人　工　知　能　106－5
（1996．10．30）

ゲーム理論における限定合理性モデル

瀧澤 弘和

東京大学大学院経済学研究科

要旨

本稿は最近再び注目を浴びつつある限定合理性アプローチが、ゲーム理論に対して持つ意義を論じる。従来、なぜ経済学者たちが限定合理性アプローチに無関心であったのかを方法論的に考察したうえで、今日のゲーム理論の状況に照らして限定合理性アプローチの復活が必然的であることが論じられる。次いで近年の限定合理性アプローチによるゲーム理論の成果を、均衡に対する演繹的アプローチと進化的アプローチ、マシーンゲーム等による均衡の縮小、その他の応用モデルの４つの分野に分けてサーベイする。前二者からは、ゲームをプレイする上で習慣や文化などの共有が重要であることが次第に明らかになってきたこと、そうした知識の共有がどのようにして生じるのかという研究が求められることが論じられる。また後二者からは、今後より一層、経済学的な文脈における応用モデルの蓄積が必要とされていることを論じる。

Bounded Rationality in Game Theory

Hirokazu Takizawa *

Graduate School of Economics
University of Tokyo

Abstract

This paper explores the significance of renewed interest in bounded rationality for the present state of game theory. The reason for the past disinterest of economists in this approach is examined. I argue that the resurgence of this approach in game theory is inevitable. I survey the state of art of this approach in 4 fields: (1) eductive approach and (2)evolutive approach to equilibrium, (3)equilibrium reduction using machine games, and (4)other economic applications. I then argue that research in these direction calls for further exploration, which can be of significance to the development of economics.

*I am grateful to Masahiro Okuno-Fujiwara and Hitoshi Matsushima for their continual encouragement and useful advice. Most of my views on bounded rationality expressed in this paper has emanated from the discussion and collaboration with Takashi Shimizu.

1 Introduction

Most people, even most economists, would agree that *true* economic agents have only limited rationality and do not act like the agents in the usual economic models, where the agents are supposed always to behave completely rationally in the face of well-defined mathematical programming problems. The last decade has seen a growing body of new literature on bounded rationality in economics, especially in the field of game theory. The intuition stated above might seem to suffice to explain this interest in bounded rationality. However, this interest is actually not new; even since the second world war we had heard the case for bounded rationality several times, but each time that it came to the fore, it faded away with no substantial progress.

This situation inevitably poses several related questions. Firstly, why has economics stuck to the rationality approach in spite of the fact that economic agents are only boundedly rational? Secondly, why has bounded rationality approach failed to revolutionize the traditional economics? Thirdly, can we expect the renewed interest in bounded rationality approach will bring about anything substantial this time?

The first two questions are answered in the next section. In answering the question, I would like to explain how economists usually think, what kind of result economists want of the bounded rationality approach, since most of the readers of this essay is supposed to be non-economists. Sections thereafter together aim to answer the third question, which is the main part of this essay. In section 3, I begin with the history of game theory. The 1980's were the turning point from fully rational approach to the boundedly rational one. I explain the driving force behind this shift, which at the same time will set some goals for bounded rationality approach in the game theory. I proceed to show some tentative results in this line of research in section 4. I divided the current research themes into 4 classes; eductive approach for explaining equilibrium, evolutive approach for explaining equilibrium, equilibrium reduction using machine games, and other applications. Finally I conclude.

2 Economics and Bounded Rationality

As I have stated in the Introduction, economists have used rationality models to explain economic phenomena, although they know that in reality agents are only limitedly rational. Milton Friedman, a Nobel laureate, gives an eloquent answer to this apparently paradoxical situation. Friedman (1953) asserts that whether assumptions of a theory are 'realistic' or not is irrelevant for the validity of the theory. The validity of a theory should be evaluated according to the predictive power of the theory, namely, relative fitness of implication deduced from the theory with the reality. Suppose we assume leaves of a tree can decide their positions freely so as to maximize the quantity of sunlight they receive. Of course this is an unrealistic assumption. Leaves cannot do such complex calculation. However, even this assumption can give reasonable explanation to wide-ranging phenomena concerning density of leaves on a tree.

Friedman goes on to argue: In economics, it is assumed that consumers maximize their utility and firms their profits. These might not be the actual case. These assumptions, however, help give the economic theory consistency and considerable explanatory power. It is entirely erroneous to reject these assumptions simply because they are unrealistic. We have to evaluate their validity by examining the validity of the theory's prediction.

In a sharp contrast, another Nobel laureate, Herbert Simon argues for the importance of reality of assumptions. According to Simon (1963) unrealistic assumptions are only "a necessary evil - a concession to the finite capacity of the scientist that is made tolerable by the principle of continuity of approximation."

I do not intend to delve into this controversy here. I would just like to comment that most economists today can be mostly said to belong to the Friedman's camp in their methodological position. Reality for the sake of reality is considered unnecessary. Although Simon received the Nobel prize, the bounded rationality approach he proposed did not revolutionize economics. The reason is twofold. On the one hand, economists have been largely satisfied with the traditional approach. On the other hand, researches in bounded rationality have not so far provided us with anything of great significance to economics.

Then what can be the motivation for studying bounded rationality in economics today? It mustn't be simply that bounded rationality models are more realistic description of human behavior, but that modelling boundedly rational human behavior provides a more plausible explanation to phenomena that could not be fully explained by the traditional rationality approach. As far as I know, at least in game theory, the *impasse* of rationality approach has been widely recognized, and there is room for the bounded rationality approach to fill this gap. To see this situation concretely, I would like to look back the history of game theory.

3 Limits of Rationality Approach in Game Theory

Since the publication of von Neuman and Morgenstern (1944), game theorists had taken it for granted that game theory deal with strategic interaction between fully rational players. In this traditional approach, players are assumed to have as much intelligence as an outside analyst, namely, a game theorist. Nash, Harsanyi and Selten received the Nobel prize in 1994 due to their theoretical contributions that made game theory more applicable to economics. Their works were largely in the traditional rationality approach.

In 1980's, as game theory became more applicable, many economists learned game theory and applied it to many fields of economics, such as industrial organization, bargaining, contracts and organizations, which had been kept relatively unexplored because of the analytical difficulties involving strategic interaction. Nash equilibrium seemed to provide us with the unified framework in which to deal with these subjects. Most of these application adopted rather naively the by then already developed solution concepts like Nash, subgame-perfect, perfect, and/or sequential equilibria.[1]

However, serious critiques began to emerge in the middle of 80's. On the one hand, experimental game theory began to report that actual behaviors observed in laboratories significantly differ from those predicted by these equilibrium concepts. Examples include finitely repeated prisoners' dilemma, ultimatum game, centipede game and so on.[2] On the other hand, theorists gradually became aware of their theories' deficiency even from introspective view point. See Rubinstein (1991) for a notable example.

I assort the problems posed by these theorists from my own point of view as follows.

1. *Why and how do we play equilibrium strategies.* This problem is important because we have multiple equilibria in many games. Without knowing how players reach some equilibrium, we can't hope to answer the question raised by multiplicity.

2. *When is the equilibrium analysis appropriate? Why do experimental results significantly differ from the prediction of equilibrium analysis?* While we cannot immediately do away with equilibrium analysis, some expeirmental results suggest its inadequacy in some important cases.

3. *Which equilibrium is most likely to be played, when there are multiple equilibria?* Only a slight complication of a game is know to be enough to bring about multiple equilibria. Examples include signaling games and repeated games.

4. *How do we settle for a specific form of protocols?* Game theory hitherto has concentrated on the analysis of well-defined games. However, as in the bargaining process, very important classes of strategic interaction involve some ambiguity of protocols. Moreover, as economists formulate and analyse institutional arrangements as games, the question of how we forge some institution has come to the fore.

[1]The notion of Nash equilibrium is due to Nash (1951). the concept of subgame-perfect equilibrium and perfect equilibrium are due to Selten (1965, 1975). The sequential equilibrium was developed by Kreps and Wilson (1982b).

[2]See for example Roth, Prasnikar, Okuno-Fujiwara, and Zamir (1991).

	C	D
C	2,2	0,0
D	0,0	1,1

Table 1: Pure Coordination Game

In the remainder of this paper, I will only deal with the first three questions, because almost nothing has been tried out as to the fourth question. At first, game theorists tried to answer these three questions by refining equilibria with the hope that imposing further rationality requirement would reduce multiple equilibria to the unique one. In the end of 1980's, however, there was disillusionment with this research program and attention gradually went to boundedly rational approach. It should be noted that some game theorists had obtained somewhat favorable, albeit not decisive, results using boundedly rational models already, such as Neyman (1985).

4 Game-Theoretic Models of Bounded Rationality

4.1 Eductive Approach

At first we are mostly concerned with the first question, namely, "how do we play equilibrium?" It is already well known that there are two approaches to this problem; ***deductive*** and ***evolutive***, to use Binmore's terminology (Binmore 1987/88).

Suppose player A and B play a pure coordination game as in figure 1, once and for all. There are two pure Nash equilibria and one mixed equilibrium in this game. Think of the equilibrium (C, C). Its self-enforcing nature does not tell us why players choose (C, C) and not (D, D). Suppose player A does not know player B intends to play C. Then she cannot choose C with confidence. It has for long been argued informally that equilibrium strategies are common knowledge when players play them. The need for rigorous formulation and analysis is obvious.

Moreover when we discuss equilibrium play of a game, we have made it a rule to assume that structure of the game (payoff, information set, strategy space etc.) is common knowledge, and that the rationality of the players is also common knowledge. The condition for an event to be a common knowledge is, however, too strong and in many actual cases hard to be satisfied.[3] Then we should ask whether we can relax the condition about the players' knowledge while guaranteeing that they play equilibrium strategies. This problem became especially important because Rubinstein (1988) gave an ingenious example in which players' action under "almost" common knowledge is completely different from the play under common knowledge.

Bernheim (1984) and Pearce (1984) gave the first analysis to this problem. They showed that if we only require the game and rationality of the players to be common knowledge, the play of the game must be 'rationalizable strategies.' In such a special case as the two-player Cournot competition, rationalizability requirement gives a sharp prediction, but in general this requirement is too weak to say something meaningful about the play of a game.

Alternative approach to this problem was made by Aumann (1987). Aumann constructed a model where each state of the world already specifies actions for each player. This is contrary to the common practice game theorists have made that strategies are decided endogenously. He assumed common prior over the set of states of the world, however. Moreover, the strategy of each players is assumed to be measurable with respect to the information partition of the player. In this setting, Aumann showed that if it is common knowledge that players are Bayesian rational, then the play of the game forms a correlated equilibrum. While this model was the first attempt to explain equilibrium play by using decision-theoretic framework, the assumption of common prior and Bayesian rationality of each player at each state of the world already embraced the factors generating equilibrium.

Aumann and Brandenburger (1995) extended this model by admitting states of the world where players can hold "erroneous" conjectures and play "irrationally". Then, by examining properties of a state where players choose Nash equilibrium strategies, they could give epistemic contidions for the equilibrium play of a game. The sufficient conditions they derived was weaker than those we have commonly made in arguing equilibrium, but only by a slight margin. Their result virtually suggests players play Nash equilibrium because they are in an equilibrium-like situation. What seems important to me is that they could not dispense with common knowledge of conjectures. If we need common knowledge of conjectures to play some equilibrium, how do we reach it? They didn't give any answers to this question. We will see that in most models pursuing "how do we play equilibrium?" nearly the same problem arises.

Binmore (1987/88) raised a related question whether there exists any algorithm for playing games that guarantees equilibrium play in the face of any opponent and any game. Behind the eductive approach lies the idea that players simulate their opponents' thought process to conclude how to play a game. Binmore's question is not about knowledge (information) of players, but about their reasoning process. Suppose that Turing machines play a game, where each Turing machine is fed as inputs with its own Gödel number, opposing machine's Gödel number and the structure of the game (appropriately coded into a natural number). This situation is deliberately constructed to eliminate problems arising from knowledge. If this question were answered affirmatively, this result would have a prescriptive meaning, namely, that we should use such an algorithm to play games.

In this setup, Anderlini (1990) considered the two-person game with at least one pure strategy Nash equilibrium and without dominant strategies. [4] He showed that there exists no algorithm that always halts and predicts the other player's play correctly, a fortiori, plays the best response. This is caused by the existence of 'diagonalizing algorithm.' This result is somehow related to the 'halting problem' that is well known in the recursive function theory. Restricting the algorithms we consider to the set of 'rational machines' does not help. Rational machines are defined such that if they both halt, they necessarily play a Nash equilibrium. For any rational machine, however, there exists a rational machine that prevents the former from halting. Anderlini attributes this result to the fact that it is not decidable whether a given machine is rational or not. The trick is that, contrary to the apparent situ-

[3] See Aumann (1976) for a formal definition. But Aumann's definition does not capture the common knowledge of the information partition of players. Modelling knowledge formally always involves this kind of problem.

[4] See also Canning (1992).

ation of common knowledge of 'thinking process' in this formulation, it cannot actually be known that both players are rational. If the rational machine is given by some outside observer the information whether the opponent is rational, they can always play Nash equilibrium.

I think this negative result suggests that "procedural rationality" on the part of players does not imply equilibrium. Some *a priori* "common knowledge", like convention or culture, between players seems to be required to attain equilibrium play. This is a similar result to that obtained in Aumann and Brandenburger (1995).

I now turn to the alternative approach with the question in mind whether learning or evolutionary process alleviates this problem.

4.2 Evolutive Approach

Generally speaking, learning approach attempts to argue that if some plausible learning rules are imposed, players will eventually learn to play equilibrium strategies. At the same time, obtained result might select some particular equilibrium.

Although evolutionary models are similar to learning models, the former differ from the latter in that explicit learning process is woven into some given dynamics in the former. Evolutionary models examine the implication of the stability of such dynamics. From the modelling point of view, while learning models consider a fixed match, evolutionary models consider the interaction in a population of players.

Learning approach dates back at least to Brown's fictitious play (Brown 1951), although it was regarded not as a learning process but as a thinking process of computing Nash equilibrium; thus called "fictitous play." It is already well known that while fictitous play generally leads to Nash equilibrium (viewed as the product of empirical marginal distributions) , some cycle can arise in the course of play. Moreover such cycles might bring the players payoffs that are substantially less than those obtained in the equilibrium play of the game. This demonstrates the main difficulty of studying learning models. Since the player in the model is assumed not to have the right model generating such cycle, she must settle for less in a sense. However, since we expect that real agents will soon realize the regularity of the path of play and respond correctly, something is thought to be missing in this story. A series of recent research, such as Fudenberg and Kreps (1993) and Fudenberg and Levine (1995), has alleviated the problem with the classical fictitious play by introducing some perturbation into the learning process, with the implication that experimentation is important to the learning process. This is parallel to the result obtained in the evolutionary models that mutations have a very important role in selecting a reasonable equilibrium as in Kandori, Mailath, and Rob (1993).

While most of the learning (evolutionary) literature has concentrated on the analysis of combination of naive assessment rule and myopic players, it seems to me that more attention should be paid to more sophisticated learning rules. When players are patient and they condition their play on histories of play, they may try to manipulate the opponents' learning process. Then we are essentially back to the strategic interaction between players.

Kalai and Lehrer (1993) gives sufficient conditions for the players playing a repeated game to predict the future path of play in the Bayesian framework. When playing the repeated game, players do not know the others' actual strategies *a priori* and they update their beliefs on the *plausible* set of opponent's strategies as the game unfolds. Essentially the condition they gave is the absolute continuity condition first discovered by Blackwell and Dubins (1962). Kalai and Lehrer then showed that if (1) players learn to predict the future path of play *and* (2) they play optimally with regard to their beliefs, then their strategies eventually converge to an ϵ-Nash equilibrium. This result looks good at first sight. However, even apart from the usual critiques against Bayesian learning, the absolute continuity condition has turned out to be problematic in the learning framework.

Nachbar (1996) showed that the above conditions (1) and (2) are hard to be satisfied at the same time. The intuition is simple. When players act optimally as Bayesian learners, their strategy might be easily forced out of the *plausible* set. Fudenberg and Levine (1996) interpret this result as showing the weakness of Bayesian approach. However, I think the same problem arises naturally in general learning rules. I conjecture that the logic behind this result parallels that obtained in Anderlini (1990) and Canning (1992). I think that by formulating a general learning rule as a Turing machine, we can extract the essence of the problem. My intuition is that even if we consider more sophisticated learning rules, this does not generally help.

On the other hand, Anderlini and Sabourian (1995a) obtained a positive result by considering the "evolution of algorithmic learning rules." The main difference between the negative and the positive results is, I think, that while the interaction of learning rules between players does not escape from the mutual "diagonalization," the existence of some objective "rule" guiding players, such as the evolutionary dynamics that determines the relative fitness, might mitigate the strategic interaction considerably. Shimizu and Takizawa (1996) explores this issue in more detail.

4.3 Equilibrium Reduction

We now turn to the third question. As the experimental game theory develops, we have come to realize that actual play can be significantly different from the prediction of theories. Moreover, in many actual cases, rational approach alone cannot give a sharp prediction as to which equilibrium is most likely to be played. Can imposing some limited computability assumption explain these phenomena? In contrast to the two approaches we have seen, we want some concrete and positive result here.

4.3.1 Irrationality

Before proceeding to models of bounded rationality, I would like to give several examples of "irrationality models" that proved to be usuful for narrowing down the set of equilibria. The so-callled "equilibrium refinement" arguments are usually supposed to be "super-rationality" approach, since they demand rationality off the equlibrium path, while Nash equilibrium demands rationality only on the equilibrium path. However it should be noted that they are all derived from introducing some "trembles," which is obviously some departure from rationality. As Aumann (1986) argues, this is not surprising, because you have to be "super-rational" to cope with irrationality of the opponents. "Thus, a more refined concept of rationality cannot feed on itself only; it can only be defined in the context of irrationality."

Also related to the equilibrium refinement literature are the reputation models, pioneered by Kreps and Wilson (1982a) and Milgrom and Roberts (1982). Reputation models consider consequences of long-run interaction between players, assuming some of the players (usually one long-run player) may be of "irrational" or "crazy" type with small probability. The "irrational" type is defined not behave optimally but be committed to some fixed strategy. Then the rational type can exploit this uncertainty to attain equilibria that are more favorable to her. This leads to significant equilibrium reduction in repeated games. For example, Kreps, Milgrom, Roberts, and Wilson (1982) showed that, in finitely repeated prisoners' dilemma, if there is small probability that one of the players is committed to "Tit-for-Tat" strategy, players cooperate until almost the end of the game in any sequential equilibrium. By restricting the set of perturbation strategies to automata with bounded recall, Aumann and Sorin (1989) showed that Pareto-efficient payoff vector is the unique equilibrium outcome in the repeated common interest game.

4.3.2 Automaton Game

One way to consider the implication of bounded rationality in games is to explore the consequence of taking account of the complexity cost of strategies players use in games. The most familiar way to do this has been to represent a strategy in the repeated game by a finite automaton, more precisely a Moore machine. Kalai and Stanford (1988) study the general problem of complexity of strategies in repeated games.

Repeated game models are know to be plagued by the multiplicity equilibria result called the "Folk Theorem." Rubinstein (1986) and Abreu and Rubinstein (1988) succeeded in reducing the set of equilibria of two-player repeated games by restricting the players' strategies to those implementable by finite automata. Actually this is a one-shot machine game where players choose a finite automaton to implement their repeated game strategy. Players care lexicographically about how much they obtain in the repeated game and about the complexity cost of the automaton, which is defined as the cardinality of the automaton's set of internal states.

Binmore and Samuelson (1992) also considered the automaton game with complexity cost consideration, but, unlike Abreu and Rubinstein, explored the appropriately modified equilibrium notion of ESS (Evolutionarily Stable Strategies). They showed that, in the repeated prisoners' dilemma game, cooperation is the unique equilibrium outcome.

4.3.3 Turing Machine Game

Automaton models are useful for considering complexity cost because its measure of complexity is very simple. But, this complexity measure is not without question. At the moment, economists have not agreed on *the* economically meaningful measure of complexity and do not like to impose some *ad hoc* bound on complexity. Thus in some cases, it may be more appropriate to utilize Turing machine, which by nits nature is not limited in the memory it uses and whose operation is supposed to cover anything that can be done with algorithms due to *Church's Thesis.*

We have already seen the models developed by Binmore, Anderlini and Canning. These models capture the idea that any human thinking process can be expressed by some algorithm, namely a Turing machine. The nagative nature of their results at least partially justifies the use of Turing machines, since their results states that *any* sophisticated thinking process cannot solve the problem.

On the other hand, Anderlini and Sabourian (1995a, 1995b) and Anderlini (1995) utilize Turing machines to derive more positive results. Anderlini and Sabourian (1995a) studies the infinitely repeated two-person common interest game with incomplete information, when players are restricted to using computable strategies, showing that the Pareto-efficient payoff vector is the unique equilibrium outcome. The logic behind this result is essentially the same as that of Aumann and Sorin (1989). What differentiates between the two is that, in Anderlini and Sabourian (1995a), the Turing machine communicates its intention to play cooperatively by appropriately choosing its actions in the early stages of the repeated game, forcing the opponent to Bayesian update his prior belief. The setting of repeated game is not essential for this logic to work. Anderlini (1995) thus obtains the same result by studying one-shot common interest game played by Turing machines where a finite sequence of pre-play communication is allowed.

We have already mentioned Anderlini and Sabourian (1995b) in the context of learning models. They think that a learning rule is representable by a Turing machine and study the consequence of (computable) evolutionary dynamics where learning rules interact one another. Due to the second recursion theorem, given some computable evolutionary dynamics, there exists an algorithm which completely simulates and calculates the consequence of dynamics and behave as the top performer at each stage.[5] Then the s-m-n theorem assures the existence of the algorithmic learning rule (the "smart type") that behaves completely in the same way as the above algorithm, thus continues to be the top performer at each stage. If the evolutionary dynamics are such that the top performer does not go extinct and the smart type is born at some stage, then the smart type will eventually prevail. Regarding a learning rule as an algorithm is very appealing. However, the main defect of this approach is that, in spite of the positive result, their model does not provide us with any insight on what kind of learning rules performs relatively well.

4.3.4 Perceptron Game

While there are many bounded rationality models using automaton and Turing machine, as far as I know, there are few models utilizing perceptron as the model of boundedly rational agent. Cho (1994) studies the infinitely repeated prisoners' dilemma game played by perceptrons. However, the obtained result is the traditional folk theorem and fails to refine the set of equilibrium payoffs.

Another model using perceptron is Rubinstein (1993). We mention this model later as an economic application of bounded rationality models.

4.3.5 Interpretation of the Results

There can be several different interpretations for the machine game models. Firstly, one may argue that a machine represents the cognitive process of human beings. But this view is too naive. If we adopt this interpretation, we have to think about how we choose an optimal machine and then we have to think about how to choose how to choose...and so on and so forth. Note that as the tier becomes deeper, the problem is getting more complicated and difficult. Lipman (1991) correctly pointed out that in any bounded rationality models we have untimately to assume a substantively rational player choosing the optimal boundedly rational procedure. He 'solved' this problem by finding conditions under which a fixed point exists in this infinite regress problem. But I think in actual cases this kind of infinite regress problem does not bother us.

Second interpretation is that these models represent situations in which a principal with substantive rationality delegate her job to an agent with limited ability to compute. Imagine a routine job in the firms. Rubinstein (1986) and Abreu and Rubinstein (1988) fit well with this interpretation.

Third interpretation, which I prefer to the others, is that the computable behavior rules in these models are just a metaphor of human behavior. In our everyday life, we seems to use some simple behavior rules. Then where do such behavior rules come from? It may have come from some evolutionary process. This interpretation fits relatively well with Binmore and Samuelson (1992).

Fourthly, one may be interested in these models just because they represent an idea that we may choose to commit ourselves to some fixed simple strategy in our daily life. Actually, in machine games, players commit themselves to a fixed strategy and analyst examines the equilibrium of this commitment game. [6]

[5] Note that the consequence of the dynamics is affected by the existence of such an algorithm. Hence recursion theorem.

[6] In this regard, the idea of Matsui (1989) is interesting. He examines the consequence of allowing the player to change her once committed repeated game strategy with exogenously given small probability.

4.4 Other Applications

I have emphasized the importance of constructing models that help us understand phenomena that could not be explained by rationality approach. However, the models we have so far seen are rather abstract and far from explaining concrete economic phenomena.

Today we have only few models of bounded rationality applied to concrete economic phenomena, but such models are now emerging. The growing literature on incomplete contracts could also be counted as application of bounded rationality. The models of incomplete contracts study the consequence of human inability to *write* contracts covering all the conceivable contingencies, but do not impose any restriction to the agents' ability to expect those contingencies and compute rational options. For this reason, I do not explain this literature in details.

Rubinstein (1993) studies the consequence of different (limited) ability to process information between players. Specifically he considered a situation where a monopolist with full rationality faces many consumers with bounded rationality. Basically the setting is that of Stackelberg's leader-follower model. The monopolist set a price and the consumers decide whether to buy or not a unit of a perishable good. The consumers' ability to process information is assumed to be limited such that the number of sets in the partition of the price space is finite. Different types of consumers are assumed to have different number of partition sets. Rubinstein showed that in equilibrium the monopolist can exploit this difference in the information processing ability among consumers to increase her profit. He also examined the model where the difference of consumers' ability to process information is expressed by the order of perceptrons they employ, obtaining essentially the same result.

Fershtman and Kalai (1993) study the behavior of a firm operating in many markets at the same time. They used the concept of strategic complexity as measured by a finite automaton and studied the consequence of bounded rationality of the firm's manager. They showed that an entry to one market may induce the firm to exit from another market to concentrate on the market where entry occurred. For different parameters, the firm may exit from this market to specialize in the other markets. In both cases the results obtained there could not have been obtained in the full rationality approach.

Anderlini and Felli (1994) considered the bounded rationality foundation of incomplete contracts. They formulated a contract as a mapping from the state of the world to the set of prescribed outcomes. Specifically they considered a sharing rule between two agents who have state-dependent utility functions. Thus the optimal sharing rule is in general sensitive to changes in the state of the world. However, in reality we usually write incomplete contracts. They explained this phenomena by requiring the written contract to be computable in order to be enforceable. However the optimal contract can be approximated arbitrarily closely by a sequence of computable contracts. They then introduce a Turing machine which selects the "better" computable contracts to show that the selected contract must be necessarily incomplete (even considering convergence) in some situation.

Radner (1993) studies a model where limited ability to process information is represented explicitly by time. The models we have seen so far almost entirely ignored this issue. Complexity cost measures such as computation in polynominal time or something like that measure complexity in terms of time. But I don't think there is a wide consensus to this approach at least in the economic professions. Radner's approach is very intuitive, however. Radner restricted an agent's ability such that she can do (associateve) operation only once in a unit time period in the face of numbers that flow in continually. An organization is formed in order to process the inflow of numbers, whose goal is to finish calculation as fast as possible. But since employing agents is costly, the organization seeks to accomplish the goal by employing as less agents as possible. Radner examined the optimal organizational structure, to find that hierarchy is "nearly" optimal.

5 Conclusion

About a decade has passed since game theorists began to explore bounded rationality models intensively and extensively. How has it fared with this approach? What do the results obtained so far indicate? Has this approach made a genuine progress in understanding human interaction?

My conclusion must be rather tentative. Before restating the gained insight, however, I would like to emphasize the possibility of paradigm change in economics. In the early part of this essay, I said that bounded rationality approach will not be widely accepted unless it brings about something new to economics. This will continue to be true in the future. However, it should also be noted that a prevalent paradigm always restricts our attentions to problems that are solvable within this paradigm. In this regard, it seems to me, economists are now increasingly becoming aware of the fact that there are many economic phenomena that cannot be explained by the traditional rationality approach.

This might be a symptom for a genuine paradigm change.

I assort what I have seen in this essay in two parts. Results obtained in the research program for explaining equilibrium play of games, whether eductive or evolutive, mainly suggest that there should be something in common *a priori* for us to play some equilibrium. This in turn suggests, it seems to me, the importance of conventions, culture, focal points and whatever may help us coordinate our beliefs. The question of how we come to such a state of consensus calls for further research.[7]

The results obtained in the analysis of machine games show us that complexity cost consideration about the strategies players use might provide a sharper prediction about the play of games. Unfortunately, at the moment, there exists no consensus as to what constitutes the economically significant measure of complexity cost. Accumulating good models that are widely applicable in economic context would promise the future of this approach.

References

ABREU, D., AND A. RUBINSTEIN (1988): "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 56, 1259–1281.

ANDERLINI, L. (1990): "Some Notes on Church's Thesis and the Theory of Games," *Theory and Decision*, 29, 19–52.

——— (1995): "Communication, Computability, and Common Interest Games," working paper.

ANDERLINI, L., AND L. FELLI (1994): "Incomplete Written Contracts: Undescribable States of Nature," *Quarterly Journal of Economics*, 109, 1085–1124.

ANDERLINI, L., AND H. SABOURIAN (1995a): "Cooperation and Effective Computability," *Econometrica*, 63, 1337–1369.

——— (1995b): "The Evolution of Algorithmic Learning Rules: A Global Stability Result," mimeo.

AUMANN, R. (1976): "Agreeing to Disagree," *Annals of Statistics*, 4, 1236–1239.

——— (1986): "Rationality and Bounded Rationality," Nancy L. Schwartz Lecture.

[7]As far as I know, there are two contrasted approaches to explaining conventions. One is evolutionay approach by Young (1993) and the other is Shin's rationality approach (Shin 1996).

——— (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

AUMANN, R., AND A. BRANDENBURGER (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1161–1180.

AUMANN, R. J., AND S. SORIN (1989): "Cooperation and Bounded Recall," *Games and Economic Behavior*, 1, 5–39.

BERNHEIM, D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007–1028.

BINMORE, K. (1987/88): "Modeling Rational Players, Part I, II," *Economics and Philosophy*, 3,4, 179–214,9–55.

BINMORE, K., AND L. SAMUELSON (1992): "Evolutionary Stability in Games Played by Finite Automata," *Journal of Economic Theory*, 57, 278–305.

BLACKWELL, D., AND L. DUBINS (1962): "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, 38, 882–886.

BROWN, G. (1951): "Iterative Solution of Games by Fictitious Play," in *Activity Analysis of Production and Allocation*, ed. by T. Koopmans. New York: Wiley, pp. 374–376.

CANNING, D. (1992): "Rationality, Computability, and Nash Equilibrium," *Econometrica*, 60, 877–888.

CHO, I. (1994): "Perceptrons Play the Repeated Prisoner's Dilemma," mimeo.

FERSHTMAN, C., AND E. KALAI (1993): "Complexity Considerations and Market Behavior," *Rand Journal of Economics*, 24, 224–235.

FRIEDMAN, M. (1953): "The Methodology of Positive Economics," in *Essays in Positive Economics*. Chicago: University of Chicago Press.

FUDENBERG, D., AND D. KREPS (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5, 320–367.

FUDENBERG, D., AND D. LEVINE (1995): "Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*.

——— (1996): *Theory of Learning in Games*.

KALAI, E., AND E. LEHRER (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61, 1019–1045.

KALAI, E., AND W. STANFORD (1988): "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, 56, 397–410.

KANDORI, M., G. MAILATH, AND R. ROB (1993): "Learning, Mutation and Long Run Equilibria in Games," *Econometrica*, 61, 29–56.

KREPS, D., P. MILGROM, J. ROBERTS, AND R. WILSON (1982): "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27, 245–252.

KREPS, D., AND R. WILSON (1982a): "Reputation and Inperfect Information," *Journal of Economic Theory*, 27, 253–279.

——— (1982b): "Sequential Equilibrium," *Econometrica*, 50, 863–894.

LIPMAN, B. (1991): "How to Decide How to Decide How to ...: Modeling Limited Rationality," *Econometrica*, 59, 1105–1125.

MATSUI, A. (1989): "Information Leakage Forces Cooperation," *Games and Economic Behavior*, 1, 94–115.

MILGROM, P., AND J. ROBERTS (1982): "Predation, Reputation and Entry Deterrence," *Journal of Economic Theory*, 27, 280–312.

NACHBAR, J. (1996): "Prediction, Optimization, and Learning in Repeated Games," mimeo.

NASH, J. (1951): "Non-Cooperative Games," *Annals of Mathematics*, 54, 286–295.

NEYMAN, A. (1985): "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma," *Economics Letters*, 19, 227–229.

PEARCE, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.

RADNER, R. (1993): "The Organization of Decentralized Information Processing," *Econometrica*, 61, 1109–1146.

ROTH, A., V. PRASNIKAR, M. OKUNO-FUJIWARA, AND S. ZAMIR (1991): "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburg, and Tokyo: An Experimental Study," *American Economic Review*, 81, 1068–1095.

RUBINSTEIN, A. (1986): "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory*, 39, 83–96.

——— (1988): "The Electric Mail Game: Strategic Behavior under 'almost common knowledge.'," *American Economic Review*, 79, 385–391.

——— (1991): "Comments on the Interpretation of Game Theory," *Econometrica*, 59, 909–924.

——— (1993): "On Price Recognition and Computational Complexity in a Monopolistic Model," *Journal of Political Economy*, 101, 473–484.

SELTEN, R. (1965): "Spieltheoretische Behandlung eines Oligopolmodels mit Nachfrageaegheit," *Zeitschrift fur die gesamte Staatswissenschaft*, 12, 301–324.

——— (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

SHIMIZU, T., AND H. TAKIZAWA (1996): "Strategic Interaction between Algorithmic Learning Rules," mimeo.

SHIN, H. (1996): "How Much Common Belief is Necessary for a Convention?," *Games and Economic Behavior*, 13, 252–268.

SIMON, H. (1963): "Discussion of Problems of Methodology," *American Economic Review*, 53, 229–231.

VON NEUMAN, J., AND O. MORGENSTERN (1944): *Theory of Games and Economic Behavior*. New York: John Wiley and Sons.

YOUNG, P. (1993): "The Evolution of Conventions," *Econometrica*, 61, 57–84.