Book review, JOURNAL OF ECONOMICS, VOL. 70, 92-95, 1999.

Ariel Rubinstein: Modeling Bounded Rationality

Abdolkarim Sadrieh, University of Bonn.

Now, almost half a century after Simon (1957) coined the term "bounded rationality", there is only little controversy left concerning the importance of the issue. The amount of empirical and experimental evidence revealing the limitations of human knowledge, cognition, and computational abilities is so large that it simply can no longer be ignored. Therefore, the debate in the field of economic theory has shifted from the question *Is bounded rationality important?* to the question *How should bounded rationality be modeled?* In this way, modeling bounded rationality has turned out to be one of those new topics in economic theory that has gained an enormous innovative potential. The quest for answers has begun and already various approaches have emerged. Most of them are being followed in parallel.

Ariel Rubinstein's book is a remarkable milestone for one of the approaches to modeling bounded rationality. Without claiming to be a complete overview of the literature under concern, the book is an extensive and excellent compilation of lecture notes on numerous studies in the field, many of which actually are superb contributions by Rubinstein, himself. The common denominator to all the presented models is their foundation in mathematical economics, specifically in decision and game theory. In each case, the framework of rational choice is amended or extended to capture one of the various aspects of bounded rationality. Thus, the approach allows for limitations of rationality within the model by *relaxing* one or more of the axioms of rational theory. Both the selection and the configuration of the imposed limitations, however, are driven more by introspective and philosophical reasoning than by experimental or empirical analysis of human cognition and behavior. This approach has found a number of other very prominent followers, such as Aumann (1997). Since the approach resembles a blend of rationality principles with plausible, but not empirically based, hypotheses on the bounds of rationality, I suggest calling it the *conjectural approach*.

It is important to note that the difference between the conjectural approach and the *behavioral approach* - as propagated by Simon (1957), Selten (1990), and others - is not the lack of a behavioral foundation *per se*. As Rubinstein correctly notes (p. 193), not all models presented in the context of the latter approach have been experimentally validated either. But yet, there is an essential difference between the two approaches. Whereas the models in the behavioral approach are explicitly meant to be subject to the empirical and/or experimental testing of their predictive and descriptive power, the models in the conjectural approach are only

"meant to establish 'linkages' between the concepts and statements that appear in our daily thinking on economic situations." (p. 191) Thus, the evaluation criteria for the models in the conjectural approach are not behavioral relevance and empirical evidence. Instead, "the test of relevance is simply the naturalness of the concepts which we study and the ability to derive, by their use, interesting analytical conclusions." (p. 193)

Rubinstein deserves much praise for having included a *dialogue* with Herbert Simon on the controversy between the behavioral and the conjectural approaches in the final chapter of the book. Both Simon's comments on a preliminary manuscript and Rubinstein's replies to that critique are worthy of being read very scrupulously. The candid and discerning style of the debate is enlightening for both newcomers and veterans. Furthermore, the discussion is a clear signal for the sincere and deep interest of both parties to foster progress in this field.

The dialog in the final chapter reveals what can and what cannot be expected of the models presented in the main body of the book. As mentioned before, mathematical elegance and intellectual stimulation are ample, but empirical underpinning is not a major issue. This, however, does not mean that the carefully presented models are irrelevant or arbitrary. There is intuitive appeal to many of the ideas, that are systematically and comprehensibly developed.

The book begins with a short presentation of the conflict between the postulates of rationality and the empirical and experimental evidence. Following this, two models of choice are discussed that consider individuals' choices based on the *similarity* of alternatives or of consequences. Next, two models of knowledge are compared and some interesting results on the temporal aspect of knowledge (*Who knows what when?*) are derived. Related this aspect of knowledge, models concerned with limited memory of decision makers are presented. Finally, two models are examined, in which agents having a *coarse* memory must choose the optimal partition of their knowledge into the few compartments. Next, three aspects of team decisions are discussed. In chapter 7, a number of boundedly rational *equilibria* in games are introduced and compared to the rational counterparts. Following that, complexity issues in infinitely repeated games are examined and the interesting result is derived that implementing a preference for *simple* strategies leads to contradictions of the Folk Theorem. In the next chapter, possible resolutions of the *finite horizon paradoxes*, such as the Chain Store Paradox and as in the Centipede Game, are investigated under assumption that complexity is costly to the players. Finally, in chapter 10, logical computability restraints are deliberated and the question is asked, whether the existence of rational players is logically feasible. Interestingly, the answer to the question is not positive in general. Since a more comprehensive overview of the material in the book is beyond the scope of this report, I limit the more detailed discussion

to three selected models which, I find, have a strong intuitive appeal and are useful for understanding the logical limitations of the "rational man paradigm".

In the model examined in chapter 2, it is assumed that decision makers employ two *similarity relations* when evaluating risky prospects, one to compare the payoffs and one to compare the probabilities. Some lottery L_1 is preferred to some other lottery L_2, if either the probabilities of winning are *similar*, but the payoff of L_1 is greater than of L_2, or if the payoffs are *similar*, but the probability of winning is greater in the lottery L_1. The underlying intuition is that probabilities (or payoffs) can be perceived as similar, although they are actually unequal. This constitutes the element of bounded rationality that is built into the model. Using a specific family of similarity relations (the *λ-ratio similarity relations* $\sim$ that are defined by $a \sim b$ if $1/\lambda \leq a/b \leq \lambda$, with $\lambda>1$) that satisfy all proposed axioms, a number of experimental observations concerning violations of expected value maximization can be explained with the model.

Another interesting example of how rationality assumptions can be relaxed is the *absent-minded driver* paradigm (chapter 4). The absent-minded driver is on his way home. At the first junction reached on the highway, he should drive straight ahead, otherwise he will end up in a dangerous neighborhood. At the second junction, he should exit to get home, otherwise he will have to drive a long way before he can return. Absent-mindedness refers to the problem that the two junctions are indistinguishable for the driver. He can neither recognize where he is, nor can he remember how many junctions he has already passed. Here the *perfect recall* assumption, which is central both to game theory and to sequential decision theory, is dropped. Formally, the two decision nodes, junction one and two, are modeled as being in a single information set, even though they are in a sequence. The paradoxical result is that if the plan of actions is made beforehand, the driver might *optimally* plan to take a different action for some decision node than the action he *optimally* takes, once he is at that node. Thus, pre-play and in-play optimality are divergent, even though the decision maker receives no new information in the game. Furthermore, under such circumstances, not all outcomes achievable with behavioral strategies can be reached with mixed strategies.

The basic assumption of the equilibrium model with *procedurally rational players* (section 7.3) is that players use a two step decision process: First, each action is connected to exactly one outcome, where each of the possible outcomes of that action can be selected with some probability. Second, after having affiliated every action with its selected outcome, the player chooses the action with the highest payoff. The probability with which a specific action-consequence relationship is selected is equal to the frequency with which this outcome is observed under the assumption that all other players also act according to the same principle. This constitutes a *recursive* structure of behavior and beliefs, similar to the notion of

equilibria in traditional game theory. But, here the mutual belief in the full rationality of the others, which in game theory leads to expecting and exercising strategic behavior, is replaced by the mutual belief in the action-consequence orientation. The concept is theoretically very elegant and can explain some phenomena of observed behavior. However, since the underlying recursive concept is at least as demanding as in traditional game theory, the question arises, why the agents solve this specific fix-point problem instead of solving the traditional one.

Aumann, Robert J. (1997) "Rationality and Bounded Rationality". *Games and Economic Behavior*, 21, 2-14.
Simon, Herbert A. (1957) *Models of Man*. New York: Wiley.
Selten, Reinhard (1990) "Bounded Rationality". *Journal of Institutional and Theoretical Economics*, 146, 649-658.