'Nature' draws particular values, V and C, from this distribution, and reveals V to B and C to S. A *strategy* for B is a rule or function, say f, that determines his bid, v, as $f(V)$. Likewise, a strategy for S is a function, say g, that determines his bid, c, as $g(C)$. Notice that there is no obligation for B to bid his true value, nor for S to bid his true cost. In fact, armchair experience would suggest that $f(V)$ would typically be less than V, and $g(C)$ would be greater than C. Finally, we postulate that each player is interested in maximising his *expected profit*, given the other player's strategy. A *Nash profile* is a pair of bidding strategies, (f^*, g^*), such that neither player can increase his expected profit by *unilaterally* changing his own strategy to some other one.

The reader may be surprised by a feature of the above formulation, namely, that the buyer must decide what he would bid for every possible realisation of his actual value, V, even though the transfer-price story begins with the buyer knowing V (and likewise for the seller). However, a little reflection should convince the reader that, in order for the buyer to predict the consequences of any particular bid, he must have some opinion about what the seller is likely to ask. One way for the buyer to formulate this opinion is (1) to predict what the seller will ask given alternative realisations of his cost, i.e., predict the seller's strategy, and (2) to impute a probability distribution to the cost, as in the Savage paradigm. Similar reasoning is applicable to the seller.

I shall not try here to give a general definition of a 'mechanism' in this situation, but here is an alternative to the sealed-bid mechanism, sometimes called the 'take-it-or-leave-it' mechanism. In this mechanism, the seller makes an asking bid, and the buyer must either accept it or reject it; in the latter case, there is no trade. (A corresponding mechanism would be obtained by interchanging the roles of the buyer and seller.)

To continue the discussion of the sealed-bid mechanism. Suppose that the probability distribution of V and C is such that *they are independent and identically distributed, each with a uniform distribution between 0 and 1*. It can be shown that the following bidding strategies form a Nash profile:

$$(N) \qquad f^*(V) = \begin{cases} V, & \text{if } V \leqslant 1/4, \\ 1/4 + (2/3)(V - 1/4), & \text{if } V > 1/4, \end{cases}$$

$$g^*(C) = 1 - f^*(1 - C).$$

Thus the buyer bids his true value, V, when V is sufficiently small, but 'shaves' his bid for larger values of V, and the seller behaves symmetrically. Although this qualitative feature of the player's strategies may be intuitively plausible, the particular form and coefficients of the (piecewise linear) functions f^* and g^* are surely not. It can also be shown that the sealed-bid mechanism, in combination with the Nash profile (N), is in fact efficient (second-best) in the class of all mechanisms that are relevant for this problem, with the assumed probability distribution of V and C. This makes the sealed-bid mechanism seem attractive in this situation, until one learns that, under the same assumptions, *the sealed-bid mechanism has an uncountable infinity of Nash profiles, which range from efficient to worthless*! (By 'worthless' I mean yielding each player an expected profit of zero.) (For this and other facts about the sealed-bid mechanism described here, see Leininger *et al.* (1989). For a survey of game-theoretic treatments of bargaining under incomplete information, see Linhart *et al.* (1992).)

In this situation, the organiser of the firm (the mechanism designer) faces some difficult problems:

1. How could we reliably say that the sealed-bid mechanism is 'efficient', if we do not know which Nash profile the players will choose?

2. How are the players to coordinate their choice of strategies to obtain a Nash profile? One answer to this question might be that the mechanism designer should tell the players which strategies to choose, and explain why. This will work, and also solve the preceding problem, if both players actually take the designer's advice. However, this does not solve the next problem.

3. The strategies described by (1) are efficient for the particular *prior* probability distribution of V and C. However, the efficient combination of mechanism and Nash profile depends on that prior, probably in a sensitive way. (Although, even for the sealed-bid mechanism, Nash profiles are known for only a small class of distributions, and even then not all the Nash profiles are known!)

4. In the above formulation, the two players agree on what the prior distribution of V and C is. What if they do not agree?

5. How can we expect the players to calculate Nash profiles of a particular mechanism, if the mechanism designer cannot do so? In fact, the calculation of a Nash profile, even in this simple problem, typically requires the solution of a difficult pair of differential equations. It is interesting to note that, in the experiments that have been done with the two-person sealed-bid mechanism,

with the experimental conditions devised to replicate the above assumptions, the experimental subjects rarely appear to use Nash-profile strategies. However, on average they do about as well as they would if they used the efficient Nash profile (1), but with different strategies. Furthermore, subject pairs use a variety of strategy profiles, which nevertheless achieve comparable average profits (see Radner and Schotter (1989) and Linhart *et al.* (1992)).

Although I have discussed here only one mechanism in any detail, it should not be surprising that, with all of these problems, it is difficult to come to definite conclusions about the best mechanism to use for transfer-pricing. This discouraging situation raises serious questions about the ultimate prospects for the game-theoretic approach to mechanism design, at least as we now know it.

III. CONCLUSION

I have argued that costly rationality alone leads to the decentralisation of information processing, decision making, and incentives in firms. I have also argued that truly bounded rationality (uncertainty about the logical implications of what one knows), together with indeterminacy of equilibrium, make the current game-theoretic approach to the theory of the firm both implausible and inadequate. Although the game-theoretic approach has proved to be a useful abstraction, in my opinion it has also missed significant features of the organisation of firms, which can only be explained by a satisfactory theory of truly bounded rationality.

Stern School of Business, New York University

REFERENCES

Arrow, Kenneth J. (1974). *The Limits of Organization*, New York: W. W. Norton & Co.
Basu, A., Bose, A. and Gosh, J. K. (1990). 'An expository review of sequential design and allocation rules,' Technical Report 90-08, Department of Statistics, Purdue University (second revision, August, unpublished).
Berle, Adolph A. and Means, G. C. (1932). *The Modern Corporation and Private Property*. New York: Harcourt, Brace and World.
Clark, John M. (1918). 'Economics and modern psychology.' *Journal of Political Economy*, vol. 26, pp. 1–30.
Dutta, Prajit K. and Radner, Roy (1995). 'Moral hazard.' In *Handbook of Game Theory*, vol. 2 (ed. Robert A. Aumann and Sergiu Hart). Amsterdam: North-Holland, pp. 869–903.
Fama, Eugene F. and Jensen, Michael C. (1983). 'Separation of ownership and control.' *Journal of Law and Economics*, vol. 26, pp. 301–25.
Fudenberg, Drew and Tirole, Jean (1991). *Game Theory*. Cambridge, MA: MIT Press.
Hart, Oliver (1989). 'An economists's perspective on the theory of the firm.' *Columbia Law Review*, vol. 89, pp. 1757–74.
Hart, Oliver (1991). 'Incomplete contracts and the theory of the firm' in Williamson and Winter (1991), pp. 138–58.
Jensen, Michael and Meckling, William (1976). 'Theory of the firm: managerial behavior, agency costs, and ownership structure.' *Journal of Financial Economics*, vol. 3, pp. 305–60.
Kim, Ki H. and Roush, Fred W. (1987). *Team Theory*. New York: Halsted Press.
Kornhauser, Lewis A. (1983). 'Reliance, repetition, and breach of contract.' *Journal of Law and Economics*, vol. 26, pp. 691–706.
Leininger, Wolfgang, Linhart, Peter B. and Radner, Roy (1989). 'Equilibria of the sealed-bid mechanism for bargaining with incomplete information.' *Journal of Economic Theory*, vol. 48, pp. 63–106.
Linhart, Peter B., Radner, Roy and Satterthwaite, Mark A. (1992). *Bargaining with Incomplete Information*. San Diego: Academic Press.

Lipman, Barton L. (1995). 'Decision theory without logical omniscience: toward an axiomatic framework for bounded rationality.' Queen's University, Canada (unpublished).

Marris, Robert L. (1964). *The Economic Theory of 'Managerial' Capitalism*, London: Macmillan.

Marris, Robert L. and Wood, A. J. B. (eds.) (1971). *The Corporate Economy*. London: Macmillan.

Marschak, J. and Radner, R. (1972). *Economic Theory of Teams*. New Haven: Yale University Press.

Osborne, Martin and Rubinstein, Ariel (1994). *A Course in Game Theory*. Cambridge, MA: MIT Press.

Radner, Roy (1992). 'Hierarchy: the economics of managing.' *Journal of Economic Literature*, vol. 30, pp. 1382–415.

Radner, Roy (1993). 'The Organisation of decentralized information processing.' *Econometrica*, vol. 61, pp. 1109–46.

Radner, Roy (1996). 'Bounded rationality, indeterminacy, and the managerial theory of the firm.' In *Organizational Decision Making* (ed. Zur Shapira,), Cambridge: Cambridge University Press.

Radner, Roy and Schotter, Andrew R. (1989). 'The sealed-bid mechanism: an experimental study.' *Journal of Economic Theory*, vol. 48, pp. 179–220.

Radner, Roy and Van Zandt, Timothy (1992). 'Information processing in firms and returns to scale.' *Annales d'Economie et de Statitique*, No. 25/26, pp. 265–98.

Savage, Leonard J. (1954). *The Foundations of Statistics*. New York: Wiley.

Schwartz, Jacob T. (1980). 'Ultracomputers.' *ACM Transactions on Programming Languages and Computers*, vol. 2, pp. 484–521.

Smith, Adam (1776). *The Wealth of Nations*; reprinted by the University of Chicago Press, Chicago, 1976 (Cannan Edition).

Van Zandt, Timothy (1995). 'The structure and returns to scale of organizations that process information,' to appear In *Organizations with Incomplete Information*, (ed. M. K. Majumdar). Cambridge: Cambridge University Press.

Wald, Abraham (1947). *Sequential Analysis*. New York: John Wiley & Sons.

Wald, Abraham (1950). *Statistical Decision Functions*. New York: John Wiley & Sons.

Wildsmith, J. R. (1973). *Managerial Theories of the Firm*, New York: Dunellen.

Williamson, Oliver E. and Winter, Sidney G. (eds.) (1991). *The Nature of the Firm*. New York: Oxford University Press.

Winter, Sidney G. (1991). 'On Coase, competence, and the corporation.' in Williamson and Winter (1991).