Forms of Bounded Rationality: The Reception and Redefinition of Herbert A. Simon's Perspective

(Article begins on next page)

DOI: 10.1080/09538259.2011.611624

Forms of Bounded Rationality: The Reception and Redefinition of Herbert A. Simon's Perspective

STEFANO FIORI*

Università di Torino, Italy

ABSTRACT *Firstly, the paper seeks to show that Herbert Simon's notion of bounded rationality should be interpreted in light of its connection with artificial intelligence. Secondly, offering four paradigmatic examples, the article presents the view that recent approaches that draw upon Simon's heterodox theory only partially accept the teachings of their inspirer, splitting bounded rationality from the context of artificial intelligence, and replacing it with different analytical tools which help give new configurations to bounded rationality. The thesis is that these events can be interpreted as an implicit (and ideal) challenge for redefining what bounded rationality is.*

1. Introduction

The thesis of this paper, firstly, is that in recent studies work on non-neoclassical approaches to economic behavior the meaning of bounded rationality (BR) has changed with respect to Herbert A. Simon's original formulation, although he is routinely acknowledged as the undisputed, fundamental reference point. In particular, I point out that Simon's original perspective, which connects BR and Artificial Intelligence (AI), has been essentially abandoned, and new theoretical tools have replaced it. This yields a number of consequences, such as for example, the reduction of the role of intentionality in favor of unconscious and automatic mechanisms in decision-making. This change is not the outcome of an explicit debate from which a new coherent view of BR has taken shape; rather, it emerges within specific contexts related to both particular subjects and the use of specific methods of inquiry, from economic psychology and experimental

* *Correspondence Address:* Stefano Fiori, Department of Economics, Università di Torino, Via Po, 53, 10124 Torino, Italy. Email: stefano.fiori@unito.it

economics to evolutionary theories of the firm. As a consequence, new applications (if not new versions) of the notion of BR appear, exhibiting differences and analogies.

Secondly, given the existence of recent and diverse representations of BR, I assume that there is an implicit challenge in redefining what constitutes BR, although in some cases these representations can be seen as complementary perspectives on the same topic. The lack of a unitary view on BR seems important, because non-neoclassical approaches (as a class of theories which strive to replace the orthodox view) include BR more or less explicitly among their fundamental categories, and this should imply possession of a clear definition of this concept.

The paper is divided into two parts. The first discusses Simon's notion of BR in the light of its connection with AI. The theoretical significance of this well-known connection has been underestimated, because insufficient consideration has been given to how Simon's AI shaped the notion of BR. This approach gave a highly structured form to BR, making it a complex notion not reducible to an evocative label simply opposed to the idea of perfect rationality. These arguments are treated in Sections 2, 3 and 4, where I examine how Simon's view of information processing has been incorporated into BR. Moreover, BR is analyzed along with the concept of 'intuition', which Simon explored in a mature phase of his research on AI, because this perspective makes it possible to emphasize certain problematic aspects of Simon's approach to rationality. In fact, according to Simon, *intentional* BR, as a decision-maker's conscious activity, and intuition (as an unintentional cognitive process) represent two extremes of rational behavior which pertain to the same information processing mechanism, and are explained by the latter. Since the notion of intuition has also been investigated by Daniel Kahneman and Gerd Gigerenzer, their approaches are compared, with the differences pointed out. In fact, these scholars do not interpret 'intuition' in light of AI as the work of symbolic operations; instead, they consider it either as an automatic mode of thinking, alternative to explicit rationality (Kahneman), or as an unconscious-adaptive capacity for decision-making developed by evolutionary processes (Gigerenzer).

The second part of the paper discusses how BR has been dealt with by some representative authors, such as March, Nelson, Winter, Kahneman, Tversky, Gigerenzer and others (Sections 5). I refer to these as paradigmatic cases because they share two features of importance for the present analysis: (1) they have acknowledged their intellectual debt to Simon, and their work in turn has been appreciated by Simon himself; (2) they have accepted Simon's non-standard approach to economics.[1] Notwithstanding these premises, examination of these views on BR suggests that Simon's teaching has been only partially accepted, and that they essentially do not refer to AI but replace it with different analytical tools. Important and stimulating perspectives emerge from such theoretical approaches, which share Simon's intent of providing an alternative to neoclassical economics, but introduce significant shifts with respect to Simon's original program. In particular, an implicit tendency to redefine the notion of BR seems to emerge as a consequence of these changes in methodologies and perspectives, the nature of which is not entirely compatible with Simon's understanding of AI. The latter is *tacitly* dealt with as a non-essential tool for analysis, since, on the one hand, it seems unable to capture certain characteristics of (bounded) rationality, while on the other, it stresses the features of instrumentality, non-ambiguity and intentionality, which delineate an image of rationality distant from that provided by new approaches. By contrast, emotions, intuitions, ambiguity, unconscious and automatic mechanisms related to neurophysiology, adaptive and evolutionary capabilities, etc connote the new views on BR, where these latter basically refer to disciplines and epistemologies which are not derived from (and sometimes are opposite to) AI assumptions. Therefore, one outcome is the appearance of attributes of rationality which show how its intentional and conscious component is *weaker* than the representation provided by Simon.

[1] Simon first put forth the essential elements of BR in his *Administrative Behavior* (1947), but he soon came to regard this version as unsatisfactory (Simon, 1979, pp. 501−502). The concept was subsequently specified as an economic notion in the mid-1950s, during the initial phase of Simon's interest in AI. This latter specification constituted a fundamental change, in that BR assumed a more definitional form that was more appealing to economists, and was more clearly conceived as an alternative to the neoclassical notion of rationality. During that period, AI and cognitive psychology had become increasingly important in Simon's scientific research.

2. Bounded Rationality: A General View

Most economists essentially know Herbert A. Simon as the theoretician of BR.[2] Moreover, most of them are familiar mainly with his contributions of the 1950s and thereafter: 'A Behavioral Model of Rational Choice' (1955) is probably the most frequently cited of Simon's articles. His intention in that paper was to replace the model of optimal rationality with another model that takes account of the limited computational capacities and limited access to information that characterize real decision environments. In short, treatment of Simon's thought generally centers on the notion of BR (and correlated concepts), taking as its reference point the version that emerged in the mid-1950s, along with some subsequent developments. Hence, I shall not analyze how Simon modified the concept of BR from *Administrative Behavior* (1947) to *Models of Man* (1957), where the term BR first appeared, although the ten years that separate the two books were very important for Simon as regards refinement of this concept (Klaes & Sent, 2005, p. 37).

The theory of BR, in its mature version, deals with the limits of 'information processing capacities'. More precisely, the definition of information processing stresses two constraints on individuals: (1) the limits of the information which is gathered and processed; and (2) the limits of computational capacities, which emerge when agents face situations perceived as complex. Given these limits, the decisional process applies when problems requiring solutions occur. More precisely, March & Simon (1958) emphasize that when actors (organizations, in this case) receive an external input or stimulus, they react either by replicating past behaviors (*routines*), if they do not encounter unforeseen situations, or by following new courses of action in order to solve new and unexpected situations. In the latter case, *problem-solving* procedures are activated, and agents treat complex problems sequentially. In particular, they reach a node (a choice point) using information collected in the previous steps, and this information, in turn, allows them to gather new information. Since uncertainty prevails in the real world, neither all the alternatives nor the

[2] Here I do not consider notions of BR as incorporated into 'standard' approaches. Moreover, I do not analyze how BR has been treated in game theory, new institutionalism or other fields of economics.

consequences that would follow from them can be evaluated, and heuristics (such as rules of thumb) are adopted to simplify search processes (Newell, Shaw & Simon, 1962; Newell & Simon, 1972; Simon, 1990). Therefore, decision-makers look for a 'satisficing', rather than an optimal, alternative; the criteria that perform this role in decision-making processes are called 'aspiration levels'. Satisficing is a 'weak method' which problem-solvers apply when task domains are ill-structured or are unknown; it allows them to halt the search process when a solution meets their aspiration threshold. Finally, (sequential) problem-solving procedures are attained by breaking down the problem into smaller components: since, problems are often too complex for the agents' computational capacities, this requires their decomposition into sub-problems that are less computationally complex.

The theory of BR is closely interwoven with Simon's research in both economics and AI, as his biography shows. In 1952, Simon met Allen Newell at the System Research Laboratory of the RAND Corporation, and in the same period started to develop a chess program inspired by the idea that chess players use not the best strategies, but satisfactory ones (Cordeschi, 2002, p. 178). By the beginning of 1955, along with Clifford Shaw, the group was working on the programming systems for the JOHNNIAC, the computer built at RAND in the early fifties, and, between 1955 and 1957, they made great advances in AI, defining languages and programs for computers. Finally, in June 1956, a seminar crucial for the further development of AI was held in Dartmouth, with the participation of leading theorists in this new field of research, including Simon. Simon (1991, p. 189) describes 1955 and 1956 as the most important years in his scientific life; and so they must have been if we consider that, besides his activity in the field of AI, two important articles on bounded rationality in economics were published in those years: Simon (1955, 1956).

Simon's research on 'human problem solving' became the core of a wide-ranging theoretical project in which AI, economics and cognitive psychology were closely intertwined. In addition, Simon connected all this to a long intellectual tradition (which included, among other things, the 'influence of formal logic', cybernetics, and Claude Shannon's information theory), whose merit was to focus on 'symbols and their manipulation in logical inference (if not in "reasoning") and decision' (Newell & Simon, 1972, p. 878).

3. AI and Human Thinking

Simon constantly emphasized the analogy between human thinking and intelligent computer behavior in light of his theory of symbolic systems, in that the computer and the 'human mind and brain' are members of the 'family of artifacts called […] physical symbol systems' (Simon, 1996, p. 21). As he maintained

> Like a modern digital computer's, Man's equipment for thinking is basically serial in organization. […] [T]here is much reason to think that the basic repertoire of processes in the two systems [human thought and computers] is quite similar. Man and computer can both recognize symbols (patterns), store symbols, copy symbols, compare symbols for identity, and output symbols. These processes seem to be the fundamental components of thinking as they are of computation. (Simon, 1976b, p. 430; see also Simon, 1964, and Simon & Shaw, 1958)

On this view, human and computer rationality are bounded because they exhibit limits as regards both computation and the manipulation of symbols (Simon, 1990, pp. 8, 17; see Newell & Simon, 1972, pp. 54–55). Simon applies BR to every sphere in which decision theory is implied, particularly economics, psychology and AI. These disciplines, in that they are involved in the analysis of information processing systems, are closely connected; it is therefore unsurprising that, for Simon, 'Economics is one of the sciences of the artificial' (Simon, 1976b, p. 441). As a consequence, economics is included within the AI approach, and this suggests that the study of BR should be considered from this perspective.

The AI perspective consists in considering the human thinking and problem-solving process in terms of an information processing system (IPS), that is, a system consisting of a set of memories, a processor, effectors and receptors (Newell & Simon, 1972, p. 20). Memories contain data, symbolized information and programs for information processing, so that the state of the system is determined by data and programs contained in these memories, with stimuli received by receptors. Symbols are the bases for both intelligent behavior and the comprehension of mind, since both computer and mind belong to 'physical symbol systems' (Simon, 1996, p. 21). Symbols are 'physical patterns' which can be components of entities called 'expressions', these being symbolic structures able to produce 'internal representations' of the external environment (Newell & Simon, 1981). More precisely, when an environment is coupled with a problem or goal, we have a 'task environment', which represents a set of facts about the world at a given moment such as, for example, the position of chess pieces on a chessboard. The subject represents the environment internally (internal representation), and the space in which his problem-solving activities take place is called 'problem space'. In addition to the current situation, problem space includes the possibilities for changing that situation, and it binds behavior in many ways: defining admissible moves (the set of possible sequential courses of action adopted to accomplish specific goals), specifying the aim, and interacting with short- and long-term memories.

Task environment, internal representation and problem space are crucial concepts, in that they determine the relation between the state of the world and individual subjectivity, since an objective world exists 'out there' but can be only represented subjectively (Newell & Simon, 1972, pp. 56, 824).

Given these premises, problem-solving implies that information is extracted from a problem space and used to seek a solution by means of heuristics. Consequently, the problem-solving procedure implies having a 'generator' of symbol structures, that is, a move generator for potential solutions, and a 'test' which evaluates these solutions. A problem will be solved if the generator produces a structure which satisfies the test (Newell & Simon, 1981, pp. 52−53). This means that we know what we want to do, but we do not know how to accomplish it. Therefore, man and computer are not omniscient, and they exhibit the same limitations as regards computation and rationality. From this perspective, intelligent systems can be treated as constituting a unique class of symbolic structures, in that they are considered on the basis of their behavior (and of their organization), regardless of their physical or biological properties (Simon, 1996, p. 21). In particular, Simon stresses that his approach does not concern homologies between a computer and the nervous system, but it describes how both intelligent behavior and its *general* properties emerge. As we shall see, this points out a difference between Simon's approach and the new ones which adopt BR and analyze how psychological attitudes, biological and neurobiological structures, skills, emotions etc highlight *specific* causes which can explain decision-making. Moreover, Simon's investigation into formal *general* characteristics of intelligent behavior implies that rationality is not a distinctive trait of man. Yet, this view raises problems when conscious rationality, however bounded, and awareness in human problem-solving activity enter the scene, because these notions imply deliberative acts, which by contrast cannot be performed by non-human intelligent systems. [3]

Whatever the case may be, when Simon focused on human BR, he specified that it was a conscious activity and that his concern was with 'intendedly rational behavior' (Newell & Simon, 1972, p. 55). In addition, he maintained that 'the first consequence of the principle of bounded rationality is that the *intended rationality* of an actor requires him to construct a simplified model of the real situation in order to deal with it' (Simon, 1957, p. 199; emphasis added), and he described problem-solving activity as a conscious endeavor to selectively explore and reduce a maze-environment to manageable proportions (Simon, 1996, pp. 54, 57).[4]

4. The 'Complex Algorithms of Thought'[5]

[3] In general, it is possible to maintain that an *intentional state* is not necessarily conscious, and that an adaptive mind can exhibit an intentional attitude toward an object or a state of the world without consciousness. For instance, the body's need for sugar prompts the eating of a cake. Therefore, the conscious desire, and the ensuing behaviour, derive from an unconscious intentional state of the body. Moreover, Simon hints that organisms adapt and consequently survive '*as if* they were rational', and this can be considered 'a *de facto* model of rationality' (Simon, 1983b, p. 35). Yet, in Simon's vocabulary, the word 'intention' (evidently used for human decision-makers) refers to deliberative acts of thought. Consequently, the distinction between rationality as a conscious or an unconscious process re-emerges.

[4] The fundamental role of intentionality has also been pointed out by two of Simon's close collaborators, who maintained that for Simon the structure of behaving systems is 'intendedly rational, subject to the limits of knowledge and computing power' (Newell, 1989, p. 411), and that 'The core notion of limited rationality is that individuals are intendedly rational' (March, 1994, p. 9).

[5] Simon (1976b, p. 442).

Intelligence (and in general IPS), as the work of symbol systems, exhibits logic and computational features (specifically limited computational capacities), and this leads to the notion of BR.

From this point of view, analogously to the concept of intelligence in computer science, BR owes an intellectual debt to formal logic. In fact, the origin of the concept according to which intelligent machines are symbol systems is explicitly ascribed to the program of Frege, Whitehead and Russell for formalizing logic. The debt to the 'logicians', considered as a 'fundamental contribution' to AI, is summarized as follows:

> The formalization of logic showed that symbols can be copied, compared, rearranged, and concatenated with just as much definiteness of process as boards can be sawed, planed, measured and glued. […] Formal logic, if it showed nothing else, showed that ideas—at least some ideas—could be represented by symbols, and these symbols could be altered in meaningful ways by precisely defined processes. (Newell & Simon, 1972, p. 877; see also Simon, 1991, pp. 192–194)

Newell & Simon specify that thought must not be confused with logic, because the former is not as rigorous as the latter; yet formal logic provided an important tool for conceiving symbol processing. From this point of view, they essentially include their work in the intellectual tradition whose characteristic was the conception that mathematics could derive from logic. In fact, in Russell's view, the complete formalization of mathematics by means of a symbolic logic system was possible. Moreover, formal logic determines a 'new operationality of symbols', although symbol manipulation refers to broader areas of knowledge with respect to deductive logic (Newell & Simon, 1972, p. 877). According to Newell & Simon, there are a number of thinkers who have contributed to linking, in different ways, formal logic, mathematics and symbol manipulation; they include Turing, Carnap, Church, Shannon, Lotka, Wiener, Pitts and McCulloch, among many others.

Simon attended the lectures delivered at the University of Chicago by Rudolf Carnap, whose theory had an important role in Simon's choice of the thesis project which culminated in *Administrative Behavior* (see

Simon 1991, p. 53). In fact, in the thesis, Simon initially intended to focus on the logical foundations of administrative science. As is well known, Carnap was among the founders of logical positivism, the philosophical school connected to the Vienna Circle in the early decades of the 20th century, which was inspired by the formal logic of Frege, Whitehead and Russell. Therefore, it is unsurprising that Simon maintained in *Administrative Behavior* that he accepted some conclusions of logical positivism (Simon, 1947, p. 45).

Crowther-Heyck (2005, p. 73) has discussed Carnap's influence on Simon, stressing how the 'intensive study of formal logic and positivist philosophy had a profound, lasting effect on [Simon's] thinking', although his criticism on excessive formalization impeded him from fully adhering to Carnap's view. Moreover, Simon followed Bridgman's operationalism, which like logical positivism was considered as a means to unify science, and he held that the 'operational definition of terms is crucial to making statements clear and testable' (ibid., 2005, p. 101). As a consequence, the operationalist approach re-appeared later in order 'to justify the use of computers to simulate human cognition' (ibid., p. 130; see Dasgupta, 2003).

Given these premises, several points can be emphasized:

(1) Rationality involves *physical symbol systems*. Consequently, rational processes can be represented in symbolic form, and as a sequence of unambiguous operations. In other words, non-conscious elements and tacit procedures for dealing with unexpected events are either not considered or are reduced to the logic of symbol manipulation. This implies that emotional or automatic responses to complex phenomena are not considered in light of their specific neurobiological or psychological dimension, in that Simon's AI examines only general and symbolic traits of intelligent behavior, irrespective of the physical features of subjects (men, computers etc).

(2) BR implies the existence of computational limits to the processing of information: agents possess limited information, and limited computational capacities.

(3) BR is instrumental in nature; it is invoked to solve problems, which often exhibit a logical form (for example, cryptarithmetic problems, proofs of theorems, etc).

(4) Studying how problem-solvers deal with a cognitive task implies that 'we are observing *intendedly rational behavior* or behavior of *limited rationality*' (Newell & Simon, 1972, p. 55; cf. Simon, 1957, p. 199). Therefore, the attention is focused on deliberate, conscious and intentional rationality, which as a matter of fact is limited. In this sense, Winograd & Flores (1986, p. 22) maintain, Simon does not contest the 'rationalistic tradition', but only the version that implies perfect knowledge, perfect foresight, and optimitization criteria (see also Gardner, 1987, p. 361; Crowther-Heyck, 2005, p. 60).

In short, decision-making emerges as a procedure connoted by the symbolic and deliberate activity of problem-solving, and can be explained in terms of information processing, where the latter implies performing a number of definite, unambiguous operations (Simon & Shaw, 1958, p. 6). As a consequence, part of Simon's monumental project was devoted to showing how phenomena such as intuition, insight and creativity (and, in general, perception and emotions) are neither mysterious nor obscure phenomena. By contrast, intentional BR and intuition can be considered the extreme poles of rational behavior, in that they are respectively conscious and unconscious activities of the decision-maker, whose explanation is based on the same fundamental information processing mechanism.

Logical and intuitive thought can be easily unified; therefore phenomena like intuition, insight, creativity, perception and emotion can be explained by means of computer programs (Simon, 1997a, pp. 174−175). Intuition appears when 'someone solves a problem or answers a question rather suddenly […] without being able to give an account of how the solution or answer was finally attained' (Simon, 1987, p. 482). As a consequence, it differs from problem-solving, which involves intentional heuristic search. Yet intuition shares with this latter a basic feature: it processes information in symbolic form. In fact, intuition is an act of 'recognition'. We recognize our past experience, i.e., we refer to the long-term memory where the knowledge of this experience is located (Simon, 1997a, p. 179). As recognition of 'stored information', intuition has a symbolized and defined form. Moreover, intuition is guided by rules, and it explains how chess grand masters, managers and other experts are able to deal with situations requiring rapid decisions (Simon, 1983a, p. 4570; March & Simon, 1958, pp. 10−13). Thus intuition entails a kind of rationality which, although limited, is instrumental, and is not connoted by ambiguity or vagueness because it processes well-defined information encapsulated in unambiguous symbols, as in problem-solving procedures. Finally, intuition as recognition implies BR. In fact,

> A major strategy for achieving intelligent adaptation with bounded rationality is to store knowledge and search heuristics in a richly indexed long-term memory in order to reduce computational requirements of problems. […] When recognition does not suffice, because a great space of possibilities must be explored, they resort to highly selective search, guided by rich stores of heuristics. (Simon, 1990, p. 17)[6]

5. Bounded Rationality Revisited?

Many scholars have acknowledged their intellectual debt to Simon, and applied the notion of BR in their work, yet only *some of them have also received scientific recognition from Simon*, as in the case of March, Nelson, Winter, Kahneman, Tversky, and Gigerenzer.[7] For these reasons, I refer to these authors as representative cases of non-standard approaches influenced by Simon. Nevertheless, in contemporary studies BR seems to undergo a shift with respect to Simon's perspective. In particular, these theories on BR prevalently describe cognitive (rational) processes as events weakly connoted in terms of computation, problem-solving ability and intentionality, whereas in Simon's view computation (as limited symbol processing), problem-solving processes and intentionality have a major role. Moreover, in these

[6] Yet, a potential problem arises. Intuition is a process which permits individuals to respond rapidly when they encounter situations requiring an immediate reaction. If there is insufficient time to evaluate alternatives, problem-solving processes cannot be applied. But, if there is no past experience upon which to draw (since the situation requiring urgent response is genuinely new), and the recognition process cannot be activated, what guides individuals' reactions?

[7] In general, Simon's scientific recognition of these scholars can be found in several articles (see Simon, 1982; 1997b). In regard to James March, suffice it to point out that he was one of Simon's collaborators and co-authored *Organization*s with him. As regards Gigerenzer and Kahneman & Tversky, see respectively Gigerenzer (2004), and footnote 12 in this paper.

contemporary approaches, unconscious, adaptive, automatic and emotional processes are considered important elements for the elaboration of judgments and decisions, and this induces re-examination of how individuals act in solving problems, given their cognitive limits. Finally, it is difficult to find in these works the perspective which relates AI and BR.

In this sense, as often happens with important concepts in science, their use has gradually led to changes with respect to the original meaning. Nevertheless, the interpretation put forward here in explanation of this process is that Simon's original theory of BR cannot be easily used in recent research which deals with BR because of some of the technical and conceptual characteristics derived from Simon's AI (although its basic assumptions are generally accepted). As a consequence, a *tacit theoretical challenge* is in fact emerging: the redefinition of a more comprehensive notion of BR, in which psychological, neurobiological, and emotional factors are appropriately included in a theoretical explanation. The consequence is a gradual departure from Simon's perspective. Much of the recent debate centers on criticism of standard explanations of rationality in economics and social sciences, yet a less evident implication is that new (and differentiated) perspectives on BR are shaped by this process. The following sections provide some examples.

5.1. James March: Problematic Aspects

A rereading of the concept of BR appears in the work of James March, Simon's well-known collaborator, who pointed out that new directions have emerged from the original Simonian context (March, 1978, p. 591). The focus of his controversy concerns the neoclassical approach, and in particular the assumption that preferences are coherent, stable, exogenous and unambiguous. On the contrary, March maintains, preferences are ambiguous and inconsistent.

> Choices are often made without respect to tastes. Human decisionmakers routinely ignore their own, fully conscious, preferences in making decisions. They follow rules, traditions, hunches, and the advice or actions of others. Tastes change over time in such a way that predicting future tastes is often difficult. Tastes are inconsistent. Individuals and organizations are aware of the extent to which some of their preferences conflict with other of their preferences; yet they do nothing to resolve those inconsistencies. (March, 1978, p. 596)

Preferences are ambivalent (at the same time, something is desired and not desired), and their inconsistency requires reversal of the perspective according to which a coherent course of action derives from preferences. In fact, 'We construct our preferences. We choose preferences and actions jointly, in part, to discover—or construct—new preferences that are currently unknown. We deliberately specify our objectives in vague terms to develop an understanding of what we might like to become. We elaborate our tastes as interpretations of our behavior' (March, 1978, p. 596). Therefore, preferences are discovered by means of decision-making processes, by experiencing our choices, since individuals often elaborate an interpretation of what they are doing in the course of their actions (March, 1988, 1994). This perspective points out the role of BR: unstable, inconsistent and imprecise preferences and goals at least partially assume this form 'because human abilities limit preference orderliness' (March, 1987, p. 598).

Yet, Simon's concept of BR seems to differ in some respects from the one generally delineated by March. In particular, Simon's BR does not include such notions as 'ambiguity' and 'inconsistency' in March's structural sense. Information, which Simonian agents process, is not ambiguous; it is simply gathered and manipulated. As a consequence, rational courses of action, such as problem-solving activities, are not guided by a fuzzy set of information from which the behavior takes shape. Generators of solutions and tests, which functionally connote every problem- solver, work according to logical schemes. Task environment, heuristics and goals (especially if related to well-structured and closed problems, like proving a theorem, achieving checkmate, solving a crossword puzzle, etc), in turn, are not vague or incoherent, and search processes have a defined structure. Computational limits in information processing and in intentional search do not imply ambiguity: the problem space, internal representation, generator of solutions and test are jointly used to identify and solve well-formulated problems, although optimal procedures are not applied. All this seems to define a framework which differs from that of March, where 'Human beings have unstable, inconsistent, incompletely evoked, and imprecise goals' (March, 1987, p. 598), and both the world and the 'self' are ambiguous. In fact, part of March's work is focused on the processes by which individual identities take shape, since the agent's self is a collection of 'multiple identities' which are often 'incompletely integrated'. From this perspective, the self—together with values, goals, wants etc—is discovered, i.e., it is the *result* of a decisional, social, process; it is not the *starting point* from which values, goals and wants are given (March, 1994, pp. 68, 191).

This view requires a distinction to be drawn between two different, although interconnected, logics of action (March & Simon, 1958, pp. 8−9): '*Pure rationality* and *limited rationality* share a *common perspective* seeing decisions as based on evaluation of alternatives in terms of their consequences for preferences. This logic of consequences *can be contrasted* with a logic of appropriateness by which actions are matched to situations by means of rules organized into identities' (March, 1994, p. 57; emphasis added). The logic of consequences (shared by pure and bounded rationality) implies analysis, calculation and a heuristic search among alternatives, while the logic of appropriateness entails that individuals who seek to discover their self and goals are rule-followers; that is, they follow 'rules or procedures that they see as appropriate to the situation in which they find themselves' (March, 1994, p. 57).

In conclusion, many elements of March's work are a direct consequence of Simon's view. Nonetheless, it is possible to note a theoretical shift (not an opposition) that, in March's case, leads to relaxation of some formal elements of BR: in one case, the self (or more precisely the problem solver) is well-structured and stable, even if computationally limited; in the other, the self does not possess a sound structure, and this is the most important limit.[8]

5.2. Nelson & Winter: Bounded Rationality and Tacit Knowledge

Nelson & Winter, in *An Evolutionary Theory of Economic Change* (1982), identify Michael Polanyi's notion of 'tacit knowing' and Simon's BR as basic assumptions for their own theoretical perspective. The 'strong' influence of Polanyi is acknowledged, especially in chapters 4 and 5; Simon's concepts of BR and of 'satisficing' are explicitly embodied in their theory. At first sight, the connection between these notions seems evident: tacit knowledge can be conceived as a limit to rationality because it means that rationality cannot be perfectly articulated, and consequently cannot be used to maximize utility. Yet some scholars have maintained that Simon's and Polanyi's epistemologies are incompatible, that tacit knowledge and BR do not imply each other and that tacit knowledge cannot be reduced to information processing in Simon's sense (Nightingale, 2003; Foss, 2003). In particular, Nightingale observes, Newell and Simon would argue that unarticulated knowledge becomes codified[9] (i.e., processed as information) in problem-solving heuristics and routines. Yet this view conflicts with the perspective according to which neurological, automatic and unconscious processes involved in problem solving and decision-making cannot be articulated.[10]

[8] In the present interpretation, Simon's model of rationality implies a unitary self. Therefore my conclusions differ from those of Mirowski (2002, p. 479), who stresses the dissolution of the notion of individual self in Simon's work.

[9] Tacitness and codification are not always opposed: for a critical reconsideration of these terms see Cowan et al. (2000).

[10] As a consequence, Simon's recourse to 'protocol analysis', which implicitly assumes that '*things that could be articulated* […] *were the only things involved in problem solving*' (Nightingale, 2003, p. 162, n. 29), is useless. It has to be pointed out that, in accordance with Simon's AI, 'protocol analysis' ('verbal protocol') implies that subjects verbally explain the steps that led them to a problem's solution. In this way, a correspondence is assumed between the conscious course followed to produce answers and the thought process.

Simon's denial of tacit knowledge emerges, in general terms, from his theory, and in particular from his criticism of Polanyi's approach. The theme can be dealt with by starting from Simon's and Polanyi's interpretations of Plato's *Meno* paradox, namely the assumption that we can discover or recognize the answer to a problem only if we already know the answer (Simon, 1962, p. 479). Polanyi deals with the paradox by resorting to tacit knowledge, since, in his view, the paradox shows that knowledge is not entirely explicit, and that the 'kind of tacit knowledge that solves the paradox of the *Meno* consists in the intimation of something hidden, which we may yet discover' (Polanyi, 1967, pp. 22−23). Simon criticized this interpretation, and the 'false' premise that 'if you know what you are looking for, there is no problem'. He specified that a problem can be posed for a formal system (containing countable set of symbols, and 'expressions') by generating a subset of 'well-formed-formulas'. Therefore 'our ability to know what we are looking for does *not* depend upon our having an effective procedure for finding it: we need only an effective procedure for *testing* candidates' (Simon, 1976a, p. 148). In short, problem-solving activity and heuristic search organized in terms of move generators and tests (i.e, devising possible solutions and then testing each one to control whether it satisfies the solution conditions of the problem) resolve the paradox without invoking tacit knowledge. This also implies that information, given its symbolic form, is processed and appropriately codified, and therefore is not vaguely tacit: simplified heuristics and searches do not rely on tacit, cognitive, components. Moreover, symbolic information can become explicit, while in Polanyi's view, tacit parts of knowledge never assume an explicit form, since human behaviors and choices can be fulfilled only if a certain amount of knowledge remains tacit.

Despite these epistemological divergences, Nelson & Winter link the two approaches, providing a peculiar explanation of rationality, which is described *at the same time as bounded and tacit*. They thus modify Simon's original perspective, and the result of this fruitful cross-fertilization is a notion of bounded rationality different from Simon's.

In Nelson & Winter's book, tacit knowing constitutes an argument against the standard notion of technological knowledge as articulated and explicit. Analysis of organizational routines in terms of skills clarifies their hypothesis, since tacit knowledge underlies skillful performances. In fact, skills are deployed without full awareness of the details of performances, and behaviors and choices are selected automatically. In short, knowledge embodied in skills is largely non-articulable, and the communication of explicit knowledge does not imply possession of the skill, since mere instructions cannot transfer ability.[11] In other words, a novelty in Nelson & Winter's approach is that tacit knowledge, as incorporated in skills and capabilities of both individuals and organizations, is *explicitly* connected with limits to rationality because, contrary to the notion of perfect knowledge, knowledge is not entirely articulated and well defined. Therefore 'there is […] a tradeoff between capability and deliberate choice, a tradeoff imposed ultimately by the fact that rationality is bounded' (Nelson & Winter, 1982, p. 85). More precisely, rationality is limited in that skills can take the place of conscious deliberation, work tacitly and automatically, and often select available options, so that deliberate choices play a restricted role. Skills are complex entities which cannot be intentionally manipulated. Therefore, 'there is inevitably some ambiguity regarding the scope of a skill'; such ambiguity can be reduced, yet part of it 'is simply irremediable' (ibid., pp. 88−89).

Given these premises, the notion of tacit knowledge incorporates and explains BR, *conferring a new dimension on it*: the bounds of rationality depend on the tacitness of knowledge. As a consequence, rationality is connoted by tacit and automatic capabilities, and this shows the limits of the conscious, *intentional*, mind. All this emphasizes the different perspective of Nelson & Winter: they mainly analyze tacit and *unintentional* rationality, whereas Simon's aim is to define the limits of *intentional* rationality.

In Simon's view, simplification of a problem permits mastery of a situation in an essentially *conscious* form, without resorting to special skills based on tacit knowledge. This activity is imposed by cognitive limits when the task is perceived as too complex, and the answer consists in exploiting intentional rationality, however limited, by means of a structured strategy (heuristic search, and reduction of problems into nearly independent parts), and it does not consist in relying on tacit elements of reason.

[11] Tacitness is also a 'matter of degree', since 'The same knowledge […] is more tacit for some people than for others', and 'much operational knowledge remains tacit because it cannot be articulated fast enough' (Nelson & Winter, 1982, pp. 78, 81−82).

In Nelson & Winter's work, BR enters the scene when the authors oppose orthodox and evolutionary theories, comparing the corresponding kinds of knowledge. In this way, BR is defined essentially in terms of a *lack of explicit knowledge*. This is a peculiar definition, because tacit knowledge and skills constitute a limit and a resource at the same time. On the one hand, tacit knowledge is seen as a boundary; on the other, it explains how non-perfect rationality actually works. In this latter sense, since orthodox models of rationality are rejected, tacit knowledge simply helps to describe individuals' and organizations' behaviors and choices. Skills (and tacit knowledge) *per se* do not denote cognitive limits; rather, they enable delineation of a distinctive functioning of rationality which is not easily made compatible with Simon's notion of information processing. Information processing, in Simon's view, involves codification work connoted by limits in the manipulation of well-defined symbols; but, in this perspective, the impossibility of articulation and ambiguity of skills cannot be dealt with. In conclusion, Simon rejects the explanation based on tacit knowledge and provides a different epistemology on which to ground the notion of BR. By contrast, Nelson & Winter define BR on the basis of tacit knowledge. This modification is seen by these scholars as a complement to Simon's theory; yet it also implies a change in the original meaning of BR.

5.3. Kahneman's Perspective[12]

In a recent article, Daniel Kahneman has summarized the most important phases of his research conducted with Amos Tversky over several decades. The 'guiding ideas' are that 'most judgments and most choices are

[12] Some scholars maintain that Kahneman and Tversky have not abandoned the neoclassical model, but rather use it as a benchmark, in order to show its empirical inadequacy (cf. Camerer, 1999; Sent, 2004; Motterlini & Guala, 2005). Others point out that Kahneman's and Tversky's studies have shown the failure of the notion of rationality associated with expected-utility theory, and 'have advanced some alternative hypotheses of behaviour' which have led to BR (Arrow, 1996, p. xv). Simon himself stated that 'Some of the most dramatic and convincing empirical refutations of the [subjective expected utility] theory have been reported by D. Kahneman and A. Tversky' (1979, p. 506). In turn, Tversky & Kahneman (1987, pp. 88–89) recognize that the results of their analysis 'are consistent with the conception of bounded rationality originally presented by Herbert Simon'.

made intuitively, [and] that the rules that govern intuition are generally similar to the rules of perception'. He adds that intuition acts '*without conscious search or computation*' (Kahneman, 2003, p. 1450; emphasis added) and distinguishes between two alternative modes of thinking and deciding: reasoning and intuition, because—as Gilovich and Griffin (2002, p. 3) maintain—intuitive judgments and rational models are 'categorically different in kind'. Moreover, intuition often prevails over reasoning. By way of contrast, as shown in Section 4, according to Simon intuition and rational thought constitute a *unique* cognitive mechanism in which computation is fundamental, even if it is characterized as a limited capacity. The only difference, Simon maintained, is that intuition appears because 'stored information' is available and searching is not necessary, while normal problem-solving processes require a search for information. In both cases, the process consists in the manipulation of unambiguous symbols contained in memories.

In particular, Kahneman, following Stanovich & West (2002), affirms that perception and intuition are collected in 'System 1', whose operations are fast, parallel, automatic, effortless, associative, slow-learning and emotional, while reasoning belongs to 'System 2', which is characterized by slow, serial, controlled, effortful, rule-governed, flexible and neutral operations (see also Zajonc, 1980; Kahneman & Frederick, 2002; Gilovich & Griffin, 2002; Sloman, 2002; Camerer, Loewenstein & Prelec, 2005).

The two systems do not necessarily always operate in a complementary fashion (monitoring intuitive answers by means of deliberate reasoning), for they may conflict and produce contradictory responses (Sloman, 2002). A distinctive outcome of this opposition arises when System 1 and System 2 are active concurrently, so that 'automatic and controlled cognitive operations *compete* for the control of overt responses, and […] deliberate judgments are likely to remain anchored on initial impressions' (Kahneman & Frederick, 2002, pp. 51−52; emphasis added). In other words, intuition (and in general System 1) is able to condition deliberate reasoning (System 2) performances, and this implies that the former performs a role which does not characterize intuition in Simon's sense. Intuition, according to Simon, neither contradicts nor competes with overt reasoning; on the contrary, it supports rational deliberations, as in the example of chess masters playing a number of games of chess simultaneously.

All this focuses the attention on both the role of heuristics and the relations between decision-making and emotions. Simonian agents are problem-solvers, and they adopt deliberate strategies (heuristics) to solve problems. Yet, as regards the approach of Kahneman & Tversky, it has been pointed out that 'deliberate choice heuristics differ substantially from the judgmental heuristics of [their] "heuristics and biases" research program, which are largely based on impressions that occur automatically and independently of any explicit judgmental goal' (Frederick, 2002, p. 549). The distinction between 'automatic' judgmental heuristics and 'deliberate' choice heuristics reflects the difference between Systems 1 and 2. Rapid and intuitive judgments, for example, based on assessments of resemblance to judge probability ('representativeness heuristics'), do not involve problem-solving processes.

The focal point is that, from this perspective, in many circumstances of everyday life people do not make decisions in terms of problem-solving processes; they do not look for 'satisficing' alternatives, given a well-defined, though not permanently fixed, 'aspiration level'. Therefore there are no sequential processes in which problems are broken down into smaller components. In fact, intuition, emotions and automatic judgments intervene *before* the formulation of any rationally limited explicit assessment. In short, *pre-analytic criteria*, connoted by psychological attitudes (explainable in terms of framing effects, status quo bias, loss aversion, endowment effects, etc), affect rational decisions and determine choices without previous rational evaluations.[13] Also intuition, which works in analogous manner to the perception system, can be observed in this perspective, since 'The central characteristic of agents is not that they reason poorly but that *they*

[13] See Kahneman, Knetsch, & Thaler (1991), where the authors explain the relations among the endowment effect, status quo bias and loss aversion. The perspective of reason-based choice implies that the decision-maker possesses 'good reasons' for and against each option: 'Decisions […] are often reached by focusing on reasons that justify the selection of one option over another' (Shafir, Simonson, & Tversky, 1993, p. 34).

often act intuitively. And the *behavior of these agents is not guided by what they are able to compute, but what they happen to see at a given moment*' (Kahneman, 2003, p. 1469; emphasis added). Sometimes intuition (not conceived as stored symbolic information), and not computation, provides good reasons for making choices. More specifically, one property of intuitive judgments is 'accessibility': that is, the ease with which some mental content or thoughts come to mind, to which the notion of substitution is correlated: 'judgment heuristics are applied in a wide variety of domains and share a common process of *attribute substitution* in which difficult judgments are made by substituting conceptually or semantically related assessments that are simpler and more readily accessible' (Kahneman & Frederick, 2005, p. 287). This perspective, Kahneman says, makes it possible to re-examine framing and prospect theories, and conceive them in a more unitary context. As a consequence, intuition and accessibility to certain thoughts matter, and even if they produce judgmental biases, they show how some decision-making processes arise from pre-analytical mechanisms.

Summarizing, in many circumstances agents do not act as problem solvers when they are faced by decisional problems. Kahneman and his collaborators do not deny the role of analytic processes, intentionally designed to simplify choices; rather, they point out that this is only one side of the decisional process, and that the continuous overlap and interaction of the two systems of mind (in general, with the prevalence of one of them) adequately describes the cognitive mechanism and its performances. In short, Simon's idea of mind and of rationality as unique, unambiguous, and instrumental mechanisms (however limited with respect to computation and information processing), differs from the two-systems theory. The latter can be interpreted as an *implicit effort to redefine the nature and limits of rationality,* that is, to borrow Kahneman's expression, provide new 'maps of bounded rationality'. Of course, it is possible specifically to analyze the properties of System 2 (deliberate reasoning); yet it seems hard to consider their functioning in isolation, even when System 2 prevails over System 1, since the two systems are necessarily integrated with different degrees of reciprocal influence. System 1 affects System 2, and vice-versa. As a consequence, System 2 cannot be merely identified with Simon's model of intentional mind (and with standard rationality models even less), because, in a certain sense, System 2 can be correctly explained *only if* System 1 functioning is presupposed, and System 2 limits are considered somehow as derivations from the interaction with System 1. As neurobiologists maintain, the mind is a whole, whose parts are closely integrated, and this issue probably is a result of how it evolved through natural selection. The two systems are complementary and cannot be dealt with as independent mechanisms, although they perform different functions. It is therefore unsurprising that Kahneman includes 'the large role of System 1' among the 'core ideas' of his essay on BR (Kahneman, 2003, p. 1469); neglecting this aspect means elaborating another model of mind.

Finally we must examine the role of emotions in understanding how judgments and decisions emerge unconsciously: 'There is now compelling evidence for the proposition that every stimulus evokes an affective evaluation, and that this evaluation can occur outside of awareness' (Kahneman & Frederick, 2002, p. 56). Analysis of emotions is a turning point in decisional theory, and Kahneman (2003), referring to a huge body of literature, points out, among other things, the concept of 'affect heuristic' to show the pervasive role of emotions in guiding judgments and decisions. According to Slovic *et al.* (2002), affect and emotional arousal influence preferences; they do not require cognitive appraisal; they generate responses which occur rapidly and automatically; and they can be explained in evolutionary terms.

In particular, Kahneman maintains, preferences must be understood in light of the psychology of emotions. Once again, attention focuses on the conflict between the Systems 1 and 2, since emotions can control preferences, although these latter may be inconsistent with the preferences of reflective reasoning (Kahneman, 2003, p. 1463). Yet, in general, the pervasive role of emotions enables Kahneman to rethink large part of his research with Tversky. For example, emotions are invoked in *prospect theory* because 'utility cannot be divorced from emotion, and emotions are triggered by changes' (ibid., p. 1457).[14] Therefore emotions play a wider role than in Simon's theory, where they are described in symbolic terms as mechanisms able to help agents select and focus conscious attention on specific objects, and consequently define priorities in the search associated with problem-solving activity. [15]

In addition, the predictions of prospect theory have received support from recent neurobiological research (in which Kahneman has taken part) based on functional magnetic resonance imaging (Breiter *et al.*, 2001). This technique has been used by De Martino and colleagues, who maintain that the framing effect is associated with amygdala activity and suggest that emotional system has 'a key role in mediating decision biases'. The experiment they conducted supports the hypothesis of two neural systems (which sometimes operate at cross-purposes) that perform diverse functions, but are also robustly correlated (De Martino *et al.*, 2006, p. 687; see Sanfey *et al.*, 2006, pp. 111−112). Kahneman's comment is telling: 'The results could hardly be more elegant' (quoted in Miller, 2006, p. 600).

In short, Kahneman & Tversky's studies received further confirmation on neurobiological grounds, not on those of Simon's AI—whose assumption was that decision-making should be studied without direct reference to the human biological structure (Simon, 1964, pp. 76−77) —and they sometimes give rise to opposite conclusions with respect to AI. This provides an argument in support of the thesis that Kahneman's and Simon's theories are basically different models.

Antonio Damasio (1994) has provided an important account of the role of emotions in decision-making. His theory of the function of somatic markers leads to the conclusion that the latter increase efficiency in decision processes. He also suggests with his collaborators that non-conscious

[14] In fact, prospect theory affirms that perception of utility depends not on states of wealth or welfare but on changes (gains and losses) relative to a neutral reference point.

[15] In Simon's view, emotion is an 'interruption mechanism', which 'allows the processor to respond to urgent needs in real time'. Therefore, it selects objects so that we can focus our attention on them. Given this framework, Simon affirms that these mechanisms have been basically embodied in 'the current information-processing theories of human cognition' (Simon, 1967, p. 38; see Simon, 1983b, pp. 29−30).

biases guide behavior prior to conscious knowledge, and that 'Without the help of such biases, overt knowledge may be insufficient to ensure advantageous behavior' (Bechara *et al.*, 1997, p. 1293). Non-conscious biases assist reasoning in a 'cooperative manner' facilitating in a more efficient way the process by which conscious decision are made. From the point of view of Systems 1 and 2, a strong connection between the two mechanisms reappears, now in terms of cooperation, and it highlights how BR is not only a product of a limited analytical mind, but of intuitive, emotional and non-conscious mind as well.

In conclusion, Kahneman's 'maps of bounded rationality' seem to imply a kind of rationality weaker than Simon's. Intuition is not a peculiar mechanism connoted by symbolic intelligence, and explained by referring to an AI context; moreover, it often prevails over computational reasoning, showing the 'unexpected weakness of System 2' (Kahneman & Frederick, 2002).

5.4. Ecological Rationality

Gerd Gigerenzer & Reinhard Selten (2002) also start from Simon's vision of BR, and maintain that their work is an elaboration of his theory. When Simon introduced the notion of 'satisficing', he pointed out two sides of BR: one cognitive and one ecological, which in turn emphasized that minds are adapted to the real world environment. The notions of 'ecological rationality' and 'adaptive toolbox' have been central elements of the research program of Gigerenzer and his group (Gigerenzer *et al.*, 1999). This perspective is summarized as follows by Gigerenzer & Selten (2002, p. 9):

> The concept of an adaptive toolbox, as we see it, has the following characteristics: First, it refers to a collection of rules or heuristics rather than to a general-purpose decision-making algorithm […]. Second, these heuristics are fast, frugal, and computationally cheap rather than consistent, coherent, and general. Third, these heuristics are adapted to particular environments, past or present, physical or social. This 'ecological rationality'— the match between the structure of a heuristic and the structures of an environment—

allows for the possibility that heuristics can be fast, frugal and accurate all at the same time by exploiting the structure of information in natural environments. Fourth, the bundle of heuristics in the adaptive toolbox is orchestrated by some mechanism reflecting the importance of conflicting motivation and goals.

Ecological rationality, as a new version of BR, exhibits some shifts with respect to the original model. In fact, it is centered less on '*intendedly rational behavior*' and more on a rationality which emerges as a strategy adapted to a changing environment scarcely characterized by analytical judgments and choices. In one case individuals are considered to be intentional problem-solvers whose explicit rationality is limited; in the other, individuals are adaptive problem-solvers with scant awareness as regards their responses to the environment. As a consequence, contrary to the idea that intelligence is a deliberate activity guided by logic rules, Gigerenzer (2007) maintains that it is based on unconscious processes—specifically, 'gut feelings or intuitions'—based on simple heuristics.

In particular, (1) heuristics are fast and frugal, because they do not involve much computation and 'only search for some of the available information' (Gigerenzer & Todd, 1999, p. 4); and (2) they exploit environmental structure to make adaptive decisions, and lead to accurate and useful inferences. In short, simple heuristics use 'minimal time and information', and this happens both in basic ecological situations (as in the prey–predator case), in which rapid decisions confer an adaptive advantage, and in complex social circumstances. More precisely, 'fast and frugal' heuristics imply a sequential search; they are characterized by recognition and ignorance; they are based on the first encountered cue; they combine a small number of cues to make categorical decisions; and they stop a sequential search after encountering only a small number of alternatives (Todd, 2002, pp. 55–56).[16]

[16] These characteristics are grouped into four main classes: 'ignorance-based decision making', 'one-reason decision making', 'elimination heuristics', and 'satisficing heuristics for sequential search'. In the first heuristic, ignorance is specifically identified as an essential element for making decisions. The second heuristic relies 'on just a single cue to make a decision'; the third too is based on little information, while the fourth is closely connected to Simon's approach (Todd & Gigerenzer, 2003, pp. 149–150).

Limits of information (for example, partial ignorance, as a lack of information, to which the 'recognition heuristic' refers) do not hamper decisional processes; on the contrary, they help them.[17] More in general: 'There is a point where too much information and too much information processing can hurt' (Gigerenzer & Todd, 1999, p. 21). Also *limits of computation* seem to be well balanced by other mechanisms which efficaciously, and unconsciously, guide choice processes. In particular, the recognition heuristic 'is the simplest of these adaptive tools. It uses a *capacity* that *evolution* has shaped over millions of years, recognition, to allow organisms to benefit from their own ignorance' (Goldstein & Gigerenzer, 1999, p. 57; emphasis added). A basic evolutionary mechanism intervenes *before deliberate reasoning* to guide the search for alternatives with minimal time and information. It has therefore been maintained that the recognition heuristic has an automatic component, although its assessments are subsequently endorsed by a deliberate strategy (Kahneman & Frederick, 2002, pp. 59–60). Not only in this case, but in general, the success of intuitions depends on evolutionary processes of the brain, and on 'the intelligence of the unconscious', that is, 'the ability to know without thinking which rule to rely on in which situation' (Gigerenzer, 2007, p. 229). Therefore, with respect to Simon, (1) intuition is not 'stored information' describable in symbolic terms, and (2) unconscious heuristics play a greater role in solving problems and influencing choices than do explicit, simple, heuristics, although both are rules of thumb.

Finally, in Simon's concept of BR, cognitive limits are not advantageous for human beings. By contrast, 'ecological rationality' delineates a perspective in which cognitive limits have some beneficial consequences; this once more suggests 'that simple heuristics had a selective advantage over more complex cognitive strategies' (Todd & Gigerenzer, 2003, p. 161).

---

[17] 'The recognition heuristic can *only* be applied when one of the two objects is not recognized, that is, under partial ignorance' (Goldstein & Gigerenzer, 1999, p. 41). In this sense, 'A decision maker who recognizes the name of one object (e.g., a city) but not of a second one (e.g., another city) can use that pattern of recognition to infer that the recognized object has a higher value on a criterion (e.g., population)' (Todd & Gigerenzer, 2003, p. 150). We make inferences from unknown aspects of the world and, from this perspective, ignorance is beneficial.

Gigerenzer and his collaborators maintain that just two forms of BR exist: 'satisficing heuristics', and 'fast and frugal heuristics'. Both continuity and differences, they note, exist between these approaches (Gigerenzer, 2008, pp. 80−91). On the one hand, Simon's heritage is widely recognized; on the other, some differences emerge as normal by-products of this autonomous research program. Yet a further theoretical shift can once again be discerned. As for Kahneman and Tversky, bounded (conscious) rationality is connoted in a weaker sense with respect to Simon's view. Computation and information have a limited role, since the most important bounds of rationality are not mental (internal) factors, as such; rather, they stem from environmental (external) pressures, and from how environmental forces have selected simple heuristics for making decisions (Todd, 2002, p. 52; see Gigerenzer, 2002, p. 39). If this is so, *internal, cognitive, limits are not so important* if adaptive processes allow the use of accurate and useful inferences: 'A computationally simple strategy that uses only some of the available information can be more robust, making more accurate predictions for new data, than a computationally complex, information-guzzling strategy that overfits' (Gigerenzer & Todd, 1999, p. 20).

In a certain sense, computational limits produce a counterintuitive effect, since 'computationally simple strategies' are more efficient than 'computationally complex' ones. But, in Simon's view, the more such computational limits are reduced, the more rational will be agents' decision-making, and the better will be their strategies for solving problems. In short, intentional rationality is weak (more so than in Simon's view), since the success of a course of action does not essentially depend on the ability to compute and process information, but relies on adaptation between the structure of heuristics and the structure of an environment. If this adaptation comes about, a successful behavioral performance will ensue. Gigerenzer's individuals adopt adaptive heuristics which enable them to make 'accurate inferences', therefore intentional rationality (in Simon's sense) is weak, while ecological rationality is strong.[18]

[18] This view, in turn, is linked to the criticisms leveled against Kahneman and Tversky. In Gigerenzer's view, simple heuristics lead to reasonable decisions and accurate inferences, while in Kahneman & Tversky's approach heuristics lead to 'systematic errors'. The debate among these scholars on this issue cannot be discussed here. See Gigerenzer (1991); Kahneman & Tversky (1996); Gigerenzer (1996).

Finally, the notion of the 'adaptive toolbox' refers to a set of concepts which seem far from Simon's original perspective; consequently, cognitive processes are not treated by recourse to symbol systems, as in the AI approach, probably because 'the mind is a biological rather than a logical entity' (Gigerenzer & Todd, 1999, p. 17), while the information processing approach implies that it is possible to study the mind 'independently of the details of the biological mechanism' (Simon, 1964, p. 77).[19] AI, Gigerenzer (2007, pp. 102, 55) maintains, is concerned with 'disembodied intelligence', and 'This suggests that the nature of thinking is logical, not psychological'. Hence the way in which an evolved human brain solves problems is essentially different from that of a computer. Nonetheless, as with Kahneman and Tversky, fast and frugal heuristics can be seen as a distinctive complement to, or a development of, Simon's work.

6. Conclusions

The simple assumption from which we started is that Simon's notion of BR is closely connected to scientific research in AI. This is not a novel observation; nonetheless it has certain, generally neglected, theoretical consequences, the most important of which is the conception of BR as a highly structured concept, in many respects characterized in terms of physical symbol processing. 'Internal representation', 'task environment' and 'problem space' constitute appropriate concepts by means of which decision-making and problem-solving take form within the perspective of symbol logic. In addition, 'generators' (of symbol structures) for potential solutions, and 'tests' which evaluate them, explain how heuristics operate in terms of sequential processes. This shows that agents do not know how to resolve a problem optimally (i.e., they are rationally bounded), but by contrast explore, step by step, only a few possible candidates, having a procedure with which to test them.

[19] This does not impede applying Simon's view to the evolutionary, Darwinian, model. From this perspective, genetic mutation and natural selection are respectively the 'generator' of variety and of new living forms, and the 'test' which selects those forms well fitted to the environment (Simon, 1996, p. 46).

Given this premise, it has been argued that formal logic is able to influence AI by providing tools with which to conceive both symbol processing and symbol operationality. This influence was coherent with Simon's endeavor to provide rigorous bases for the social sciences, a perspective which connects his scientific work with the *Zeitgeist* of the mid-20th century. Therefore, this heritage matters, as Simon himself recognized.

Simon sought to explain phenomena such as intuitions, insights, perceptions and emotions by means of AI. The unitary vision of bounded rationality emerges from this perspective. Rationality works in essentially the same way in any circumstance, and it exhibits a homogeneous structure when it functions by using either problem-solving procedures or intuition. Moreover, rationality is connoted by a set of well-defined symbolic operations; it does not deal with ambiguous information; it is distinguished by search processes which can take the form of explicit verbal protocols; it is instrumental because it has to solve problems; it is computational; and it implies intentionality when it refers to human beings. All these features are consistent with the most important one: rationality is bounded.

The second part of the paper has suggested that some shifts in the conceptualization of BR are apparent in certain theories which share Simon's criticism of standard economic theory, although they declare that they assume his view on BR. These works are not homogeneous: James March deals with the pervasive ambiguity of preferences, and discusses the weak structure of decisional processes; Nelson & Winter adopt a perspective which relates BR and tacit knowledge, in the context of an evolutionary theory of the behavior of business firms; Kahneman and Gigerenzer, with their collaborators, treat decision making in a perspective which connects psychology and economics, although they reach very different conclusions (for the former, agents commit systematic errors, while for the latter they elaborate accurate inferences).

This implies that we observe many applications (if not many versions) of BR. Yet, some common elements make it possible both to consider these contributions jointly and to maintain that the notion of BR has been revisited, since they diverge from the original version in some respects:

(1) A common thread seems to be the split between BR and AI. There is no reference to the rigorous structure of AI which contributed to defining BR; consequently the unitary view of rationality derived from the AI perspective disappears. Simon's view included intuition, emotion and perception as coherent parts of the information processing system, while in the new theories these elements assume a non-symbolic dimension, and work in ways different from 'reasoning'.

(2) In dealing with BR, the post-Simonian scholars emphasize the role of *unconscious/intuitive* mechanisms more than *conscious/intentional* ones, whereas the latter view was essentially adopted by Simon.

(3) Computational limits in processing information (and agents' limits of information) play a less important role with respect to the original version.

The second and third points are closely related, since Simon maintains that individuals, given their limits of computation and information, intentionally make decisions. By contrast, in the view of Kahneman and his collaborators, computation is not so important if individuals act largely intuitively, that is, in a rapid, parallel, non-reflexive and emotional way. Intuition is apparently not only the recognition of past experience reduced to the form of physical symbols contained in long-term memories of processors. Rather, intuition is a cognitive tool which is alternative to, but not separated from, deliberate reasoning, since it cooperates and may compete with the latter.

For Gigerenzer as well, computation is less important if 'fast and frugal' heuristics allow more accurate predictions than computationally complex strategies, and it is these (and other simple) heuristics that explain intuition. On the contrary, in Simon's view, computation matters: rationality is bounded because computational capacities are limited, and in a trivial sense, the less computation capacities are limited, the more rational is the individual.

Limits of information and information processing also play a minimal role in recent approaches to BR. Judgments and decisions are often generated by mechanisms which differ from the information processing mechanism described by Simon. As Kahneman and his colleagues show, emotions, intuitions and various pre-analytic assessments often intervene before any rational, however limited, information processing, and they influence decision-making, often giving rise to systematic errors. In turn, Gigerenzer & Todd (1999, p. 21) maintain, 'too much information processing can hurt', and partial ignorance can help in making correct inferences if the structure of heuristics is adapted to the structure of the environment. Therefore, limits in information processing do not necessarily engender limits in performances of rationality. In fact, Gigerenzer and his colleagues contend, limits of information can constitute an advantage, and 'ignorance makes us smart'. On both these views, despite the opposite issues, there emerges a residual role of intentionality as a conscious and reflexive act. In fact, for Kahneman, System 1 often prevails over System 2 and, for Gigerenzer and his group, adaptive (unconscious) rationality helps in making correct decisions, which are often better than deliberative ones.

Similar problems emerge in Nelson & Winter's theory, where BR is explained in terms of tacit knowledge, though Simon rejected this concept. In this case too, the focus is on automatic and non-reflexive responses which acquire form from the operation of skills.

Finally, March's contribution starts from the ambiguity and inconsistency of preferences to delineate the ambiguity and inconsistency of the self. This seems to modify Simon's view of individuals, according to which the structure of decisional processes is not weak, but only limited as regards computation and information.

The four approaches discussed here are very different, and some of them have provoked fierce discussion. Yet, the revision of the concept of BR is not the outcome of a specific debate which has highlighted its several dimensions. Rather, BR has been incorporated into different theories (characterized by distinctive methods and aims), and assumes distinctive meanings in each of them. For this reason, we may assume that there is an implicit challenge in redefining what constitutes BR, given the different applications of this notion.[20] This perspective can be appreciated once more by observing how Simon,

[20] There are many reasons this issue merits exploration, from the distinctive reception of Simon's view of BR among economists over time, to the emergence of new-connectionism in the mid-1980s, based on neural

Kahneman, and Gigerenzer differently describe the consequences of following simple heuristics in order to deal with problems. For Simon, heuristics, as rules of thumb, help agents to find 'satisficing' alternatives, and to reach 'good enough' solutions; for Kahneman, simple heuristics generally give rise to systematic errors; for Gigerenzer, 'fast and frugal' heuristics enable the formulation of accurate inferences, and the solution of a number of complex problems.[21]

References

Arrow, K. (1996) Preface, in: K. Arrow, E. Colombatto, M. Perlman & C. Schmidt (Eds) *The Rational Foundations of Economic Behaviour* (London: Macmillan).

Bechara, A., Damasio, H., Tranel, D. & Damasio, A.R. (1997) Deciding advantageously before knowing the advantageous strategy, *Science*, 275 (28 February), pp. 1293–1295.

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A. & Shizgal, P. (2001) Functional imaging of neural responses to expectancy and experience of monetary gains and losses, *Neuron*, 30, pp. 619–639.

Camerer, C. (1999) Behavioral economics, *CSWEP Newsletter*, Winter [*http://www.cswep.org/camerer.htlm*].

Camerer, C., Loewenstein, G. & Prelec D. (2005) Neuroeconomics: how neuroscience can inform economics, *Journal of Economic Literature*, 43, pp. 9–64.

Cordeschi, R. (2002) *The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics* (Dordrecht: Kluwer).

---

networks, which induced abandonment of the traditional AI (included Simon's one). But these topics are beyond the scope of the present article.

[21] One can remark that Simon's perspective is subjective; that is, it describes how individuals do not refer to optimal criteria, but consider decisions and outcomes as 'good enough' from their point of view, while an external observer could make a different assessment. By contrast, Kahneman and Gigerenzer describe an objective point of view. Nonetheless, this distinction is not explicitly delineated.

Cowan, R., David, P.A. & Foray, D. (2000) The explicit economics of knowledge codification and tacitness, *Industrial and Corporate Change,* 9, pp. 211−253.

Crowther-Heyck, H. (2005) *Herbert A. Simon. The Bounds of Reason in Modern America* (Baltimore: John Hopkins University Press).

Damasio, A. R. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Putnam).

Dasgupta, S. (2003) Multidisciplinary creativity: the case of Herbert A. Simon, *Cognitive Science*, 27, pp. 638−707.

De Martino, B., Kumaran, D., Seymour, B. & Dolan, R. J. (2006) Frames, biases and rational decision-making in the human brain, *Science*, 313, pp. 684−687.

Foss, N. J. (2003) Bounded rationality and tacit knowledge in the organizational capabilities approach: an assessment and a re-evaluation, *Industrial and Corporate Change*, 12, pp. 185−201.

Frederick, S. (2002) Automated choice heuristics, in: T. Gilovich, D. Griffin, & D. Kahneman (Eds) *Heuristics and Biases: The Psychology of Intuitive Judgment* (New York: Cambridge University Press)

Gardner, H. (1987) *The Mind's New Science: A History of the Cognitive Revolution* (New York: Basic Books).

Gigerenzer, G. (1991) How to make cognitive illusions disappear: beyond 'heuristics and biases,' *European Review of Social Psychology*, 2, pp. 83−115.

Gigerenzer, G. (1996) On narrow norms and vague heuristics: a reply to Kahneman and Tversky (1996), *Psychological Review*, 103, pp. 592−596.

Gigerenzer, G. (2002) The Adaptive Toolbox, in: G. Gigerenzer & R. Selten (Eds) *Bounded Rationality. The Adaptive Toolbox* (Cambridge, MA: MIT Press).

.Gigerenzer, G. (2004) Striking a blow for sanity in theories of rationality, in: M. Augier & J. G. March (Eds) *Models of a Man. Essays in Memory of Herbert A. Simon* (Cambridge, MA: MIT Press).

Gigerenzer, G. (2008) *Rationality for Mortals. How People Cope with Uncertainty* (Oxford: Oxford University Press).

Gigerenzer, G. & Goldstein, D.G. (1996) Reasoning the fast and frugal way: models of bounded rationality, *Psychological Review*, 103, pp. 650−699.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999) *Simple Heuristics That Make Us Smart* (Oxford: Oxford University Press).

Gigerenzer, G. & Todd, P. M. (1999) Fast and frugal heuristics. The adaptive toolbox, in: G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple Heuristics That Make Us Smart* (Oxford: Oxford University Press).

Gigerenzer, G. & Selten, R. (2002) Rethinking rationality, in: G. Gigerenzer & R. Selten (Eds) *Bounded Rationality. The Adaptive Toolbox* (Cambridge, MA: MIT Press).

Gilovich, T. & Griffin, D. (2002) Introduction—heuristics and biases: then and now, in: T. Gilovich, D. Griffin, & D. Kahneman (Eds) *Heuristics and Biases: The Psychology of Intuitive Judgment* (New York: Cambridge University Press)

Goldstein, D. G. & Gigerenzer, G. (1999) The recognition heuristic: how ignorance makes us smart, in: G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple Heuristics That Make Us Smart* (Oxford: Oxford University Press).

Kahneman, D. (2003) Maps of bounded rationality: psychology for behavioral economics, *American Economic Review*, 93, pp. 1449−1475.

Kahneman, D. & Frederick, S. (2002) Representativeness revisited: attribute substitution in intuitive judgment, in: T. Gilovich, D. Griffin, & D. Kahneman (Eds) *Heuristics and Biases: The Psychology of Intuitive Judgment* (New York: Cambridge University Press)

Kahneman, D. & Frederick, S. (2005) A model of heuristic judgment, in: K. J. Holyoak & R. G. Morrison (Eds) *The Cambridge Handbook of Thinking and Reasoning* (Cambridge: Cambridge University Press).

Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1991) Anomalies: the endowment effect, loss aversion and status quo bias, *Journal of Economic Perspectives*, 5, pp. 193−206.

Kahneman, D. & Tversky, A. (1996) On the reality of cognitive illusions, *Psychological Review*, 103, pp. 582−591.

Klaes, M. & Sent, E.-M. (2005) A conceptual history of the emergence of bounded rationality, *History of Political Economy*, 37, pp. 27−59.

March, J. G. (1978) Bounded rationality, ambiguity, and the engineering of choice, *Bell Journal of Economics*, 9, pp. 587−608.

March, J. G. (1988) *Decisions and Organizations* (Oxford: Basil Blackwell).

March, J. G. (1994) *A Primer on Decision Making. How Decisions Happen* (New York: The Free Press).

March, J. & Simon, H. A. (1958) *Organizations*, 2nd edition (Cambridge: Blackwell, 1993).

Miller, G. (2006) The emotional brain weighs its options, *Science*, 313, pp. 600−601.

Mirowski, P. (2002) *Machine Dreams. Economics Becomes a Cyborg Science* (Cambridge: Cambridge University Press).

Motterlini, M. & Guala, F. (2005) Psicologia ed esperimenti in economia, in: M. Motterlini, & F. Guala (Eds) *Economia cognitiva e sperimentale* (Milano: Università Bocconi Editore).

Nelson, R. R. & Winter, S. G. (1982) *An Evolutionary Theory of Economic Change* (Cambridge, MA: Harvard University Press).

Newell, A. (1989) Putting it all together, in: D. Klahr, & K. Kotovsky (Eds) *Complex Information Processing: The Impact of Herbert A. Simon* (Hillsdale: Lawrence Erlbaum Associates).

Newell, A. & Simon, H. A. (1972) *Human Problem Solving* (Englewood Cliffs: Prentice-Hall).

Newell, A. & Simon, H. A. (1981) Computer science as empirical inquiry: symbols and search, in: J. Haugeland (Ed.) *Mind and Design. Philosophy, Psychology, Artificial Intelligence*. (Cambridge, MA: MIT Press).

Newell, A., Shaw, J. C. & Simon, H. A (1962) The process of creative thinking, in: H. A. Simon, *Models of Thought*, Vol. I (New Haven: Yale University Press, 1979).

Nightingale P. (2003), If Nelson and Winter are only half right about tacit knowledge, which half? A Searlean critique of 'codification', *Industrial and Corporate Change*, 12, pp. 149−183.

Polanyi, M. (1967) *The Tacit Dimension* (Garden City: Doubleday Anchor).

Sanfey, A. G., Loewenstein, G., McClure, S. M. & Cohen, J. D. (2006) Neuroeconomics: cross-currents in research on decision-making, *Trends in Cognitive Sciences*, 10, pp. 108−116.

Sent, E.-M. (2004) Behavioral economics: how psychology made its (limited) way back into economics, *History of Political Economy*, 36, pp. 735−760.

Shafir, E., Simonson, I. & Tversky, A. (1993) Reason-based choice, *Cognition*, 49, pp. 11−36.

Simon, H. A. (1947) *Administrative Behavior* (New York: Macmillan).

Simon, H. A. (1955) A behavioral model of rational choice, in: H. A. Simon, *Models of Bounded Rationality. Behavioral Economics and Business Organization*, Vol. 2 (Cambridge, MA: MIT Press, 1982).

Simon, H.A. (1956) Rational choice and the structure of the environment, in: H. A. Simon, *Models of Bounded Rationality. Behavioral Economics and Business Organization*, Vol. 2 (Cambridge, MA: MIT Press, 1982).

Simon, H. A. (1957) *Models of Man* (New York: Wiley).

Simon, H.A. (1964) Information processing in computer and man, in: H.A. Simon, M. Egidi, R. Marris & R. Viale (Eds) (1992) *Economics, Bounded Rationality and the Cognitive Revolutions* (Aldershot: Edward Elgar).

Simon, H. A. (1962) The architecture of complexity, *Proceedings of the American Philosophical Society*, 106, pp. 467−482.

Simon, H. A. (1967) Motivational and emotional controls of cognition, in: H. A. Simon, *Models of Thought*, Vol. I (New Haven: Yale University Press, 1979).

Simon, H.A. (1976a) Bradie on Polanyi on the Meno paradox, *Philosophy of Science*, 1, pp. 147−151.

Simon, H. A. (1976b) From substantive to procedural rationality, in: H. A. Simon, *Models of Bounded Rationality*. *Behavioral Economics and Business Organization*, Vol. 2 (Cambridge, MA: MIT Press, 1982).

Simon, H.A. (1979) Rational decision making in business organizations, *American Economic Review*, 69, pp. 493−513.

Simon, H.A. (1982) *Models of Bounded Rationality*. *Behavioral Economics and Business Organization*, Vol. 2 (Cambridge, MA: MIT Press).

Simon, H. A. (1983a) Discovery, invention, and development: human creative thinking, *Proceedings of the National Academy of Sciences of the United States of America*, Part 2: Physical Sciences, Vol. 80, pp. 4569−4571.

Simon, H.A. (1983b) *Reason in Human Affairs* (Oxford: Basil Blackwell).

Simon, H. A. (1987) The information-processing explanation of Gestalt phenomena, in: H. A. Simon, *Models of Thought*, Vol. II, (New Haven: Yale University Press, 1989).

Simon, H. A. (1990) Invariants of human behavior, *Annual Review of Psychology*, 41, pp. 1−19.

Simon, H.A. (1991) *Models of My Life* (Cambridge, MA: MIT Press).

Simon, H. A. (1996) *The Sciences of the Artificial*, 3rd edn (Cambridge, MA: MIT Press).

Simon, H.A. (1997a) Spiegare l'ineffabile: l'intelligenza arificiale a proposito di intuizione, insight e ispirazion, *Sistemi intelligenti*, 9, pp. 169−192.

Simon, H. A (1997b), *Models of Bounded Rationality*. *Empirically Grounded Economic Reason,* Vol. 3 (Cambridge, MA: MIT Press).

Simon, H. A. & Shaw, J. C. (1958) Elements of a theory of human problem solving, in: H. A. Simon, *Models of Thought*, Vol. II (New Haven: Yale University Press, 1989).

Sloman, S. A. (2002) Two systems of reasoning, in: T. Gilovich, D. Griffin, & D. Kahneman (Eds) *Heuristics and Biases: The Psychology of Intuitive Judgment* (New York: Cambridge University Press)

Slovic, P., Finucane, M., Peters, E. & MacGregor, D. G. (2002) The affect heuristic, in: T. Gilovich, D. Griffin, & D. Kahneman (Eds) *Heuristics and Biases: The Psychology of Intuitive Judgment* (New York: Cambridge University Press)

Stanovich, K. E. & West, R. F. (2002) Individual differences in reasoning: implications for the rationality debate?, in: T. Gilovich, D. Griffin, & D. Kahneman (Eds) *Heuristics and Biases: The Psychology of Intuitive Judgment* (New York: Cambridge University Press)

Todd, P.M. (2002) Fast and frugal heuristics for environmentally bounded minds, in: G. Gigerenzer & R. Selten (Eds) *Bounded Rationality. The Adaptive Toolbox* (Cambridge, MA: MIT Press).

Todd, P. M. & Gigerenzer, G. (2003) Bounding rationality to the world, *Journal of Economic Psychology*, 24, pp. 143−165.

Tversky, A. & Kahneman, D. (1987) Rational choice and the framing of decisions, in: R. M. Hogarth, & M. W. Reder (Eds) *Rational Choice. The Contrast between Economics and Psychology* (Chicago: University of Chicago Press).

Winograd, T. & Flores, F. (1986) *Understanding Computers and Cognition. A New Foundation for Design* (Norwood: Ablex).

Zajonc, R. B. (1980) Feeling and thinking: preferences need no inferences, *American Psychologist*, 35, pp. 151−175.