Rationality and Bounded Rationality*

Robert J. Aumann

The Hebrew University of Jerusalem

INTRODUCTION

Economists have long expressed dissatisfaction with the complex models of strict rationality that are so pervasive in economic theory. There are several objections to such models. First, casual empiricism or even just simple introspection leads to the conclusion that even in quite simple decision problems, most economic agents are not in fact maximizers, in the sense that they do not scan the choice set and consciously pick a maximal element from it. Second, such maximizations are often quite difficult, and even if they wanted to, most people (including economists and even computer scientists) would be unable to carry them out in practice. Third, polls and laboratory experiments indicate that people often fail to conform to some of the basic assumptions of rational decision theory. Fourth, laboratory experiments indicate that the conclusions of rational analysis (as distinguished from the assumptions) sometimes fail to conform to "reality." And finally, the conclusions of rational analysis sometimes seem unreasonable even on the basis of simple introspection.

From my point of view, the last two of the above objections are more compelling than the first three. In science, it is more important that the conclusions be right than that the assumptions sound reasonable. The assumption of a gravitational force seems totally unreasonable on the face of it, yet leads to correct conclusions. "By their fruits ye shall know them" (Matthew 7, 16).

In the following, though, we shall not hew strictly to this line; we shall examine various models that, between them, address all the above issues.

To my knowledge, this area was first extensively investigated by Herbert Simon (1955, 1972). Much of Simon's work was conceptual rather than formal. For many years after this initial work, it was recognized that the area was of great importance, but the lack of a formal approach impeded its progress. Particular components of Simon's ideas, such as satisficing, were formalized by several workers, but never led to an extensive theory, and indeed did not appear to have significant implications that went beyond the formulations themselves.

There is no unified theory of bounded rationality, and probably never will be. Here we examine several different but related approaches to the problem. We will not survey the area, but discuss some of the underlying ideas. For clarity, we may sometimes stake out a position in a fashion that is more one-sided and extreme than we really feel; we have the highest respect and admiration for all the scientists whose work we cite, and beg them not to take offense.

From the point of view of the volume of research, the field has "took off" in the eighties. An important factor in making this possible was the development of computer science, complexity theory, and so on, areas of inquiry that created an intellectual climate conducive to the development of the theory of bounded rationality. A significant catalyst was the experimental work of Robert Axelrod (1984) in the late seventies and early eighties, in which experts were asked to prepare computer programs for playing the repeated prisoner's dilemma. The idea of a computer program for playing repeated games presaged some of the central ideas of the later work; and the winner of Axelrod's tournament—*tit for tat*—was, because of its simplicity, nicely illustrative of the bounded rationality idea. Also, repeated games became the context of much of the subsequent work.

The remainder of this lecture is divided into five parts. First we discuss the evolutionary approach to optimization—and specifically to game theory—and some of its implications for the idea of bounded rationality, such as the development of truly dynamic theories of games, and the idea of "rule rationality" (as opposed to "act rationality"). Next comes the area of "trembles," including equilibrium refinements, "crazy" perturbations, fail-

ure of common knowledge of rationality, the limiting average payoff in infinitely repeated games as an expression of bounded rationality, ε-equilibria, and related topics. Part 3 deals with players who are modeled as computers (finite state automata, Turing machines), which has now become perhaps the most active area in the field. In Part 4 we discuss the work on the foundations of decision theory that deals with various paradoxes (such as Allais, 1953, and Ellsberg, 1961), and with results of laboratory experiments, by relaxing various of the postulates and so coming up with a weaker theory. Part 5 is devoted to an open problem.

Most of this lecture is set in the framework of noncooperative game theory, because most of the work has been in that framework. Game theory is indeed particularly appropriate for discussing fundamental ideas in this area, because it is relatively free from special institutional features. The basic ideas are probably applicable to economic contexts that are not game-theoretic (if there are any).

1. EVOLUTION

a. Nash Equilibria as Population Equilibria

One of the simplest, yet most fundamental ideas in bounded rationality —indeed, in game theory as a whole—is that no rationality at all is required to arrive at a Nash equilibrium; insects and even flowers can and do arrive at Nash equilibria, perhaps more reliably than human beings. The Nash equilibria of a strategic (normal) form game correspond precisely to population equilibria of populations that interact in accordance with the rules—and payoffs—of the game.

A version of this idea—the evolutionarily stable strategy—was first developed by John Maynard Smith (1982) in the early seventies, and applied by him to many biological contexts (most of them animal conflicts within a species). But the idea applies also to Nash equilibria—not only to interaction within a species, but also to interactions between different species. It is worthwhile to give a more precise statement of this correspondence.

Consider, then, two populations—let us first think of them as different species—whose members interact in some way. It might be predator and prey, or cleaner and host fish, or bees and flowers, or whatever. Each interaction between an individual of population A and one of population B results in an increment (or decrement) in the fitness of each; recall that

the fitness of an individual is defined as the expected number of its offspring (I use "its" on purpose, since, strictly speaking, reproduction must be asexual for this to work). This increment is the payoff to each of the individuals for the encounter in question. The payoff is determined by the genetic endowment of each of the interacting individuals (more or less aggressive or watchful or keen-sighted or cooperative, etc.). Thus one may write a bimatrix in which the rows and columns represent the various possible genetic endowments of the two respective species (or rather those different genetic endowments that are relevant to the kind of interaction being examined), and the entries represent the single encounter payoffs that we just described. If one views this bimatrix as a game, then the Nash equilibria of this game correspond precisely to population equilibria; that is, under asexual reproduction, the proportions of the various genetic endowments within each population remain constant from generation to generation if and only if these proportions constitute a Nash equilibrium.

This is subject to the following qualification: in each generation, there must be at least a very small proportion of each kind of genetic endowment; that is, each row and column must be represented by at least some individuals. This minimal presence, whose biological interpretation is that it represents possible mutations, is to be thought of as infinitesimal; specifically, an encounter between two such mutants (in the two populations) is considered impossible.

A similar story can be told for games with more than two players, and for evolutionary processes other than biological ones; e.g., economic evolution, like the development of the QWERTY typewriter keyboard, studied by the economic historian Paul David (1986). It also applies to learning processes that are perhaps not strictly analogous to asexual reproduction. And though it does not apply to sexual reproduction, still one may hope that, roughly speaking, similar ideas may apply.

One may ask, who are the "players" in this "game?" The answer is that the two "players" are the two populations (i.e., the two species). The individuals are definitely *not* the "players"; if anything, each individual corresponds to the pure strategy representing its genetic endowment (note that there is no sense in which an individual can "choose" its own genetic endowment). More accurately, though, the pure strategies represent kinds of genetic endowment, and not individuals. Individuals indeed play no explicit role in the mathematical model; they are swallowed up in the proportions of the various pure strategies.

Some biologists object to this interpretation, because they see it as implying group or species selection rather than individual selection. The player is not the species, they argue; the individual "acts for its own good," not the good of the group, or of the population, or of the species. Some

even argue that it is the gene (or rather the allele) that "acts for its own good," not the individual. The point, though, is that *nothing* in this model really "acts for its own good"; nobody "chooses" anything. It is the process as a whole that selects the traits. The most we can do is ask what it is that corresponds to the player in the mathematical model, and this is undoubtedly the population.

A question that at first seems puzzling is what happens in the case of interactions within a species, like animal conflicts for females, etc. Who are the players in this game? If the players are the populations, then this must be a one-person game, since there is only one population. But that doesn't look right, either, and it certainly doesn't correspond to the biological models of animal conflicts.

The answer is that it is a two-person symmetric game, in which both players correspond to the same population. In this case we look not for just any Nash equilibria, but for symmetric ones only.

b. *Evolutionary Dynamics*

The question of developing a "truly" dynamic theory of games has long intrigued game theorists and economic theorists. (If I am not mistaken, it is one of the conceptual problems listed by Kuhn and Tucker (1953) in the introduction to Volume II of "Contributions to the Theory of Games"—perhaps the last one in that remarkably prophetic list to be successfully solved.) The difficulty is that ordinary rational players have foresight, so they can contemplate all of time from the beginning of play. Thus the situation can be seen as a one-shot game, each play of which is actually a long sequence of "stage games," and then one has lost the dynamic character of the situation.

The evolutionary approach outlined above "solves" this conceptual difficulty by eliminating the foresight. Since the process is mechanical, there is indeed no foresight; no strategies for playing the repeated game are available to the "players."

And indeed, a fascinating dynamic theory does emerge. Contributions to this theory have been made by Young (1993), Foster and Young (1990), and Kandori *et al.* (1993). A book on the subject has been written by Hofbauer and Sigmund (1988) and there is an excellent chapter on evolutionary dynamics in the book by van Damme (1987) on refinements of the Nash equilibrium. Many others have also contributed to the subject.

It turns out that Nash equilibria are often unstable, and one gets various kinds of cycling effects. Sometimes the cycles are "around" the equilibrium, like in "matching pennies," but at other times one gets more

complicated behavior. For example, the game

	0		5		4
0		4		5	
	4		0		5
5		0		4	
	5		4		0
4		5		0	

has $((1/3, 1/3, 1/3), (1/3, 1/3, 1/3))$ as its only Nash equilibrium; the evolutionary dynamic does not cycle "around" this point, but rather confines itself (more or less) to the strategy pairs in which the payoff is 4 or 5. This suggests a possible connection with correlated equilibria; this possibility has recently been investigated by Foster and Vohra (1997).

Thus evolutionary dynamics emerges as a form of rationality that is bounded, in that foresight is eliminated.

c. "*Rule Rationality*" *vs* "*Act Rationality*"

In a famous experiment conducted by Güth *et al.* (1982) and later repeated, with important variations, by Binmore *et al.* (1985), two players were asked to divide a considerable sum of money (ranging as high as DM 100). The procedure was that P1 made an offer, which could be either accepted or rejected by P2; if it was rejected, nobody got anything. The players did not know each other and never saw each other; communication was a one-time affair via computer.

"Rational" play would predict a 99–1 split, or 95–5 at the outside. Yet in by far the most trials, the offered split was between 50–50 and 65–35. This is surprising enough in itself. But even more surprising is that in most (all?) cases in which P2 was offered less than 30 percent, he actually *refused*. Thus, he *preferred* to walk away from as much as DM 25 or 30. How can this be reconciled with ordinary notions of utility maximization, not to speak of game theory?

It is tempting to answer that a player who is offered 5% or 10% is "insulted." Therefore, his utilities change; he gets positive probability from "punishing" the other player.

That's all right as far as it goes, but it doesn't go very far; it doesn't explain very much. The "insult" is treated as exogenous. But obviously the "insult" arose from the situation. Shouldn't we treat the "insult" itself endogenously, somehow explain *it* game-theoretically?

I think that a better way of explaining the phenomenon is as follows: Ordinary people do not behave in a consciously rational way in their

day-to-day activities. Rather, they evolve "rules of thumb" that work in general, by an evolutionary process like that discussed above (Section 1*a*), or a learning process with similar properties. Such "rules of thumb" are like genes (or rather, alleles). If they work well, they are fruitful and multiply; if they work poorly, they become rare and eventually extinct.

One such rule of thumb is "Don't be a sucker; don't let people walk all over you." In general, the rule works well, so it becomes widely adopted. As it happens, the rule doesn't apply to Güth's game, because in that particular situation, a player who refuses DM 30 does not build up his reputation by the refusal (because of the built-in anonymity). But the rule has not been consciously chosen, and will not be consciously abandoned.

So we see that the evolutionary paradigm yields a third form of bounded rationality: rather than consciously maximizing in each decision situation, players use rules of thumb that work well "on the whole."

2. PERTURBATIONS OF RATIONALITY

a. *Equilibrium Refinements*

Equilibrium refinements—Selten (1975), Myerson (1978), Kreps and Wilson (1982), Kalai and Samet (1984), Kohlberg and Mertens (1986), Basu and Weibull (1991), van Damme (1984), Reny (1992), Cho and Kreps (1989), and many others—don't really sound like bounded rationality. They sound more like super-rationality, since they go beyond the basic utility maximization that is inherent in Nash equilibrium. In addition to the Nash equilibrium, which demands rationality on the equilibrium path, they demand rationality also off the equilibrium path. Yet all are based in one way or another on "trembles"—small departures from rationality.

The paradox is resolved by noting that in game situations, one player's irrationality requires another's super-rationality. *You* must be super-rational in order to deal with *my* irrationalities. Since this applies to all players, taking account of possible irrationalities leads to a kind of super-rationality for all. To be super-rational, one must leave the equilibrium path. Thus, a more refined concept of rationality cannot feed on itself only; it can only be defined in the context of irrationality.

b. *Crazy Perturbations*

An idea related to the trembling hand is the theory of irrational or "crazy" types, as propounded first by the "gang of four" (Kreps, Milgrom, Roberts, and Wilson, 1982), and then taken up by Fudenberg and Maskin (1986), Aumann and Sorin (1989), Fudenberg and Levine (1989), and no doubt others. In this work there is some kind of repeated or other dynamic

game setup; it is assumed that with high probability the players are "rational" in the sense of being utility maximizers, but that with a small probability, one or both play some one strategy, or one of a specified set of strategies, that are "crazy"—that is, have no a priori relationship to rationality. An interesting aspect of this work, which differentiates it from the "refinement" literature and makes it particularly relevant to the theory of bounded rationality, is that it is usually the crazy type, or a crazy type, that wins out—takes over the game, so to speak. Thus, in the original work of the gang of four on the prisoner's dilemma, there is only one crazy type, who always plays tit-for-tat no matter what the other player does, and it turns out that the rational type must imitate the crazy type—he must also play tit-for-tat, or something quite close to it. Also, the "crazy" types, while irrational in the sense that they do not maximize utility, are usually by no means random or arbitrary (as they are in refinement theory). For example, we have already noted that tit-for-tat is computationally a very simple object, far from random. In the work of Aumann and Sorin, the crazy types are identified with bounded recall strategies; and in the work of Fudenberg and Levine, the crazy types form a denumerable set, suggesting that they might be generated in some systematic manner, e.g., by Turing machines. There must be method to the madness; this is associated with computational simplicity, which is another one of the underlying ideas of bounded rationality.

c. Epsilon-Equilibria

Rather than playing irrationally with a small probability (as in Sections $2a$ and $2b$), one may deviate slightly from rationality by playing so as almost, but not quite to, maximize utility; i.e., by playing to obtain a payoff that is within ε of the optimum payoff. This idea was introduced by Radner (1980) in the context of repeated games, in particular of the repeated prisoners' dilemma; he showed that in a long but finitely repeated prisoner's dilemma, there are ε-equilibria with small ε in which the players "cooperate" until close to the end (though, as is well-known, all exact equilibria lead to a constant stream of "defections").

d. Infinitely Repeated Games with Limit-of-the-Average Payoff

There is an interesting connection between ε-equilibria in finitely repeated games and infinitely repeated games with limit of the average payoff ("undiscounted"). The limit of the average payoff has been criticized as not representing any economic reality; many workers prefer to use either the finitely repeated game or limits of payoffs in discounted games

with small discounts. Radner *et al.* (1986), Forges *et al.* (1986), and perhaps others, have demonstrated that the results of these two kinds of analysis can indeed be quite different.

Actually, though, the infinitely repeated undiscounted game is in some ways a simpler and more natural object than the discounted or finite game. In calculating equilibria of a finite or discounted game, one must usually specify the number n of repetitions or the discount rate δ; the equilibria themselves depend crucially on these parameters. But one may want to think of such a game simply as "long," without specifying *how* long. Equilibria in the undiscounted game may be thought of as "rules of thumb," which tell a player how to play in a "long repetition," independently of how long the repetition is. Whereas limits of finite or discounted equilibrium payoffs tell the players approximately how much *payoff* to expect in a long repetition, analysis of the undiscounted game tells him approximately how to *play*. See Aumann and Maschler (1995), pp. 131–134.

Thus, the undiscounted game is a framework for formulating the idea of a duration-independent strategy in a repeated game. Indeed, it may be shown that an equilibrium in the undiscounted game is an approximate equilibrium simultaneously in all the n-stage truncations, the approximation getting better and better as n grows. Formally, a strategy profile ("tuple") is an equilibrium in the undiscounted game if and only if, for some sequence of ε_n tending to zero, each of its n-stage truncations is an ε_n-equilibrium (in the sense of Radner described above) in the n-stage truncation of the game.

e. Failure of Common Knowledge of Rationality

In their paper on the repeated prisoners' dilemma, the Gang of Four pointed out that the effect they were demonstrating holds not only when one of the players believes that with some small probability, the other is a tit-for-tat automaton, but also if one of them only believes (with small probability) that the other believes this about him (with small probability). More generally, it can be shown that many of the perturbation effects we have been discussing do not require an actual departure from rationality on the part of the players, but only a lack of common knowledge of rationality (Aumann, 1992).

3. AUTOMATA, COMPUTERS, AND TURING MACHINES

We come now to what is probably the mainstream of the newer work in bounded rationality, namely, the theoretical work that has been done since the mid-eighties on automata and Turing machines playing repeated games. The work was pioneered by A. Neyman (1985) and A. Rubinstein

(1986), working independently and in very different directions. Subsequently, the theme was taken up by Ben-Porath (1993), Kalai and Stanford (1988), Zemel (1989), Abreu and Rubinstein (1988), Ben-Porath and Peleg (1987), Lehrer (1988), Papadimitriou (1992), Stearns (1989), and many others, each of whom made significant new contributions to the subject in various different directions. Different branches of this work have been started by Lewis (1985) and Binmore (1987, 1988), who have also had their following. We do not even touch on the more recent work in this area, which has been very active lately.

It is impossible to do justice to all this work in a reasonable amount of time, and we content ourselves with brief descriptions of some of the major strands. In one strand, pioneered by Neyman, the players of a repeated game are limited to using mixtures of pure strategies, each of which can be programmed on a finite automaton with an exogenously fixed number of states. This is reminiscent of the work of Axelrod, who required the entrants in his experiment to write the strategies in a Fortran program not exceeding a stated limit in length. In another strand, pioneered by Rubinstein, the size of the automaton is endogenous; computer capacity is considered costly, and any capacity that is not actually used in equilibrium play is discarded. The two approaches lead to very different results. The reason is that Rubinstein's approach precludes the use of "punishment" or "trigger" strategies, which swing into action only when a player departs from equilibrium, and whose sole function is precisely to prevent such departures. In the evolutionary interpretation of repeated games, Rubinstein's approach may be more appropriate when the stages of the repeated game represent successive generations, whereas Neyman's may be more appropriate when each generation plays the entire repeated game (which would lead to the evolution of traits having to do with reputation, like "Don't be a sucker").

The complexity of computing an optimal strategy in a repeated game, or even just a best response to a given strategy, has been the subject of works by several authors, including Gilboa (1988), Ben-Porath (1990), and Papadimitriou (1992). Related work has been done by Lewis (1992), though in the framework of recursive function theory (which is related to infinite Turing machines) rather than complexity theory (which has to do with finite computing devices). Roughly speaking, the results are qualitatively similar: finding maxima is hard. Needless to say, in the evolutionary approach to games, nobody has to find the maxima; they are picked out by evolution. Thus, the results of complexity theory again underscore the importance of the evolutionary approach.

Binmore (1987, 1988) and his followers have modeled games as pairs (or n-tuples) of Turing machines in which each machine carries in it some kind of idea of what the other "player" (machine) might look like.

Other important strands include work by computer scientists who have made the connection between distributed computing and games (''computers as players,'' rather than ''players as computers''). For a survey, see Linial (1994).

4. RELAXATION OF RATIONALITY POSTULATES

A not uncommon activity of decision, game, and economic theorists since the fifties has been to call attention to the strength of various postulates of rationality, and to investigate the consequences of relaxing them. Many workers in the field—including the writer of these lines—have at one time or another done this kind of thing. People have constructed theories of choice without transitivity, without completeness, violating the sure-thing principle, and so on. Even general equilibrium theorists have engaged in this activity, which may be considered a form of limited rationality (on the part of the agents in the model). This kind of work is most interesting when it leads to outcomes that are qualitatively different —not just weaker—from those obtained with the stronger assumptions, but I don't recall many such cases. It can also be very interesting and worthwhile when one gets roughly similar results with significantly weaker assumptions.

5. AN OPEN PROBLEM

We content ourselves with one open problem, which is perhaps the most challenging conceptual problem in the area today: to develop a meaningful formal definition of rationality in a situation in which calculation and analysis themselves are costly and/or limited. In the models we have discussed up to now, the problem has always been well defined, in the sense that an absolute maximum is chosen from among the set of feasible alternatives, no matter how complex a process that maximization may be. The alternatives themselves involve bounded rationality, but the process of choosing them does not.

Here, too, an evolutionary approach may eventually turn out to be the key to a general solution.

REFERENCES

Abreu, D., and Rubinstein, A. (1988). ''The Structure of Nash Equilibrium in Repeated Games with Finite Automata,'' *Econometrica* **56,** 1259−1281.

Allais, M. (1953). ''Le Comportement de l'Homme Rationnel devant le Risque: Critiques des Postulats et Axioms de l'Ecole Americaine,'' *Econometrica* **21,** 503−546.

Aumann, R. J. (1992). ''Irrationality in Game Theory,'' in *Economic Analysis of Markets and Games, Essays in Honor of Frank Hahn* (P. Dasgupta, D. Gale, O. Hart, and E. Maskin, Eds.), pp. 214–227. Cambridge/London: MIT Press.

Aumann, R. J., and Maschler, M. (1995). *Repeated Games with Incomplete Information*. Cambridge/London: MIT Press.

Aumann, R. J., and Sorin, S. (1989). ''Cooperation and Bounded Recall,'' *Games Econ. Behav.* **1,** 5–39.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Basu, K., and Weibull, J. W. (1991). ''Strategy Subsets Closed under Rational Behavior,'' *Econ. Lett.* **36,** 141–146.

Ben-Porath, E. (1990). ''The Complexity of Computing Best Response Automata in Repeated Games with Mixed Strategies,'' *Games Econ. Behav.* **2,** 2–12.

Ben-Porath, E. (1993). ''Repeated Games with Finite Automata,'' *J. Econ. Theory* **59,** 17–32.

Ben-Porath, E. and Peleg, B. (1987). ''On the Folk Theorem and Finite Automata,'' Center for Research in Mathematical Economics and Game Theory, Hebrew University, *Res. Mem.* **77.**

Binmore, K. G. (1987). ''Modelling Rational Players, I,'' *Econ. and Philos.* **3,** 179–214.

Binmore, K. G. (1988). ''Modelling Rational Players, II,'' *Econ. and Philos.* **4,** 9–55.

Binmore, K., Shaked, A. and Sutton, J. (1985). ''Testing Noncooperative Bargaining Theory: A Preliminary Study,'' *Amer. Econ. Rev.* **75,** 1178–1180.

Cho, I.-K., and Kreps, D. (1987). ''Signaling Games and Stable Equilibria,'' *Quart. J. Econ.* **102,** 179–221.

David, P. A. (1986). ''Understanding the Economics of QWERTY: The Necessity of History,'' in *Economic History and the Modern Economist* (W. N. Parker, Ed.), Chap. 4. New York: Blackwell.

Ellsberg, D. (1961). ''Risk, Ambiguity and the Savage Axioms,'' *Quart. J. Econ.* **75,** 643–669.

Forges, F., Mertens, J.-F. and Neyman, A. (1986). ''A Counter Example to the Folk Theorem with Discounting,'' *Econ. Lett.* **20,** 7.

Foster, D., and Young, H. P. (1990). ''Stochastic Evolutionary Game Dynamics,'' *Theoret. Popul. Biol.* **38,** 219–232.

Foster, D., and Vohra, R. (1997). ''Calibrated Learning and Correlated Equilibrium,'' *Games Econ. Behav.* **21,** 40–55.

Fudenberg, D., and Levine, D. K. (1989). ''Reputation and Equilibrium Selection in Games with a Patient Player,'' *Econometrica* **57,** 759–779.

Fudenberg, D. and Maskin, E. (1986). ''The Folk Theorem in Repeated Games with Discounting and Incomplete Information,'' *Econometrica* **54,** 533–554.

Gilboa, I. (1988). ''The Complexity of Computing Best Response Automatan in Repeated Games,'' *J. Econ. Theory* **45,** 342–352.

Güth, W., Schmittberger, R. and Schwarze, B. (1982). ''An Experimental Analysis of Ultimatum Bargaining,'' *J. Econ. Behav. Organization* **3,** 367–388.

Hofbauer, J. and Sigmund, K. (1988). *Theory of Evolution and Dynamical Systems*, Cambridge: Cambridge University Press.

Kalai, E. and Samet, D. (1984). ''Persistent Equilibria,'' *Int. J. Game Theory* **13,** 129–144.

Kalai, E., and Stanford, W. (1988). ''Finite Rationality and Interpersonal Complexity in Repeated Games,'' *Econometrica* **56,** 397–410.

Kandori, M., Mailath, G. and Rob, R. (1993). ''Learning, Mutation, and Long Run Equilibria in Games,'' *Econometrica* **61,** 29–56.

Kohlberg, E., and Mertens, J.-F. (1986). "On the Strategic Stability of Equilibria," *Economet-rica* **54,** 1003–37.

Kreps, D., and Wilson, R. (1982). "Sequential Equilibria," *Econometrica* **50,** 863–894.

Kreps, D., Milgrom, P., Roberts, J., and Wilson, R. (1982). "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *J. Econ. Theory* **27,** 245–252.

Kuhn, H. W., and Tucker, A. W. (Eds.) (1953). *Contributions to the Theory of Games, Vol. II,* Annals of Mathematics Studies, Vol. 28. Princeton: Princeton University Press.

Lehrer, E. (1988). "Repeated Games with Stationary Bounded Recall Strategies," *J. Econ. Theory* **46,** 130–144.

Lewis, A. (1985). "On Effectively Computable Realizations of Choice Functions," *Math. Social Sci.* **10,** 43–80.

Lewis, A. (1992). "Some Aspects of Effectively Constructive Mathematics that are Relevant to the Foundations of Neoclassical Mathematical Economics and the Theory of Games," *Math. Social Sci.* **24,** 209–236.

Linial, N. (1994). "Game Theoretic Aspects of Computing," in *Handbook of Game Theory with Economic Applications*, Vol. 2, (R. J. Aumann and S. Hart, Eds.), Chap. 38. Amsterdam: North-Holland.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Myerson, R. B. (1978). "Refinements of the Nash Equilibrium Concept," *Int. J. Game Theory* **7,** 73–80.

Neyman, A. (1985). "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoners' Dilemma," *Econ. Lett.* **19,** 227–229.

Papadimitriou, C. H. (1992). "On Players with a Bounded Number of States," *Games Econ. Behav.* **4,** 122–131.

Radner, R. (1980). "Collusive Behavior in Noncooperative Epsilon-Equilibria of Oligopolies with Long but Finite Lives," *J. Econ. Theory*, **22,** 136–154.

Radner, R., Myerson, R., and Maskin, E. (1986). "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Rev. Econ. Studies* **53,** 59–69.

Reny, P. J. (1992). "Backwards Induction, Normal Form Perfection and Explicable Equilibria," *Econometrica* **60,** 627–649.

Rubinstein, A. (1986). "Finite Automata Play the Repeated Prisoners' Dilemma," *J. Econ. Theory* **39,** 83–96.

Selten, R. (1975). "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *Int. J. Game Theory* **4,** 25–55.

Simon, H. (1955). "A Behavioral Model of Rational Choice," *Quart. J. Econ.* **64,** 99–118.

Simon, H. (1972). "Theories of Bounded Rationality," in *Decision and Organization* (C. McGuire and R. Radner, Eds.). Amsterdam: North Holland.

Stearns, R. E. (1989). "Memory-Bounded Game Playing Computing Devices," Tech. Rep. 547, IMSSS, Stanford University.

van Damme, E. (1984). "A Relation between Perfect Equilibria in Extensive Form Games and Proper Equilibria in Normal Form Games," *Int. J. Game Theory* **13,** 1–13.

van Damme, E. (1987). *Stability and Perfection of Nash Equilibria*. Berlin: Springer-Verlag.

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

Zemel, E. (1989). "Small Talk and Cooperation: A Note on Bounded Rationality," *J. Econ. Theory* **49,** 1–9.