“Bounded Rationality”

By Coralio Ballester, Universidad de Alicante

and Penélope Hernández, Universidad de Valencia.

Abstract

The observation of the actual behavior by economic decision makers in the lab and in the field justifies that bounded rationality has been a generally accepted assumption in many socio-economic models. The goal of this paper is to illustrate the difficulties involved in providing a correct definition of what a rational (or irrational) agent is. In this paper we describe two frameworks that employ different approaches for analyzing bounded rationality. The first is a spatial segregation set-up that encompasses two optimization methodologies: backward induction and forward induction. The main result is that, even under the same state of knowledge, rational and non-rational agents may match their actions. The second framework elaborates on the relationship between irrationality and informational restrictions. We use the beauty contest (Nagel, 1995) as a device to explain this relationship.

1. Introduction

Bounded rationality is a concept used in different fields such as economics, psychology and computer science, among others. Many models of human behavior in the social sciences assume that humans can be reasonably approximated or described as "rational" entities that would never fail to behave in a rational way according to their preferences. The concept of bounded

rationality revises this assumption in order to account for the fact that perfectly rational decisions are often not feasible in practice due to the finite computational resources available for making them. Herbert Alexander Simon (1916–2001) was the first to coin the term “bounded rationality”. This leading scientific thinker, whose research ranged across the fields of cognitive psychology, computer science, economics and sociology, among other disciplines, stressed the limitations of the concept of a rational agent against a real human being. In his two seminal papers, one of which was written for the Rand Organisation and the other of which was published in the Quarterly Journal of Economics, Simon postulates that most people are only partly rational, and that they are in fact emotional in the remaining part of their actions.

Simon describes a number of dimensions along which "classical" models of rationality can be made somewhat more realistic. These include a class of utility functions, which may be multi-valued functions or recognizing the costs of processing information.

From a more theoretical point of view–particularly in game theory where the Nash equilibrium is the most important part–there is a vast amount of literature that examines the consequence of bounded rationality on the set of Nash equilibrium outcomes. Following Aumman’s suggestion to use simple strategies as an approximation to bounded rational strategies, Ariel Rubinstein (1986, 1988, 1990) and Abraham Neyman (1985,1998) proposed modeling such simple strategies as finite machines. After this seminal approach, an active branch of research was born, giving rise to a new approach to modeling repeated games. These studies have enhanced our understanding of the impact of limiting the set of possible strategies and provided answers on how

cooperation emerges or how to refine equilibria among the multiplicity correspondence from the well-known Folk Theorem. Actually, cooperation is justified under the assumption of bounded rationality.

Another way to capture the idea of bounded rationality is by assuming that individuals are on average rational, and that a large amount of people can be approximately modeled to act as boundedly rational agents by specifying explicit decision-making procedures. This puts the study of decision procedures on the research agenda.

Daniel Kahneman proposes bounded rationality as a model to overcome some of the limitations of the rational-agent models in economic literature. In a joint work, Kahneman and Twersky refute the standard use of the economic decision-making paradigm. Moreover, they prove the emergence of emotional and procedural elements. The authors explore the psychology of intuitive beliefs and choices, while examining aspects of bounded rationality. More specifically, they generate a map of bounded rationality by exploring the systematic biases that separate the beliefs that people have, the choices they make from optimal beliefs, and the choices assumed in rational-agent models. This work is a pioneering approach that lies at the foundation of behavioral economics. In addition, bounded rationality suggests that economic agents employ heuristics to make decisions rather than a strict rigid rule of optimization in light of the complexity of the situation, or the inability to process and compute all the possible alternatives due to deliberation costs and the presence of other economic activities.

Nowadays, it is widely accepted among the scientific community that human

beings are limited either by the information they have, by their computational ability, or by the cognitive limitations of their minds.

If we accept the fact that boundedly rational agents are limited in formulating and solving complex problems and in processing information, then we can accept computational approaches for understanding the decision-making process. In particular, models of bounded rationality help construct inference models and simulate human behavior by using computers. Edward Tsang (2008) argues that the effective rationality of agents is determined by their computational intelligence. Therefore, we assume that, in many circumstances, decision makers lack the ability and resources to arrive at the optimal solution, and instead apply their rationality only after having greatly simplified the choices available in a pre-processing stage.

A common thread weaves through all of these branches, namely the idea that the rational man is a “rara avis”. Behavioral economics, in particular, is the branch where bounded rationality is the central theme. Indeed, the notability of behavioral economics stems precisely from the fact that it connects the assumption of bounded rationality with other disciplines.

2. Bounded Rationality and Segregation

2.1. Schelling’s segregation model

Tomas Schelling was a forerunner in the study of segregation. In his first paper (1969, 1971a), he points out the impact of aggregating individual preferences on the final landscape of the society. Schelling’s model starts by assuming a social context with n agents located in a geographical society. In his model,

there are two types of agents: black agents and white agents. All agents have a preference over the mixed structure of their neighborhood, but not over the specific configuration. This preference establishes when agents are happy in such a society. More specifically, a society consists of a set of agents located on a line or in a circle. Agents' utility is only affected by the structure of their local neighborhood, i.e., by the agents to the right and to the left of them. Agents are defined by "type" (e.g., black or white) and by a number denoted as "tolerance", which specifies the minimum ratio of close neighbors that must be of their same type in order to reach a state of happiness. For simplicity sake, we assume that the utility of each agent can be one or zero when the agent is happy or unhappy, respectively. For instance, an intolerant agent would be one who demands that all the neighbors next to her be of her same type, while a moderately tolerant agent would accept that half of her neighbors were like her. Schelling, moreover, allows unhappy agents to move across the geographical society to improve their individual levels of happiness. Specifically, each unhappy agent (following an exogenously given order) will move to the nearest place where she becomes happy (if any), that is, to the nearest position on the line or in the circle where she attains the minimum fraction of like neighbors as determined by her tolerance level. In other words, Schelling models economic agents as simple machines that can compute the nearest location where the agent's new mixed neighborhood satisfies her tolerance level of happiness, and is able to act according to this computation: when it is an unhappy agent's turn, the agent will either decide to move or to stay.

2.2. From machines to more rational agents

Let us now consider a completely different framework from the above scenario: a rational agent. We will examine an instance of a society with eight agents: four black agents and four white agents. The agents are displayed as a ring (a circle) with alternative colors (Figure 1a). Let us fix a mild tolerance level, in other words, each agent wants to share at least one neighbor like her (out of the two actual neighbors). Suppose that agents must find their closest matching location only clockwise. Notice that the society in this example is an unhappy society since no individual exceeds her tolerance level. Consequently, starting from the top individual and following in a clockwise direction, the eight players should compute the nearest place and move to that place. Following the bounded rational dynamics suggested by Schelling, the final configuration is a fully segregated society. Figure 2 illustrates the movements and how a fully segregated society is reached.

Figure 1: Integration (a) versus segregation (b) with N=8 subjects

Figure 2: Movements in the sequential game with bounded rational agents

But what happens if all players are rational? This new situation can be modeled as an extensive game with eight players. Each player may choose two possible actions: to move or to stay. The final configuration depends on the path played by all players and leads to a positive payoff (payoff 1) for each happy player and zero otherwise. Given this game, we can compute the subgame perfect equilibrium. Players will choose their optimal action (to move or to stay) by reasoning backwards, yielding the maximum grade of rationality since at any stage the individual should anticipate what would be the best response of the

remaining players and play accordingly. Actually, this problem could be considered difficult as the number of computations needed to act rationally can be exponential in the number of players. Moreover, we may find a multiplicity of equilibria of equilibrium strategies, but only two final possible configurations: either the fully segregated society or a mixed configuration.

Now let us enrich the payoff structure in the following way. Suppose that each player incurs a positive cost c if she decides to move. In such a case, a player may guarantee a positive payoff 1+c if she reaches a happy society and stays in her original position. Nevertheless, if she reaches the happy position but moves, then she only guarantees the happiness payoff 1. In contrast, if she obtains an unhappy situation at the end of the game, she will earn c or 0 depending on whether she stays or moves, respectively. Under this incentive structure, as Benito, Brañas, Hernández and Sanchis (2010) prove, there is a unique subgame perfect equilibrium, and therefore a unique equilibrium path identified by the movements of only players 4 and 8.

In this new “costly” environment with rational agents, we are able to characterize what is called a rational player. For instance, a rational player 1 should stay in her initial position even when she is not happy. This is so because she can anticipate the best response of player 2, player 3, until player 8. In particular, the movement of player 8 as her best response will force the happiness of player 1. In the same way, player 2 will react identically, and so on. But what should be the prediction of a rational player off the equilibrium path?

Suppose that player 1 has already moved between players 2 and 3. Therefore, both players become happy since player 2 is closer to player 8, both are white, and player 1 and 3 (both black) stay together. This action is not on the equilibrium path. When it is player 4's turn to play, she has to choose an action given the above history that conveys information about the past agents' choices in the society. First, player 1 did not behave rationally. Second, she is not able to discriminate the rationality level of players 2 and 3. Hence, player 4 faces a dilemma: either the remaining players are all rational or they are not. If she considers the first assumption, then her best response following the backward induction methodology is to move. Nevertheless, she could consider a different approach to tackling this problem: she could react according to a forward induction rather than a backward induction, in other words, player 4 assumes that the remaining players are rational depending on the rationality observed in the previous stages.

When players have to face unexpected events, they need to attach meaning to such events (see Govindan and Robson,1998). Namely, in our example, player 4 has to interpret player 1's action which may convey information about the agent's level of rationality. Therefore, player 4 may generate different beliefs such as "no player is rational" or "with probability 2/3, players 5, 6, 7 and 8 are rational". The first belief is consistent with the fact that player 1 is not rational and the actions of players 2 and 3 do not convey any information. The second belief assumes that player 2 and 3 were rational and that player 1 was the only irrational player. Therefore the rest of the society should have the same distribution of irrational individuals. When player 4 computes her optimal strategy, her best response will naturally depend on these beliefs.

This stylized example stresses the difficulties involved in defining bounded rationality. The same action of player 4 could be explained as being that of a fully rational agent or exactly the opposite! Therefore different actions may convey a different level of rationality or not if we assume a change in reasoning.

2.3. Complexity theory and tractability of problems

Complexity theory has been shown to be a useful tool for determining the significance of bounded rationality in economic decisions. There are many approaches in computer science and mathematics that address the notion of complexity, which refers to the difficulty of a general task to be solved (see Kolgomorov 1998, Garey and Johnson 1979, Solomonov 2009). We focus on a particular dimension of this notion of complexity: time complexity.

From the computational point of view a *problem* is a general question regarding some abstract object. For instance, a problem could be to determine whether there exists a Nash equilibrium in a game when players are not allowed to randomize in the choice of their strategies (that is, the problem of determining the existence of a pure-strategy Nash equilibrium).

An *instance* of a problem is the particularization of this general question. For example, to determine whether there is a pure-strategy Nash equilibrium in a "beauty contest" game (Nagel, 1995) with three players. The beauty contest is a guessing game where all agents must simultaneously announce a number between 0 and 100. The player(s) whose announcement is closest to two-thirds of the mean wins a prize (for instance 90 euros), which is split equally in the case of ties. For instance, suppose that there are n=3 players who simultaneously announce 20, 30 and 40, respectively. In this case, 2/3 of the

mean is 20 and player 1 gets the prize. Note that player 3 (who makes the highest announcement) regrets her announcement, since given the other players' announcements she would have won the prize by announcing, for instance, 13. In that case, 2/3 of the mean would have been 14 and she would have won the prize. In this respect, the announcements 20, 30 and 40 do not constitute a Nash equilibrium. Now consider the new announcements where player 3 anticipated the other players' actions and the announcements are: 20, 30 and 13. Here, player 2 also regrets her decision: if she had announced 12, 2/3 of the mean would have been 10 and she would have won the prize. Hence, the announcements 20, 30 and 13 do not constitute a Nash equilibrium either. It turns out that the only Nash equilibrium occurs when every player announces 0. Here, everyone gets the split prize (30 euros each) and no individual has incentives to deviate from her announcement. Intuitively, a strictly positive announcement by any player would trigger regret by some other player and only a zero announcement by all players can be sustained as an equilibrium.

Even though the answer to this particular instance is yes (since all players guessing zero is an equilibrium in pure strategies), there are instances of other games where players cannot achieve an equilibrium in pure strategies.

Another problem is to determine the best-response to other players' strategies in the beauty contest. A particular instance of this problem would be a set of strategies (guesses) by the other players (for instance, 30 and 40): the solution for this particular instance would be 20. In this particular problem, note that any instance or situation can be handled by a player with a relatively cheap amount of resources: just a small set of basic operations must be performed in order to give a best-reply to opponents' strategies. More specifically, this amount of

computation is not expected to increase more than linearly as the problem increases with the number of opponents. In this sense, we say that best-replying in the beauty contest is a problem of linear (polynomial) complexity; an easy problem. In contrast, we may find other problems that are exponentially difficult in which only relatively small instances will be handled (solved) in a reasonable amount of time.

A usual way of measuring the difficulty or complexity of an instance is to measure the time that the most efficient machine (or algorithm) would take to solve that particular instance. Finally, the time-complexity of a problem is measured as the worst time among these efficient solutions (worst-case complexity) or the average time among these efficient solutions (average-case complexity). In other words, the (worst case) time-complexity of a problem is the time that the best machine would take to solve the most difficult instance of that problem. Naturally, complexity must be measured as a function of the size of the instance, given that larger instances will need a longer time to be processed.

Surprisingly, it has been shown that some problems are very difficult in nature, that is, machines will encounter difficulties to solve certain instances of these problems in a reasonable period of time. Or even worse, there are problems that can never be solved by regular computers! (for a detailed study of computational complexity, see Garey and Johnson 1979). This proposition provides a very sharp intuition of the natural limitations of human brains in dealing with problems.

Problems of such intrinsic complexity include the computation of the Nash equilibrium or best responses in some games. This makes it difficult to associate human behavior with that of a completely rational agent playing the Nash equilibrium, and calls for a model of bounded rationality. In fact, many other economic environments, such as the formation of coalitions or the formation of economic and social networks involve decisions by players or social planners that are complex in nature. Ballester (2004), for instance, studies the difficulty of problems faced by decision makers (social planners) who must achieve the stable organization of society in the sense of minimizing the moves across different groups by “unhappy” agents. He shows that many general problems of this type are NP-complete, a computational notion of complexity that includes a vast amount of well-known problems that are very unlikely to be solved in polynomial time.

3. Rationality and Information

As we already pointed out, bounded rationality is related to situations where an agent’s decision-making process does not completely adhere to classical rationality assumptions. A further aspect of bounded rationality is incomplete information, which takes into account the possibility of scarce information about the actual state of the world when agents are faced with economic decisions. In principle, this latter concept is independent of rationality, that is, we can have fully rational agents making decisions in a world with uncertainty, or completely informed agents that are not fully rational, or both.

In order to understand this distinction more clearly, we will illustrate it by means of the beauty contest explained in section 2.

3.1. An incomplete information setting

Consider now a situation in which each player *i* is privately informed about the set of players for which she must guess 2/3 of their mean, that is, player *i* receives a guessing assignment that only she knows. This assignment is generated by some random rule that is commonly known by all players. A player's strategy consists of a guess that is contingent on the assignment that she receives.

One example could be a situation with n=2 players where every player is in her opponent's guessing assignment with a probability of 1/2. Each player *i* would have to guess about all 2 players with a probability of 1/2; and only about herself with a probability of 1/2 (in this case, the player has a clear advantage: her guess would be straightforward and she would earn something).

The interesting point in this new incomplete information framework is that each player privately knows the set of players she must make predictions about (her guessing assignment), but she does not know what the other player must guess. Hence, in one particular situation, player 1 could be randomly assigned to guess about players 1 and 2 and she would have this information, but she would not know which set of players player 2 must make predictions about. Here, a player must form beliefs about the actual guessing assignment of the other player. Moreover, she must also assess what beliefs the other player has about her own guessing assignment. And so on.

In this new context of incomplete information, the equilibrium only occurs if both players announce 0, independently of the guessing assignment that they each received according to the assignment rule. Intuitively, equilibrium behavior under complete information also turns out to be an equilibrium under incomplete information. More importantly, in this particular example, it is the unique optimal behavior.

3.2. Information and rationality

We now turn our attention to the complete information case in order to relate our previous example to a bounded rationality setting. We use the same beauty contest framework, where bounded rationality is defined using the concept of k-level rationality as in Nagel (1995). To simplify matters, we define a 0-rational player as one whose guess is a focal point such as 50. A 1-rational player is a player that best-replies to 0-rational opponents. For instance, in the case of 3 players, a 1-rational player would choose the best guess assuming that her opponents guess is 50, that is, her guess would be around 29. A 2-rational player would be a player that best-replies to 1-rational opponents (playing 29), so that she would play around 16 (for the sake of simplicity, we adopt a slightly different notion from Nagel's definition of k-level rationality). Continuing this reasoning ad infinitum, it would be easy to verify that an infinitely rational player would play 0, which is the Nash equilibrium. In many situations, infinitely rational players may not play some equilibria that are unstable or not reachable through this process.

Yet can we relate levels of information to levels of rationality? For instance, in the case of incomplete information, let us define (as in Ballester, Ponti and van der Leij, 2010) a 1-informed player as a player who knows only her assignment, a 2-informed player as a player who knows her own assignment and the assignment of the players appearing in her own assignment, a 3-informed agent would be defined accordingly, etc… It is important to note that in our example with incomplete information, two 2-informed players correspond to a game with complete information as long as the level of information possessed by each player is common knowledge: each player knows her own assignment and her opponent's, and this is common knowledge.

The question that arises is whether there is some common pattern in the behaviors of a boundedly-rational agent (for instance, a 1-level rational payer) and a partially informed agent (for instance, a 1-informed player). This problem has been addressed in Ballester et al. (2010), who study the behavior of experimental subjects under both frameworks. Even though they focus on a different game structure, they show that subjects in the lab tend to behave similarly under incomplete information and under rationality limitations. These two notions are different and it is difficult to provide a satisfactory theory for this feature. This is mainly due to the fact that the task of mapping levels of information to levels of rationality involves building an ex-ante link between these two notions.

In order to make this important point clear, let us return to our examples with the beauty contest. Note that a minimum level of information (like 1) by all players can lead to optimal equilibrium behavior (that is, guessing 0), while all players are required to have a sufficiently high level of rationality in order to make

guesses that are close to the unique zero-equilibrium. Partially informed agents are able to reason ad infinitum within their limited informational environments, while boundedly rational agents are only able to perform limited calculations in a complete information set-up.

A completely different approach to establishing this relationship between information and rationality has been proposed by computer scientists. The rationality-information pair can be easily mapped to the time-space pair in computer science. Suppose that a computer has to derive its optimal strategy (such as guessing in the beauty contest) against the other computers. The input of the machine is the information available about the game payoffs and possibly about the other machines' configurations. Here the level of rationality may correspond to the complexity of the machine circuits, while the level of information may correspond to the input size or the memory size. In this computational context, at least one direct relationship arises between both concepts: information bounds rationality in the sense that in order to read or process all the available information, we at least need computational resources in order to read it! This fundamental relationship is not considered under the game-theoretical approach where public or private information is assumed to be known by players, but the process by which agents come to acquire this knowledge is not specified.

4. Conclusion

In this paper, we have explored some of the fundamental difficulties involved in identifying bounded rationality in economics. On the one hand, these difficulties have to do with the distinction between rational behavior and rational thinking as shown through an example using the Schelling's segregation game. On the other hand, there seems to be a link between rationality and information, but again it may be difficult to disentangle both dimensions from the observation of individual behavior.

References:

- Abreu, D., and A. Rubinstein (1988): "The Structure of Nash Equilibrium in Repeated Games with Finite Automata", Econometrica(56) 1259 -1282.
- Ballester, C. (2004). NP-completeness in hedonic games. Games and Economic Behavior, 49, 1-30.
- Ballester, C., Ponti, G., and van der Leij, Marco (2010). Bounded rationality vs. incomplete information in network games. Mimeo.
- Benito, J., P. Brañas, P. Hernández, and J. Sanchis (2010): "Strategic behavior in Schelling dynamics: A new result and experimental evidence", Working Paper ERI-CES.
- Garey, Michael R. and Johnson, David S. (1979). Computers and intractability. A guide to the theory of NP-completeness. W.H. Freeman.

- Govindan, S., and A. J. Robson (1998): “Forward Induction, Public randomization and Admissibility,” Journal of Economic Theory, 82, 451–457.
- Kahneman, Daniel (2003). Maps of bounded rationality: psychology for behavioral economics. The American Economic Review. 93(5). pp. 1449–1475.
- Kolmogorov (1998). On tables of random numbers. Theoretical Computer Science, 207(2), 387-395.
- March, James G. (1994). A Primer on Decision Making: How Decisions Happen. New York: The Free Press.
- Nagel, R. (1995). “Unraveling in Guessing Games: An Experimental Study,” *American Economic Review* **85** (5): 1313-1326.
- A. Neyman. (1985). “Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner’s Dilemma”, Economic Letters, (19), 227-229.
- A. Neyman. (1998). “Finitely Repeated Games with Finite Automata” Mathematics of Operations Research,(23) 513-552.
- Rubinstein, A. (1986). “Finite Automata Play the Repeated Prisoner Dilemma”. Journal of Economic Theory (39) 83-96.
- Rubinstein, A. (1998). Modeling bounded rationality, MIT Press.
- Schelling, T.C. (1969). “Models of Segregation”, American Economic Review, Papers and Proceedings 59: 488-493.
- Schelling, T.C. (1971a). “Dynamic Models of Segregations”, Journal of Mathematical Sociology 1 (2): 143-186.
- Simon, Herbert (1957). "A Behavioral Model of Rational Choice", in Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting. New York: Wiley.

- Simon, Herbert (1990). A mechanism for social selection and successful altruism, Science 250 (4988): 1665-1668.
- Simon, Herbert (1991). Bounded Rationality and Organizational Learning, Organization Science 2(1): 125-134.
- Solomonoff, R.J. (2009). Algorithmic probability: Theory and applications, Information theory and statistical learning. Springer NY.
- Tsang, E.P.K. (2008). Computational intelligence determines effective rationality, International Journal on Automation and Control, Vol.5, No.1, 63-66.
- Williamson, Oliver (1981). The economies of organization: the transaction cost approach. American Journal of Sociology 87 (3): 548-577.

Tables and Figures:

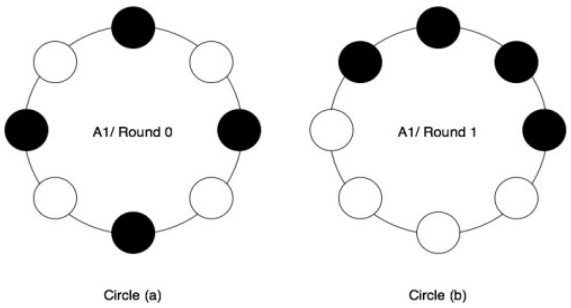

Figure 1: Integration (a) versus segregation (b) with N=8 subjects

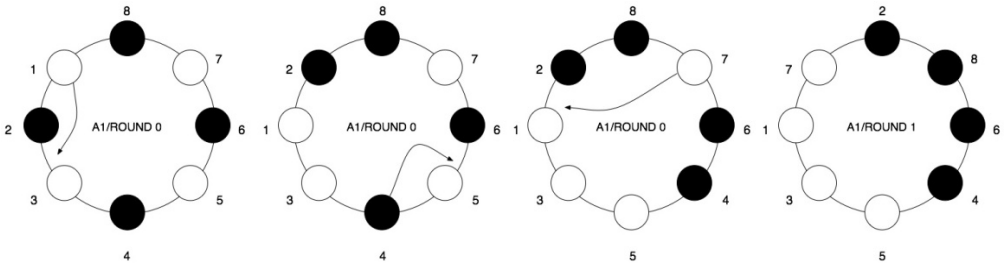

Figure 2: Movements in the sequential game with bounded rational agents