Book review for *Economics and Philosophy*, October 2000.

Modeling Bounded Rationality, ARIEL RUBINSTEIN. MIT Press, 1998, xii + 208 pages.

Most of economic theory presumes that decision makers are rational: they correctly conceive of a well-defined set of alternatives, have well-defined preferences over these alternatives, and choose a best alternative. Evidence from many sources—experiments, interviews, field studies, introspection—indicates that such an idealistic picture in many cases does not accurately describe behavior. The reason is that decision makers simplify, misunderstand, lack ability, miscalculate, forget, and make evaluations of alternatives that depend on seemingly irrelevant details about how a problem is framed. A leading figure in calling economists' attention to these matters is Herbert Simon, who in 1978 received the Nobel Memorial Prize for his contributions. He introduced the term "bounded rationality" which refers to a decision procedure by which humans deliberate and arrive at decisions.

Usually the work of Nobel laureates has had profound effect on research practice. However, in his Nobel Lecture (published in the *American Economic Review* in 1979) Simon noted that ideas about bounded rationality had made little mark on mainstream economic theory. Twenty years later not much has changed, as explained by Reinhard Selten, also a Nobel laureate, in his 1997 Presidential Address to the *European Economic Association* (published in the *European Economic Review* in 1998): "The picture of rational decision making underlying most of contemporary economic theory is far away from observed behavior. It is therefore necessary to

develop theories of bounded rationality".[1] The appearance of Ariel Rubinstein's book is a signal that this is starting to happen. The author explains that his goal is modest: "This book is not intended to be a triumphal march of a field of research but a journey into the dilemmas faced by economic theorists attempting to expand the scope of the theory in the direction of bounded rationality" (p. 5).

Rubinstein argues that to analyze bounded rationality, details regarding *how* decision makers deliberate and arrive at decisions must be taken into account. In chapter 1, he presents a model of rational choice, as a benchmark of comparison for what follows. He formally shows that some forms of boundedly rational behavior can be alternatively understood "as if" they were rational. In particular, this is the case with Simon's famous *satisficing procedure*, where a decision maker does not pick a best alternative but rather looks around until he finds one that is good enough. As Simon himself showed in the mid-fifties, that procedure can alternatively be viewed as an optimization endeavor, if search costs are taken into account. To Simon, this result shows how an interesting conclusion can be derived with more realistic (and parsimonious) assumptions than those of rationality. To many (traditional) economists it provides a justification for using rationality assumptions, since these may do the trick even if the decision maker does not consciously optimize. Rubinstein notes that besides satisficing there are not many procedures that admit a description in terms of "as if rationality", and he suggests that the study of bounded rationality should consider other forms of behavior.

---

[1] Selten shared the 1994 Nobel Memorial Prize in economics with John Harsanyi and John Nash, "for their pioneering analysis of equilibrium in the theory of non-cooperative games". Over the past two decades, non-cooperative game theory has become an increasingly important part of mainstream economics. Selten's award-winning contribution was to develop the concept of (subgame) perfect equilibrium, which is a cornerstone of game theory and many economic models. It is interesting to note that he seems to view this contribution as a philosophical inquiry with no a priori relevance for describing human economic behavior.

Rubinstein's approach is to selectively relax a feature of the standard rationality assumptions, and to see what behavior is implied. Chapters 2-6 concern situations with single decision makers. What happens if a person simplifies decision problems where alternatives are similar? What if he makes imperfect inferences, or if he has limited memory? What if his ability to process information is limited? And what if the decision maker is an organization, restricted in some way by its internal communication structure? In chapters 7-10 the analysis deals with strategic interaction, situations where the decisions of multiple agents interact. How can the notion of equilibrium be developed in games where the players follow certain boundedly rational procedures? What happens if they cannot do backward induction, or if they are concerned with reducing the complexity of the strategies they choose? What if the players are Turing machines that condition their strategic choices on some description of the other players? Rubinstein treats these issues using a variety of rather different models, developed by himself and others. The presentation is always crisp and clear. In each case Rubinstein explains how the model with bounded rationality differs from some benchmark with rational decision makers. Each chapter ends with a set of "projects", a rich collection of challenging exercises and suggested readings. The reader who studies the material carefully is bound to discover many interesting research topics.

Here is an example of a model covered in the book. In section 7.3, a new equilibrium concept for "procedurally rational" play in games, developed by Rubinstein and Martin Osborne, is presented. One feasible interpretation of this equilibrium concept is the following: There is a population of individuals and every now and then a small group of individuals gets matched and play a given game. If the game is asymmetric, then a given individual always takes the same player position in

the game. Sometimes a newcomer will enter the population. The first few times (as many times as his total number of pure strategies) this fellow plays the game, he tries out each of his possible pure strategies once and records and remembers his payoff. He then clings forever to the strategy that led to the best result in these trials (if there was more than one best strategy, he selects one of these with equal probability). The equilibrium probability assigned to strategy x is the probability with which an individual entering the population (in the relevant position) ends up using strategy x. This probability is also equal to the fraction of the population (in the relevant position) that uses strategy x.

The interpretation of this new equilibrium is quite special. The interacting individuals are boundedly rational in that they follow a certain procedure that is not conceptualized as an optimization effort. An individual must correctly perceive his own realized payoffs, but he need not realize that he interacts strategically with others. By contrast, the usual interpretation of the well-known game theoretic concept of Nash equilibrium entails that the players understand the full game, and even have correct expectations concerning the behavior of other players.

The new concept is easy to apply. Imagine some kind of economic exchange where two individuals simultaneously choose activities *R* or *S*. Each individual's payoff equals the total productivity of the exchange *plus* the intrinsic satisfaction the individual derives from his chosen activity. Assume that there are gains to specialization, so the total productivity is 3 if the individuals choose different activities, but only 1 if they choose the same activity. Assume that the row player's intrinsic satisfaction is 1 from activity *R* and 0 from activity *S*. The column player's intrinsic satisfaction is 0 from activity *R* and 1 from activity *S*. The situation is described by the following game:

	R	***S***
R	2, 1	4, 4
S	3, 3	1, 2

To find all equilibria, suppose that in some equilibrium the proportion of column players choosing *R* is *p*. Now consider the behavior of a newcomer who plays in the row player's position. He first tries out his strategy *R*, and then his strategy *S*. If on both these occasions the column player chooses *R*, then the resulting payoffs for the row player will be 2 when he tries out *R* and 3 when he tries out *S*. Since 3 is greater than 2, the row player will subsequently choose *S* forever. However, if the column player's choices follow any other pattern, then the resulting payoff for the row player will be higher when he tries out *R* than when he tries out *S*. (For example, if the column player chooses *S* when the row player tries out *R*, and *R* when the row player tries out *S*, then the resulting payoffs for the row player will be 4 and 3 respectively. Since 4 is greater than 3, the row player will subsequently choose *R* forever.)

Hence the probability that a newcomer who plays in the row player's position ends up choosing *S* forever equals the probability that the column player chooses *R* in both the trial rounds. This probability is $p \cdot p = p^2$, since on each of the two trial rounds the probability that the column player chooses *R* is *p*. Thus, in equilibrium, p^2 is also the proportion of row players choosing *S*. By analogous reasoning, one sees that a newcomer who plays in the column player's position would end up using strategy *R* with probability $(p^2)^2 = p^4$. In equilibrium the proportion of column players choosing *R* equals the probability that a newcomer who plays in the column player's position would end up using strategy *R*, so $p=p^4$. This can only hold if $p=0$ or if $p=1$. One sees

that the game has two equilibria in pure strategies, (R,S) and (S,R), but no equilibrium in mixed strategies. By contrast, there exists a Nash equilibrium also in (non-degenerate) mixed strategies, so the example shows that the new equilibrium concept differs mathematically, not only in its interpretation, from that of Nash equilibrium.

In motivating the new equilibrium notion, Rubinstein does not refer to any empirical evidence that shows that the particular form of behavior assumed fits the facts of the world. It is clear that he finds the assumptions intuitive and interesting, and given this he is happy to work through the theory. Rubinstein's attitude towards the other models presented in the book is similar, although he stresses that the assumptions that go into a theoretical exercise should not be completely detached from reality. He writes: "I have to agree that an understanding of the procedural aspects of decision making should rest on an empirical or experimental exploration of the algorithms of decision. Too many routes diverge from the rational man paradigm, and the input of experimentation may offer some guides for moving onward" (p. 16). Rubinstein cites some experimental findings from which he takes inspiration.

One might think that Simon would welcome Rubinstein's initiative to write a book on bounded rationality. I wonder what Rubinstein thought while he was writing. He sent an early version to Simon who responded with very critical comments, calling Rubinstein's research methodology into question. Simon objects to the rather loose connection between the assumptions Rubinstein makes about boundedly rational behavior, and empirical findings about the behavior of humans. In the final chapter Rubinstein quotes from Simon's letters. Simon writes: "Aside from the use you make of the Tversky-Kahneman experiments, for which I applaud you and them, almost the only reference to empirical matters I detect in your pages is occasional statements like 'a casual observation' and 'the phenomenon exhibited here is quite common.' ...

[C]asual empiricism does not provide a firm foundation for the theories that fit the facts of the world. Facts do not come from the armchair, but from careful observation and experimentation. ... Using the rubric of 'bounded rationality' to expand the arena of speculation misses the intent of my nagging at the economics profession. At the moment we don't need more models; we need evidence that will tell us what models are worth building and testing". Simon furthermore charges Rubinstein with neglecting work in artificial intelligence and cognitive psychology that succeeds in describing human behavior quite well. He mentions formal theories that take the "form of computer programs that demonstrably simulate in considerable detail...a wide range of both simple and complex human behaviors. Little of the behavior that has been studied is explicitly economic, but that provides no excuse for ignoring its relevance to economic analysis." (All quotes here come from chapter 11.)

Rubinstein answers that his goal is not to predict and he writes: "The models are perceived as patterns of views adopted about the world. ... [W]e try to examine the logic of a variety of principles that guide decision makers. ... We are interested in a model only if it refers to concepts and considerations that make sense in the context of social interactions. ... A model with this approach does not have to be verifiable in the way models in the sciences must be." Rubinstein draws an analogy from a theory that does not presume bounded rationality: "From Hotelling's 'main street' model, we learn that the desire to attain as large a share of the market as possible is a force that pushes vendors (or political parties, or the makers of soft drinks) towards positioning themselves or their products in the center. In real life, the many other motives that influence a vendor's choice will cause him sometimes not to be located at the center. It is nonetheless insightful to identify the exact logic that leads an economist to the conclusion that the desire to maximize the share of the market leads a vendor to be

located in the center". Rubinstein explains that clear insights of this nature are not delivered by those complicated models in the artificial intelligence literature which Simon favours: "Those models may be capable of producing imitations of human behavior, but they are not convenient components for analytical work." (All quotes here come from chapter 11.)

I find that the methodologies of Rubinstein and Simon are best viewed as complementary. Simon points to interesting tools developed by researchers in other fields that can be usefully incorporated into economics, with the aim of developing good descriptive theories. In his Presidential Address (mentioned above) Selten reports several experimental results that he argues may be useful for developing an empirical-based microeconomic theory. The methodology Simon and Selten favors seems viable and valuable. I cannot, however, see that this in any way diminishes the value of Rubinstein's approach. In the introduction to his forthcoming book *Economics and Language*, Rubinstein explains that "all my academic research has been motivated by my childhood desire to understand the way that people argue. ... I continued to explore formal models of game theory and economic theory, though not in the hope of predicting human behavior...and without any illusion about the ability of capturing all of reality in one simple model." I think this is fine. Rubinstein's work is always intriguing and refreshing. It is full of insights about how to model and understand complicated social phenomena. Moreover, Rubinstein's methodological bent leads him to explicitly interpret and philosophize about the theories he considers, and his discussions are always thought-provoking and interesting.

In the end, I wonder if not the specific game I invented above doesn't indeed summarize rather well the interaction between people like Rubinstein and people like Simon. Consider a population of two sorts of researchers who have different opinions

on what research methodology is best, *R* or *S*. These fellows get matched and argue, defending a particular methodology. With Rubinstein as a row player, and Simon as a column player, I think they have coordinated on the (*R*,*S*) equilibrium. Given the payoffs I proposed, this seems like a happy state of affairs. I wonder if the players involved would agree with the payoffs I have given them though. In Simon's case, I see that this cannot be the case. He is clearly trying to make researchers move to the profile (*S*,*S*), which would not make sense for the given payoffs. With Rubinstein, however, I am not so sure. It seems clear that he enjoys his own methodology the most. Rubinstein writes: "The crowning point of making microeconomic models is the discovery of simple and striking connections between concepts...that initially appear remote" (p. 191) Moreover, he does not reject others' methodologies altogether, stating that "[t]he economics profession has several legitimate goals" (p. 194).

Martin Dufwenberg

Stockholm University

Biography:

Martin Dufwenberg is Associate Professor in Economics at Stockholm University. His research uses non-cooperative game theory and experimental methods to examine strategic interaction.