"S. Cognetti de Martiis"

Via Po, 53 – 10124 Torino (Italy)
Tel. (+39) 011 6704917 - Fax (+39) 011 6703895
URL: http//www.cesmep.unito.it

WORKING PAPER SERIES

Herbert A. Simon and Contemporary Theories of Bounded Rationality

Stefano Fiori

Dipartimento di Economia "S. Cognetti de Martiis"

Centro di Studi sulla Storia e i Metodi dell'Economia Politica
"Claudio Napoleoni"
(CESMEP)

Working paper No. 02/2008

Università di Torino

Herbert A. Simon and Contemporary Theories of Bounded Rationality[*]

STEFANO FIORI[**]

Abstract. Firstly, the paper seeks to show that Simon's notion of bounded rationality should be interpreted in light of its connection with artificial intelligence. This connection points out that bounded rationality is a highly structured concept, and sheds light on several implications of Simon's general view on rationality, especially as regards intentionality and instrumentality. Secondly, offering four paradigmatic examples, the article presents the view that recent approaches, which refer to Simon's heterodox theory, only partially accept the teachings of their inspirer, splitting bounded rationality from the context of artificial intelligence, and substituting it with different analytical tools which help give new configurations to bounded rationality. In particular, consideration of the role of automatic, neurobiological, and unconscious mechanisms (and their interactions with deliberate reasoning) provides new perspectives on how (bounded) rationality works. The thesis is that these events can be interpreted as an implicit (and ideal) challenge for redefining what bounded rationality is.

JEL Classification: B4; B21; D80; D89

Key words: bounded rationality; artificial intelligence, intuition, heuristics, tacit knowledge

Introduction

This paper is divided into two parts. The first discusses Simon's notion of bounded rationality (henceforth BR), starting from his juvenile works (sect. 1), and subsequently in the light of its connections with artificial intelligence (henceforth AI), when it acquired a more definitional, theoretical, dimension (sects. 2; 3; 4).

BR is analyzed along with the concept of "intuition", which Simon explored in a mature phase of his research on AI, since, in my opinion, this perspective allows one to emphasize certain problematic aspects of Simon's approach to rationality (sect. 5). In fact, intentional BR, as a decision-maker's conscious activity, and intuition (as an unintentional, cognitive process) represent two extremes of rational behavior which belong to the same information processing mechanism. Since the notion of intuition has been treated by Daniel Kahneman as well, both the approaches are compared, with their differences pointed out.

[*] This is a revised, and very different, version of *Simon's Bounded Rationality. Origins and use in Economic Theory*, CESMEP Working Paper /2005. Earlier versions of this paper have been presented at the 32nd Annual Meeting of the History of Economics Society (HES), Tacoma (USA), June 24-27, 2005 and at the Tenth Annual Conference of the European Society for the History of Economic Thought (ESHET), Porto, April 28-30, 2006. I thank discussants and participants for helpful comments.

[**] Department of Economics "S. Cognetti de Martiis", Università di Torino, e-mail: stefano.fiori@unito.it

The second part of the paper discusses how BR has been used by some representative authors, such as March, Nelson, Winter, Kahneman, Tversky, Gigerenzer, and others, who have acknowledged their intellectual debt to Simon and whose work in turn has been appreciated by Simon himself (sects. 7; 7.1; 7.2; 7.3; 7.4).

The concept of BR emerged in its essential terms in *Administrative Behavior* (1947), and was subsequently specified as an economic notion in the mid-1950s, during the initial phase of Simon's interest in AI. This latter constituted a fundamental change, in that BR assumed a more definitional form, to which economists prevalently refer, and was more clearly conceived as an alternative to the neoclassical notion of rationality. The general impression is that over time the problem of how individuals make decisions was considered in the light of those disciplines (administrative, organization, economic, AI theories, and later cognitive psychology) which jointly contributed to providing a unitary answer, although AI and cognitive psychology had become increasingly important in Simon's scientific search. In this way, concepts such as the "limited computational capacities" of decision makers and "information processing", were conceived in an interdisciplinary perspective. Yet the notion of "information processing" was accurately defined within the context of AI; therefore, it seems plausible to analyze how AI contributed to providing new tools for Simon's economic theory, and, more precisely, how AI sheds light on some aspects of BR. This perspective points out that BR is a highly structured concept, embodying some features partially shared with those disciplines which Simon included as an essential part of the "Zeitgeist" of the 1940s and early 1950s. In many respects, Simon considered his scientific experience to be the result of this climate, in which cybernetics, formal logic, information theory and many other scientific approaches emerged.

Finally, given these premises, two problems will be discussed in this paper: 1) how BR is dealt with by some contemporary authors who have accepted Simon's non-standard approach to economics; 2) how they have received the implications arising from the distinctive interrelations between different disciplines in Simon's work. A survey, presenting four paradigmatic cases, suggests that Simon's teaching has been only partially accepted, and many recent studies essentially do not refer to AI and substitute it with other, 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 non-marginal shifts with respect to the original program of their inspirer. 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 AI**.** This 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, whilst 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 (despite their reference to Simon's thought). 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 a derivation of (and sometimes are opposite to) AI assumptions. Therefore, one outcome is the appearance of a BR *weaker* with respect to Simon's, and essentially based on different conceptual frameworks.

1. Bounded rationality: the emergence of the concept

Most economists know Herbert A. Simon as the theoretician of BR.

Some of them have sought to include BR in standard approaches, or to incorporate it into specific schools of thought, from game theory to new institutionalism,[1] whilst others have preferred to assume BR as a basic concept with which to elaborate alternative visions of rationality, following (and developing) Simon's program.[2] However, most of them refer to Simon's contributions of the 1950s and thereafter: for example *A Behavioral Model of Rational Choice* (1955), probably the most frequently cited of Simon's articles (Sent, 2000, p. 395; Mirowski, 2002, p. 460), where he declared his intention to give "drastic revision" to the model of the optimal rationality of economic man, and replace it with another model which considered the limited computational capacities and access to information in the real environment (Simon, 1955, p. 239). In short, treatment of Simon's thought generally centers on the notion of BR (and correlated concepts), and in a number of cases assumes as its reference point the version that emerged in the mid-1950s, and some of its subsequent developments.

The term BR first appeared in *Models of Man* (Simon, 1957a, p. 196), where Simon, among other things, maintained that the principle was described in *Administrative Behavior* (1947) (Simon,

[1] Here I am not considering notions of BR as incorporated into "standard" approaches like those of Stigler (1961) and Sargent (1993). On this see Conlisk (1996); Sent (1997); (Gigerenzer and Selten, 2002). Moreover, I do not analyze how BR has been dealt with in game theory, new institutionalism, behavioral economics and others fields of economics.

[2] Judgments have sometimes been more severe: "Much of Simon's work was conceptual rather than formal […] it was recognized that the area was of great importance, but the lack of a formal approach impeded its progress" (Aumann, 1997, p. 3; see also Arrow, 2004). Camerer connects Simon's unenthusiastic reception to the success of standard economics in explaining economic decision and equilibria in "elegant mathematical terms" (Camerer, 1999, p. 1). Professor Esther-Mirjam Sent (personal communication) suggests that the diversity in Simon's theoretical production influenced its reception (cf. Sent, 2000). Moreover, the potentialities of the notion of BR, as an approach alternative to neoclassical economics, were not entirely developed, and this may explain why some authors have tried to include it in their standard theories.

1957a, p. 199). The ten years that separate the two books were very important for Simon, since a number of events induced him to modify his initial interest in administrative theory, so that at the end of this period his research focused mainly on AI and cognitive science. One outcome of this process was that from 1947 to 1957 Simon "consciously refined and replaced concepts such as 'approximate rationality' and 'limited rationality' until settling for 'bounded rationality'" (Klaes and Sent, 2005, p. 37).

In regard to this topic, Augier (2000), and Augier and March (2002) point out a "considerable continuity" in Simon's writings, and highlight his constant effort to clarify the real processes of human decision making; an endeavor which must be viewed in terms of "gradual transformation" of his theory. This interpretation is in polemic with Sent's and Mirowski's view (Sent, 2000; Mirowski, 2002), according to which the Cold War, the climate generated by cyborg science, and the experience at RAND Corporation in the mid-1950s exerted a strong influence which culminated in a "seachange in Simon's interest". More precisely, Sent writes, this radical change was connoted as a shift of his focus from the analysis of human decision-making in organizations and public administration to problem solving analysis (Sent, 2000).

On the one hand, the constancy of his scientific interest is emphasized by Simon himself, who described himself as a monomaniac, who throughout his life was "obsessed with human decision making and problem solving activity" (Simon, 2001, p. 501, see Feigenbaum, 1989, p. 180), and it is confirmed by his friend and collaborator Allen Newell, who specified that BR was the central idea in Simon's scientific life as a whole (Newell, 1989, p. 400). On the other hand, the concept of BR as limited "information processing capacity", which focuses on the limits of information and of computational capacities, is the result of an intellectual process whose turning point is well-defined. In fact, Simon writes that his life as a scientist "changed radically in the last months of 1955", and "Soon I was transformed professionally into a cognitive psychologist and computer scientist, almost abandoning my earlier professional identity", that of a scholar of administration and economics (Simon, 1991, p. 189).

Given these premises, a radical change did not take place in Simon's basic interests; rather, it appeared in theoretical tools which he adopted over time in order to explain human decision-making and problem solving. Therefore, Augier is right to point out the continuity in Simon's research interests, but she does not adequately capture the radical change caused by the use of new analytical instruments over time. Sent, in her turn, is certainly right to stress a profound change in Simon's scientific life in the mid-1950s, but she seems to undervalue the fact that the new analytical tools were applied to an essentially invariant object: decision making (even in the form of problem solving).[3]

Administrative Behavior is the juvenile work in which the concept of "limits of rationality" appears most extensively, and its analysis enables comparison with the subsequent treatment of BR. As Simon stated, in this work "bounded rationality [was] largely characterized as a residual category". It required a more formal treatment which appeared, as "first step" in that direction, in two famous papers (Simon, 1955, 1956). As a consequence, only "by the middle 1950's, a theory of bounded rationality had been proposed as an alternative to classical omniscient rationality" (Simon, 1979, pp. 502-503), i.e., when he met Allen Newell at the RAND Corporation and began to devote large part of his research to AI (see sect. 2)[4]. More specifically, with respect to the "rather vague and general initial formulation of the idea of bounded rationality", he writes, "During the decade that followed the publication of *Administrative Behavior*, substantial progress was made in both direction ["formalization" and "empirical verification"]" (Simon, 1979, p. 501). The same judgment (that is, limited rationality as a "residual category" in the first version of *Administrative Behavior*) is re-affirmed in subsequent editions (Simon, 1957b, 1976c, 1997a), where he left the original text largely unchanged but equipped it with new *Introductions* (and "new commentaries", in the fourth edition) (cf. Golembiewski, 1988, p. 294). Limited rationality in this book, he writes in these comments, was not "a departure from [standard vision of] rationality", and "the positive characterization of the process of choice was very incomplete" (Simon, 1957b, p. xxv; 1976c, p. xxix; 1997a, p. 118).

The lack of "positive characterization" raises some problems. Firstly, as we have seen, it emphasizes the scant "formalization" and "empirical verification" of the initial formulation of BR. Yet it also probably implies that a more robust definition was subsequently accomplished by means of AI tools. Secondly, this expression seems to suggest that the first edition of *Administrative Behavior* provided a *negative characterization*, so to speak; that is, a choice process which merely implied a description of differences between the rationality of the ideal "economic man" and the empirical (limited) rationality of the "administrative man", (Simon, 1957b, p. xxv; 1976c, p. xxix; 1997a, p. 118). In this sense, a positive model based on bounded rationality standing as an alternative to the omniscient model was not constructed.

---

[3] Both interpretations refer to the relation between Simon's intellectual life and his experience at RAND, at Carnegie Mellon University, and at the Cowles Commission, during and after World War II. In particular, Augier contests the central role of the RAND with respect to Carnegie Mellon in shaping Simon's interests. By contrast, my perspective focuses more on the theoretical tools that Simon adopted over time, and much less on the influence of certain institutions on his work.

[4] See Egidi and Marengo (2004, p. 339).

In particular, Simon in his *Introductions* specified that chapters IV and V (which deal with these topics) are closely connected, in that the former describes the optimal rationality of "economic man", who maximizes, while the latter depicts "how actual behavior deviates from the economic model", and describes the administrative man, who "satisfices", i.e., selects a "good enough" course of action (Simon, 1957b, p. xxv; 1976c, p. xxviii-xxix). Behind this framework, the original text exhibits the presence of neoclassical categories (Bartlett, 1988, p. 306) which account for Simon's subsequent doubts about the role of BR, and related concepts, in the first edition of this work. For example, the notion of efficiency is strongly emphasized. It "must be the guiding criterion" of administrative organizations (Simon, 1947, p. 36), and it "requires that results be maximized with limited resources" (Simon, 1947, p. 197, cf. p. 65).[5] In short, the standard criterion of efficiency is a *benchmark* for the administrator. All this also emerges when Simon discusses perfect rationality and concludes that his aim was "to explore the anatomy of decision with a view to establishing a terminology and a framework of analysis that permit a realistic investigation of administrative decision" (Simon, 1947, p. 77). As a consequence, the conduct of "realistic investigation" depends on elaboration of a context and of a vocabulary based on full rationality, assumed as the reference point, against which to compare the "limits of rationality". In this perspective, limited rationality is not a radical departure from standard economics; and not surprising is Simon's complaint, with respect to chapter IX of the book, that he "yielded too much ground to the omniscient rationality of economic man" (Simon, 1957b, p. xxxv).

The relation between chapters IV and V should be interpreted in the terms delineated above: the former describes economic "principles and theories" ("how *good* administrators decide"), i.e., the benchmark; the latter delineates empirical, rationally limited behavior ("how administrators [empirically] decide"), i.e., the distance from the model (Simon, 1947, p. 62). This framework evokes the current idea in standard economics concerning the relation between abstract models and empirical facts, with the consequence that BR, incorporated in administrator behavior, is not the unit of analysis from which to develop a new paradigm. The "administrative man" is not the opposite of the "classical economic man", but rather "takes his place alongside" him (Simon, 1947, p. 39). Given these premises, chapter V introduces the limits of rationality more as empirical bounds[6] than as the basis for an alternative model, whilst in Simon (1955) the task will be "to replace the global rationality of economic man" (p. 239). Therefore, limited rationality and

[5] Efficiency is more a "definition" than a "principle", and it "states that this maximization is the aim of administrative theory, and that administrative theory must disclose under what conditions the maximization takes place" (Simon, 1947, p. 39). These concepts, as Simon reminds us (1947, p. 39, note 16) were treated in Ridley and Simon (1943); in particular see p. 3. See also Simon (1991, p. 64).

[6] Specifically, lack of complete knowledge, difficulties in anticipating consequences of choices, physical and biological limitations of possible behaviors (Simon, 1947, pp. 80-84; see also pp. 39-41).

"satisficing" in *this* book have been viewed as concepts emerging within the (Kuhnian) "normal science", rather than as new principles for a revolutionary paradigm (cf. Brown, 2004, p. 1248).

The *Introduction* to the second edition of *Administrative Behavior* (1957), as we have seen, evinced Simon's awareness that in the 1940s limited rationality exhibited an unsatisfactory, theoretical, dimension. This point of view appears in other writings of the same period, and reveals a constant effort to modify his description of rationality. Hence, in *A Formal Theory of the Employment Relationship* (1951), where Simon develops a model based on a distinction between sales contract and employment contract, he maintains: "The most serious limitations of the model lie in the assumptions of rational utility-maximizing behavior incorporated in it" (Simon, 1951, p. 305).

This perspective highlights Simon's intellectual process. In particular, the insistence on the impossibility "for the behavior of a single, isolated individual to reach any high degree of rationality" (Simon, 1947, p. 79) was a fundamental step in re-thinking the notion of rationality. This latter took shape over time, starting from empirical realization of its limits, and culminated in a more radical vision which was the opposite to that of omniscient economic man. Finally, when Simon wrote the 1957 *Introduction*, some important events had modified his early perspective; and BR, conceived in more precise, rigorous, terms, was now a category ready to challenge the standard view of rationality in economics. This latter version is the one most widely known by economists (although they have often referred only to a few parts of Simon's theory), and by other scholars.

To summarize, two correlated elements characterize Simon's view on the limits of rationality in the 1950s: 1) the transformation of the BR notion, which became the basic assumption for a paradigm alternative to standard theories of decisional processes; 2) the use of new instruments, drawn from AI apparatus, to explain how BR works.

In what follows (sects. 2; 3; 4; 5), I shall seek to show how this mature version of BR emerged in close connection with AI. Subsequently, I will discuss its reception, showing how the concept of BR has tacitly changed its features in recent studies inspired by Simon's theory, where the AI approach is replaced by new perspectives and methodologies.

2. Bounded rationality and Artificial Intelligence: some interrelations.

Theories of BR, Simon maintains, deal with the limits of "information processing capacities". More precisely, the definition of information processing stresses two constraints on individuals: 1) the limits of information, which is gathered and processed; 2) the limits of computational capacities, which emerge when agents face situations perceived as complex (Simon, 1955; 1972; 1976b; 1978; 1996).

Given these limits, the decisional process applies when problems requiring solutions occur. More precisely, Simon and March emphasize that when actors (organizations, in this case) receive an external input or stimulus, they react by either 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 (March and Simon, 1958). 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 latter, in turn, allows them to gather new information. Since uncertainty prevails in the real world, neither all the alternatives nor the consequences that would follow from them are evaluated, and heuristics (as "rules of thumb") are adopted to simplify search processes (Newell, Shaw, Simon, 1962; Newell, Simon, 1972; Simon, 1990, p. 9). Therefore, decision-makers look for a "satisficing", rather than an optimal, alternative, and the criteria, which perform this role in decision-making processes, are called "aspiration levels". In particular, "satisficing" is a "weak method" which problem-solvers apply when task domains are ill-structured or are unknown, and this allows them to halt the search process when a solution meets their expectations (Simon, 1983a, p. 4570; 1990, p. 9). Finally, problem-solving (sequential) procedures are attained by breaking down the problem into smaller components. In fact, problems are often too complex for the agents' computational capacities, and 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 rather 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 computer science: specifically, the definition of *Information Processing Languages* (IPL), and the creation of both *Logic Theorist* (LT) and *General Problem Solver* (GPS) programs. Finally, in June 1956, a seminar crucial for the further development of AI was held in Dartmouth, with the participatation, among others, of J. McCarthy, M. Minsky, C. Shannon, H.A. Simon and A. Newell. Simon describes 1955 and 1956 as the most important years in his scientific life (Simon, 1991, p. 189); and so it must have been if we consider that, besides his activity in the field of AI, two very important articles on (bounded) rationality in economics were published in those years: *A Behavioral Model of Rational Choice* (1955) and *Rational Choice and the Structure of the Environment* (1956). In a note to the latter essay, as if he wished to stress his unitary research program, Simon declared himself "indebted to Allen Newell for numerous enlightening conversations on the subject of this paper" (Simon, 1956, p. 259)[7].

In short, 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 (see McCorduck, 1979; Crevier, 1993; Edwards, 1997). In addition, Simon connected all this to a long intellectual tradition which included, among other things, the "influence of formal logic", cybernetics and the information theory of Wiener and Shannon. This tradition, as regards the decades prior to World War II, has been described in terms of

> a powerful, growing Zeitgeist, having its origins around the turn of the century, involving a deep faith in mathematics as the language of science, but, more crucially, focused upon symbols and their manipulation in logical inference (if not in "reasoning") and decision. Symbols became, for the first time, tangible – as tangible as wood or metal. (Newell and Simon, 1972, p. 878)

3. AI and Human Thinking

Simon constantly emphasized the analogy between human thinking and computer intelligent behavior.[8] From this perspective, the first step consists in considering the human thought 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 and 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

---

[7] A similar acknowledgment is made in *Organizations*, where Allen Newell's influence on this work is mentioned in regard to "the treatment of human problem-solving" (March and Simon, 1993, p. viii).

[8] "Like a modern digital computer's, Man's equipment for thinking is basically serial in organization […] there 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 and Shaw, 1958, and Simon (1964)).

I assume that a close relation exists among the terms "intelligence", "thought" and "rationality", as well as between "intelligent behavior" and "bounded rationality". Simon explains the link, as follows: "Because of the limits on their computing speeds and power, intelligent systems must use approximate methods. Optimality is beyond their capabilities; their rationality is bounded." Because of this fact Simon uses the expression "limits of human and computer rationality" (Simon, 1990, p. 17 and p. 8). In other words, the computer must operate with BR. In general, Simon applies BR to every sphere in which decision theory take place, particularly economics, psychology, and AI, in fact, "Economics is one of the sciences of the artificial" (Simon, 1976b, p. 441). See also, Newell and Simon (1981, pp. 52-54).

comprehension of mind, since both computer and mind belong to "physical symbol systems". 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 and 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 chess-board. 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 binds behavior in many ways: defining admissible moves (the set of possible sequential courses of action adopted for accomplishing specific goals), specifying the aim, and interacting with short- and long-term memories.

Task environment, internal representation and space problem 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.[9]

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 "test" (the solution to a problem, even if we do not know how to achieve it)[10], and a "generator" of symbol structures, that is, a move generator for a potential solution. A problem will be solved if the generator produces a structure which satisfies the test (Newell and Simon, 1981, pp. 52-53).

4. The "complex algorithms of thought"[11]

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.[12]

---

[9] "In talking about the task environment we must maintain clear distinctions among the environment itself (the Kantian *Ding an sich*, as it were), the internal representation of the task environment used by the subject (the problem space), and the theorist's 'objective' description of that environment" (Newell, Simon, 1972, p. 56). See Newell and Simon, (1972, p. 824).

[10] This means that we know what we want to do, but we do not know how to accomplish it. If symbol systems were not characterized by limited computational capacities, for example, the test for checkmate would be a strategy that achieves this aim for all counter strategies of the other player. But this is impossible in the real world. Once again, man and computer exhibit the same limitations as regards computation, intelligence and rationality (Newell and Simon, 1981, p. 53).

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

[12] BR can be explained in terms of a symbolic system: "It is precisely when we begin to ask *why* the properly motivated subject does not behave in the manner predicted by the rational model that we recross the boundary again from a theory

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 (Newell and Simon, 1972 p. 877; 1981, p. 42; Simon, 1991, pp. 192-194)[13]. The debt regarding the "logicians" is summarized as follows:

> The fundamental contribution was to demonstrate by example that the manipulation of symbols (at least *some* manipulation of *some* symbols) could be described in terms of specific, concrete processes quite as readily as could the manipulation of pine boards in a carpenter shop. 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 and Simon, 1972, p. 877)[14]

Newell and 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[15]. 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. More generally, they seem to accept the project for the construction of a human thinking alphabet whose origin was in "Leibniz's dream", that is, the dream of a general science of reasoning which can be applied in every field of knowledge (Kline 1972; Davis 2000). Simon and collaborators stress that their approach does not concern homologies between a computer and the nervous system. On the contrary, it describes how both intelligent behavior and its *general* properties emerge, and this implies explaining and treating, so to speak, the *language of thought* by means of symbol logic. Moreover, formal logic determines a "new operationality of symbols", although symbol manipulation refers to broader areas of knowledge with respect to deductive logic (Newell and Simon, 1972, p. 877). According to Newell and Simon, there are a number of thinkers who have contributed to linking, in different ways, formal logic, mathematics, and symbol

---

of the task environment to a psychological theory of human rationality. The explanation must lie inside the subject: in limits of his ability to determine what the optimal behavior is, or to execute it if he can determine it. In simple concept attainment experiments, for example, the most important mechanism that prevents from adopting an efficient strategy is usually the limit on the number of symbols he can retain and manipulate in short-term memory." (Newell and Simon, 1972, pp. 54-55). See Newell and Simon (1981, pp. 53-54)

[13] This point is also emphasized in Dreyfus H.L. and Dreyfus S.E. (1990); see also Bartlett (1988); Brown (2004).

[14] See Newell and Simon (1981, pp. 42-43).

[15] As regards distances between Simon and the formal logic approach, see Mirowski (2002, p. 462).

manipulation; they include Turing, Carnap, Church, Shannon, Lotka, Wiener, Pitts, McCulloch, and many others.

To be noted is that Simon attended Carnap's lectures at the University of Chicago. The philosopher, he pointed out, "was particularly important to me, for I had a strong interest in the logic of the social sciences" (Simon 1991, p. 53). This influence was not marginal, since Simon affirms that Carnap's theory had an important role in deciding his thesis project, which culminated in *Administrative Behavior*. In fact, this study initially concerned the logical foundations of administrative science[16].

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 twentieth century, which was inspired by the formal logic of Frege, Whitehead and Russell. Therefore, the following statement, quoted from *Administrative Behavior* is unsurprising:

> the conclusions reached by a particular school of modern philosophy – logical positivism – will be accepted as a starting point, and their implications for the theory of decisions examined. (Simon, 1947, p. 45)

Crowther-Heyck 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 (2005, p. 73). Moreover, Simon followed Bridgman's operationalism, which like logical positivism was considered as a means of unifying science, and he held that "operational definition of terms is crucial to making statements clear and testable" (Crowther-Heyck, 2005, p. 101). As a consequence, the operationalist approach re-appeared later in order "to justify the use of computers to simulate human cognition" (Crowther-Heyck, 2005, p. 130).

Given these premises, several points can be emphasized:

1) rationality involves intelligence, and in its turn intelligence implies *physical symbol systems*. Consequently, rationality 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.[17]

[16] A further sign of this influence was the dedication to Carnap and Henry Schultz, another university teacher, of his book *Models of Discovery*, in 1977. However, it is evident that Simon's intellectual history cannot be reduced to the elements above mentioned, for a historical and theoretical interpretation, see Mirowski (2002).

[17] Simon, for example, criticized Michael Polanyi's "tacit knowledge", with reference to the so-called "Meno Paradox", (Simon, 1976a); see sect. 7.2.

2) BR is instrumental in nature; it is essentially evoked to solve problems, which often exhibit a logical form (for example, cryptarithmetic problems, proofs of theorems, etc.) (Hargreaves Heap *et al.*1992).

3) Studying how problem solvers deal with a cognitive task implies that "we are observing *intendedly rational behavior* or behavior of *limited rationality*" (Newell and Simon, 1972, p. 55; Simon, 1976c, p. xxviii).[18] Therefore, the attention is focused on deliberate, conscious, and intentional rationality, which as a matter of fact is limited. In this sense, Winograd and Flores maintain, Simon does not contest the "rationalistic tradition", but only the version that implies perfect knowledge, perfect foresight, and optimitization criteria (Winograd and Flores, 1986, p. 22, see also O'Neill, 2005, p. 296, 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 and Shaw, 1958, p. 6).

This interpretation gives rise to conclusions different from those of Mirowski, who stresses the dissolution of the notion of individual Self due to the influence of cyborg sciences (to which Simon's AI was connected) on modern economics. On this view, the cognitive approach in economics has proved to be "a debacle for the deliberative Self", and Simon "abandoned the old-fashioned self in favor of program of constructing simulacra of people" (Mirowski, 2002, p. 452). Finally, Simon's dissolution of the Self is described as follows:

> There is more than one way to demolish the Self, Herbert Simon initiated his own process of deconstruction through the twin techniques of decomposing the self into little hierarchies of problem-solving modules, in passing blurring the boundaries between organism and environment, and by pursuing the method of simulation, which effaced the external integrity of the functioning self. (Mirowski, 2002, p. 479)

It emerges from our previous analysis that the agent's bounded rationality is instrumental, intentional, and connoted by unambiguous symbolic processes, and hence the related idea that the self reflects such features. This self, although rationally limited, is strongly structured; it coherently follows simple rules (of thumb) in its decisional processes, which produce neither internal conflicts nor disintegration of personality. Moreover, it is not divided among conflicting and incoherent

[18] Newell maintains that for Simon the structure of behaving systems is "intendedly rational, subject to the limits of knowledge and computing power" (Newell, 1989, p. 411). See Augier and March (2002, p. 7)

components driven by contradictory instances. The "techniques of decomposing the self into little hierarchies of problem-solving modules", do not imply a deconstruction of self, but only a deconstruction of problems into smaller parts, an action accomplished by a unitary self, however rationally limited. In addition, "the boundaries between organism and environment" are not blurred; on the contrary, agents and environment are distinct. The external world *per se* is unknown, as Simon points out on assuming a Kantian perspective (see note 10), and must consequently be represented in a subjective model (internal representation); it imposes tasks (task environment) and problems to solve (within the space problem). According to Simon, human beings are adaptive systems that respond to the task environment, given their limited capacities. Hence "men's goals define the interface between their inner and outer environment" (Simon, 1996, p. 53, cf. p. 12), and BR is part of this interface. Organisms and environment are not indistinct; rather, they are adaptively correlated.

Simon's model of rationality implied a unitary self, since the former had to embody any manifestation of intelligent behavior. In fact, part of Simon's monumental project was devoted to showing how phenomena such as intuition, insight, "aha" and creativity (and, in general, perception and emotions) were not mysterious, and that they implied neither vagueness nor tacit procedures (that is to say, procedures that cannot be explained in terms of symbolic information processes).[19] More generally, intentional BR (as a conscious activity of the decision-maker) and intuition can be considered the opposite poles of rational behavior. Yet, they only appear to be opposing cognitive tools; on the contrary, they share the same fundamental information processing mechanism.

5. Intuition and Insight: embodying the ineffable

Logical and intuitive thought can be easily unified; therefore phenomena like intuition, insight, "aha" and creativity can be explained by means of computer programs (Simon, 1997b, p. 174-5; 1983b, chap. 1).

Insight and intuition, which are near synonyms, appear when "someone solves a problem or answers a question rather suddenly […] without being able to give an account of how the solution

[19] Creativity is considered to be "a special class of problem-solving activity characterized by novelty, unconventionality, and difficulty in problem formulation" (Newell, Shaw, Simon, 1962, pp. 145-146). As regards perception, Simon maintains: "a wide range of the known phenomena of visual perception can be accounted for by a mechanism that encodes information from the stimulus into an internal representation having certain efficient characteristics from an information processing and retrieval stand-point." (Simon, 1967b, p. 361). In turn, emotion is considered to be 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

or answer was finally attained" (Simon, 1987, p. 482). Intuition is not a mysterious phenomenon, and consists in an act of "recognition" (see Frantz, 2003, p. 266). In particular, we recognize something of our previous experience, and as consequence of this recognition we arrive at the long-term memory where the knowledge of this phenomenon is located (Simon, 1997b, p. 179). Knowledge and information (about the familiar pattern and the cues for recognition) are stored and have a suitable and defined form.

Recognition (and intuition), as a process that is not accessible to consciousness and not reportable, characterized the EPAM program, which was able to report the symbols which reached its memory and, for example, to recognize a pathology.[20] In fact, the symbol representing the stimulus that has been recognized is placed in the EPAM short-memory, and hence provides access to knowledge about the stimulus that had been previously stored in the long-term memory (Simon, 1987, p. 482). In short, intuition is the recognition of past experience which takes the form of "stored information", it is guided by rules, and it explains the ability of chess grand masters, managers and, in general, experts who face situations requiring rapid decisions (Simon, 1983a, p. 4570; March and Simon, 1993, pp. 10-13).

Finally, 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)

In conclusion: because intuition is based on "stored information", it is not something else with respect to explicit information, which is normally processed by physical symbol systems. Therefore, intuition implies 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.

Yet, a potential problem arises. Intuition (as recognition of past experience or familiar patterns) is a process which permits individuals to respond rapidly when they encounter situations requiring an immediate reaction. In fact, since there is insufficient time to carefully evaluate alternatives, problem-solving processes cannot be applied. Yet, if there is no past experience to evoke (since the situation requiring urgent response is genuinely new), and the recognition process cannot be

---

mechanisms have been basically embodied in "the current information-processing theories of human cognition." (Simon, 1967a, p. 38). See Simon (1983b, pp. 29-30).

[20] EPAM program was in the process of construction in 1958 and was successively modified.

activated, what do individuals' reactions guide?[21] As a consequence, intuition probably involves different dimensions (see sect. 7.3).

6. The turn of 1980's

Simon, Newell and collaborators belonged to the first generation of AI scientists. Here it is not possible to examine the history of this discipline and compare the success of Simon's work both in economics and in cognitive science. Yet it should be borne in mind that a new strand of research, conceived as a paradigm alternative to AI, appeared in the mid-1980s: the so-called (new) connectionism, an approach based on neural networks.[22] Connectionism (first developed during the 1940s in studies by Hebb, McCulloch, Pitts, and others) reversed some basic perspectives of traditional AI. In particular, cognition ceased to be represented by means of symbolic processing, and the assumption inspiring LT and GPS programs – namely that intelligent behavior can be performed by following general strategies and principles of reasoning – was abandoned. In general terms, neural networks introduced a new way to treat mental representations. These latter were conceived as distributed, because on this view each element involves several units, and each unit participates in the representation of more than an element. Therefore, individual units codify information at subsymbolic levels, and units of networks work in parallel. Finally, the subsymbolic approach conceived cognitive processes as emergent, macroscopic, phenomena.

Connectionism contributed both to the decline of the computer metaphor and to establishing new relations with neurosciences. However, theoretical changes in both cognitive sciences and other disciplines contributed to modifying the image of rationality;[23] and this climate probably influenced scholars who had not been trained in AI theories, but were sensitive to the notion of BR.

All this can be summarized by citing Gardner's opinion (1987, p. 361) that Newell, Simon, and the first generation of AI scientists adopted a "decidedly rationalistic" model of man, since they were concerned with human capacities to reason correctly. But empirical research and new perspectives on cognition showed the inadequacy of rationalistic approaches, and Gardner includes Kahneman and Tversky, whose point of view will be treated in next sections, among the scholars who "developed stunning demonstrations of human departures from rationality".

[21] On these issues see Simon et al. (1988, pp. 27-30); Winograd and Flores (1986, p. 146).
[22] Fodor and Pylyshyn criticized connectionism, which nonetheless prevailed over other approaches.

7. Is Bounded Rationality Revisited?

Many scholars, from Nelson, Winter, and Williamson to March, Kahneman and Tversky have acknowledged their intellectual debt to Simon, and applied the notion of BR in their work. In turn, *some of them have received scientific recognition by Simon*, as is the case of Kahneman, Tversky, Nelson, Winter, March, and Gigerenzer. However, in contemporary studies BR seems to undergo a shift with respect to Simon's approach. In particular, these theories on BR prevalently describe cognitive (rational) processes as events weakly connoted in terms of computation, problem-solving ability, and instrumentality, where in Simon's view computation (as limited symbol processing), problem-solving processes and instrumentality have a major role. Moreover, 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, providing that highly structured form for the latter.

In this sense, as often happens with important concepts in science, their use has gradually led to changes with respect to the original meaning. In this case, the outcome is a notion of BR that is much more bounded than Simon's, and largely uninfluenced by the formal structure of *physical symbol systems*, although in many respects the latter provided the foundation for BR. 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 its technical and conceptual characteristics (although its basic assumptions are generally accepted), and that 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 survey provides some examples.

[23] For example, very influential was the work of Eleanor Rosch (especially her prototypical theory) and her criticism of the classical notion of categorization, a view, this latter, "consonant with the aura of cognitive science – a science built upon the unambiguity of the computer" (Gardner, 1987, p. 342).

7.1 James March: problematic aspects

A re-reading of the concept of BR appears in the work of J. March, Simon's well-known collaborator, who pointed out that new directions have emerged from the original Simonian context.[24] 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 induces us to reverse 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 behaviors, since individuals often elaborate an interpretation of what they are doing in the course of their actions (March 1988; 1994; see Slovic 1995; Tversky, 1996). 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

[24] "The original [Simon's] articles suggested small modifications in a theory of economic behavior, the substitution of bounded rationality for omniscient rationality. But the ideas ultimately have led to an examination of the extent to which theories of choice might subordinate the idea of rationality altogether to less intentional conceptions of the causal determinants of action" (March, 1978, p. 591), and, in this sense, he indicates: March and Olsen (eds.), *Ambiguity and Choice in Organizations* (1976).

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, 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 space problem, 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: "[h]uman beings have unstable, inconsistent, incompletely evoked, and imprecise goals" (March, 1987, p. 598), and both the world and the "self" are ambiguous (March, 1994). Many elements of March's work are a direct consequence of Simon's view; both authors, for example, emphasize the role of search processes. Nonetheless, it is possible to note a theoretical shift that, in March's case, leads to relaxation of some formal elements of BR: on the one hand, the "self" (or more precisely the problem solver) is well-structured and stable, even if computationally limited; on the other hand, the "self" does not possess a sound structure, and this is the most important limit.

7.2 Nelson and Winter: bounded rationality and tacit knowledge

Nelson and Winter, in their famous book *An Evolutionary Theory of Economic Changes* (1982), refer to the notions of Michael Polanyi's "tacit knowing", and of Simon's BR, as basic assumptions for their theoretical perspective. On the one hand, the "strong" influence of Polanyi is largely acknowledged, especially in chapters 4 and 5; on the other, Simon's concepts of BR and of "satisficing" are explicitly embodied in their theory (Nelson and Winter, 1982, pp. 35-36).[25]

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, and tacit knowledge and BR do not imply each other (Foss, 2003; Nightingale, 2003). Since tacit knowledge cannot be expressed in words (or by other symbols), it cannot be reduced to information processing in Simon's sense.

[25] Nelson states: "the assumption of bounded rationality of course lies at the heart of the evolutionary theory of economic change that I developed with Sidney Winter (Nelson and Winter, 1982)" (Nelson, 2005, p. 3; see Conlisk, 1996, pp. 677-678).

"Objectivists"[26] like Newell and Simon would argue that unarticulated knowledge becomes codified[27] (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. As a consequence, Simon's recourse to "protocol analysis"[28], which implicitly assumes that "*things that could be articulated* […] *were the only things involved in problem solving*" (Nightingale, 2003, p. 162, note 29), is useless.

These arguments can be analyzed in more detail. In particular, Simon's opposition to 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 (1962), where Simon provides an answer to a *Meno* paradox different from Plato's (a topic recurrent in his work). According to Plato, Simon says, we can discover or recognize the answer to a problem only if we already know the answer; therefore learning is remembering. Simon's explanation of the paradox is different:

> We pose a problem by giving the state description of the solution. The task is to discover a sequence of processes that will produce the goal state from an initial state. Translation from the process description to the state description enables us to recognize when we have succeeded. The solution is genuinely new for us […] (Simon, 1962, p. 479)

In terms of Simon's AI, once the "problem space" has been defined, problem solving consists in a heuristic (selective) search with a move "generator" and a "test" for potential solutions.

In his turn, Polanyi (1967, pp. 22-24) resolves the *Meno* paradox by resorting to tacit knowledge, i.e., the concept according to which "*we can know more than we can tell*" (Polanyi, 1967, p. 4), since most of our knowledge cannot be expressed in words. More specifically, the paradox shows that knowledge is not entirely explicit, and "[t]hat kind of tacit knowledge that solves the paradox

[26] Nightingale uses this term in reference to Gerald Edelman's view that "objectivism" is characterized by models "which are seen in mathematical logic [and they] consist of symbolic entities appearing singly or in sets, together with their relationship. Symbols in these models are made meaningful (or are given semantic significance) in a unique fashion by assuming that they correspond to entities and categories in the world" (Edelman, 1992, p. 232). As is well-known, Edelman's criticism is addressed to cognitive sciences: in particular, its tenets that 1) cognition consists of the manipulation of symbols governed by a unambiguous sets of rules (a syntax); 2) syntactic rules constitute a form of computation independent from the actual organization of the nervous system; 3) language can be represented in terms of an universal grammar in accordance with an intellectual tradition which includes Frege, Peano, Whitehead, Russel, Turing and other logicians and mathematicians (Edelman, 1992, pp. 13-15).

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

[28] In accordance with Simon's AI theoretical hypotheses, "verbal protocols" (protocol analysis) that the subjects of an experiment produce when they solve formal problems have special relevance. By means of this method, subjects verbally explain the steps that led them to the problem's solution, following the rules provided by the experimenter. In this way, a correspondence is assumed between the conscious (and declared) course followed to produce answers and the thought process. Criticisms on this method are in LeDoux (1996).

of the *Meno* consists in the intimation of something hidden, which we may yet discover" (Polanyi, 1967, pp. 22-23)[29].

Simon (1976a) criticized this interpretation of the paradox, and the "false" premise that "if you know what you are looking for, there is no problem"[30]. 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, as we shall see, human behaviors and choices can be fulfilled only if a certain amount of knowledge remains tacit.

Despite these epistemological divergences, Nelson and 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 both Polanyi's and Simon's original perspectives, and the result of this fruitful cross-fertilization is a notion of (bounded) rationality different from Simon's.

In Nelson's and 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. Yet, tacitness is also "matter of degree", since "The same knowledge […] is more tacit for some people than for others" (Nelson and Winter, 1982, p. 78), and this poses a problem if knowledge is only silent, but capable of articulation (see Cowan et al. 2000). Moreover,

[29] "[T]he *Meno* shows conclusively that if all knowledge is explicit, i.e., capable of being clearly stated, then we cannot know a problem or look for its solution. And the *Meno* also shows, therefore, that if problems nevertheless exist, and discoveries can be made by solving them, we can know things, and important things, that we cannot tell" (Polanyi, 1967, p. 22).

[30] This is Polanyi's summary of the first part of Plato's paradox, the second part, Polanyi says, affirms that if "you do not know what are looking for […] then you cannot expect to find anything" (Polanyi, 1967, p. 22).

much operational knowledge remains tacit because it cannot be articulated fast enough, because it is impossible to articulate all that is necessary to a successful performance, and because language cannot simultaneously serve to describe relationships and characterize the thing related" (Nelson and Winter, 1982, pp. 81-82)

Note that the above argument of rapidity for knowledge articulation is not central to Polanyi's theory, while other interesting arguments, which show further differences between Polanyi and Simon, are not examined by the authors. In Polanyi's view, ability and skills depend on the cooperative action of "subsidiary awareness" and "focal awareness" which characterizes even very simple courses of action such as, for example, hammering a nail into a wall. The former kind of awareness denotes the feeling of the fingers, the latter refers to conscious components of the action. Subsidiary awareness unconsciously helps control a number of details (i.e., those unconsciously perceived by the fingers) which are not (and cannot be) the object of our deliberate (focal) attention. The two kinds of awareness are "mutually exclusive", in fact, focal awareness cannot replace subsidiary awareness: the attempt consciously to control every particular would produce paralysis. A pianist or an orator engaged in deliberately checking every detail will necessarily stop their performance, since they will lose the sense of context indispensable for achieving their course of action. In short, the success of an action depends on cooperation between the two mechanisms, whose functions are distinct, while their overlap generates dangerous effects (Polanyi, 1962, pp. 55-57). Moreover, all this explains why some parts of our knowledge must remain tacit, and others not, although they are comprised in a unified system.

From this perspective, tacitness is rooted in our neurological system and explains how knowledge actually works, allowing the accomplishment of courses of action and behaviors. A novelty in Nelson's and Winter's approach with respect to Polanyi's 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 in a sense a tradeoff between capability and deliberate choice, a tradeoff imposed ultimately by the fact that rationality is bounded" (Nelson and Winter, 1982, p. 85).

More precisely, rationality is limited in that skills can substitute conscious deliberation,[31] work tacitly, and often select available options, so that deliberate choices play a narrow role.[32] Skills are complex entities which cannot be intentionally manipulated; therefore, "there is inevitably some

[31] "[M]uch of the knowledge that underlies the effective performance is tacit knowledge of the organization, not consciously known or articulable by anyone in particular" (Nelson and Winter, 1982, p. 134).
[32] "We […] emphasize that *automaticity* of skillful behavior and the suppression of choice that this involves. In skillful behavior, behavioral options are selected, but they are not deliberately chosen" (Nelson and Winter, 1982, p. 94).

ambiguity regarding the scope of a skill"; such ambiguity can be reduced, yet part of it "is simply irremediable" (Nelson and Winter, 1982, pp. 88-89).

Given these premises, the notion of tacit knowledge incorporates and explains BR, *conferring it with a new dimension*: 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 authors' different perspective: Nelson and Winter prevalently 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.[33] 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.

In Nelson's and 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 *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 to found the notion of BR. On the contrary, Nelson and Winter define BR on the base of tacit knowledge. This modification is seen as a complement to Simon's theory; yet it also implies a change in the original meaning of BR.[34] Now rationality, however limited, is an entity which must be represented as a complex mechanism involving tacit elements incorporated in skills, and not only as the product of the intentional mind.

---

[33] "[T]he 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, 1957a, p. 199; emphasis added).

[34] For further discussion on the "alternative interpretation of the concept" of BR in Nelson's and Winter's work see Klaes and Sent (2005, p. 48).

7.3 Kahneman's perspective

In 2002 the Nobel Prize for economics was awarded to Daniel Kahneman, a psychologist who, with Amos Tversky (who died in 1996), made fundamental contributions to the theory of human judgment and decision-making. This award implies that cognitive (and behavioral) economics has been recognized as an important part of economic science, and that this latter has included within its movable boundaries new theoretical tools and methodologies. [35]

In a recent article, Kahneman has summarized the most important phases of his research conducted with Amos Tversky for several decades. The "guiding ideas" are that "most judgments and most choices are made intuitively, [and] that the rules that govern intuition are generally similar to the rules of perception"[36]. He adds that intuition acts "*without conscious search or computation*" (Kahneman, 2003, p. 1450, emphasis added). In fact, he distinguishes between *two alternative* modes of thinking and deciding: reasoning and intuition, where intuition often prevails over reasoning. By way of contrast, as shown in Section 5, according to Simon intuition and rational thought are a *unique* cognitive mechanism in which computation is fundamental, even if it is characterized as a limited capacity.[37] 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 affirms, perception and intuition are collected in "System 1", whose operations are fast, parallel, automatic, effortless, associative, slow-learning, emotional, while reasoning belongs to "System 2", which is characterized by slow, serial, controlled, effortful, rule-governed, flexible, neutral operations (see also Zajonc, 1980; Khaneman and Frederick, 2002;

[35] Some scholars maintain that Kahneman and Tversky have not abandoned the neoclassical model, rather they use it as a benchmark, showing its empirical inadequacy (cf. Camerer, 1999; Sent, 2004; Motterlini and Guala, 2005). Other scholars 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"; therefore this evidence has "led to hypotheses of bounded rationality" (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" (Simon, 1979, p. 506), and deemed that "They are certainly important for the future of behavioral economics" (Simon 1999, p. 27, quoted in Augier and Kreiner, 2000, p. 664, note 2). In turn, Tversky and Kahneman recognize that the results of their analysis "are consistent with the conception of bounded rationality originally presented by Herbert Simon" (Tversky and Kahneman, 1987, pp. 88-89). Yet, in my view, the recent developments of Kahneman's research raise interesting new questions with regard to both the neoclassical and Simon's approaches.

[36] This perspective also appeared in "framing theory", where "framing effects resemble visual illusion more than computational errors" (Tversky and Kahneman, 1987, p. 76).

[37] "Khaneman and Tversky were convinced that the processes of intuitive judgments were not merely simpler than rational models demanded, but were categorically different in kind." (Gilovich and Griffin, 2002, p. 3).

Gilovich and Griffin, 2002; Sloman, 2002; Stanovich and West, 2002; Camerer, Loewenstein and Prelec, 2005).

The two systems do not only have complementary functions (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" (Khaneman and 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. Hence, there are no reasons for a divergence between the two sides (automatic and deliberate) of rationality, which, on the contrary, may emerge if these sides are viewed as distinct (although correlated) mechanisms whose origin probably depends on evolution (Zajonc 1980; Slovic et al. 2002; Stanovich and West, 2002; Kahneman, 2003; Greene et al. 2004; De Martino et al. 2006; Sanfey et al. 2006).

All this focuses the attention on both the role of heuristics and the relations between decision-making and emotions.

In Simon's view, heuristics are "rules of thumb" used to simplify choices, given the limits of rationality. Simonian agents are problem-solvers, and they adopt deliberate strategies to solve problems. This perspective has been treated by many authors. Yet, as regards Kahneman's and Tversky's approach, it has been pointed out that "deliberate choice heuristics differ substantially from the judgmental heuristics of the [Kahneman's and Tversky's] '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 (cf. Shafir and LeBoeuf, 2002, p. 494). Moreover, this perspective has recently led to a new, more comprehensive, and unified version of Tversky's and Kahneman's previous studies on judgment heuristics which has culminated in the notion of “attribute substitution”[38] (Kahneman and Frederick, 2002, p. 81).

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 “aspiration level”.[39] 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 and explicit) assessment. They govern, so to speak, both the “internal representation” of the problem, and the ensuing answers, with no recourse to deliberate and sequential problem-solving processes. In short, pre-analytic factors affect rational decisions; and *good reasons*, that is, pre-analytic criteria connoted by psychological attitudes (explainable in terms of “framing” effect, “status-quo bias”, “loss aversion”, “endowment effect”, etc.), determine choices without previous rational evaluations.[40] Also intuition can be observed in this perspective, since “[t]he central characteristic of agents is not that they reason poorly but that *they 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).[41] Sometimes intuition (not conceived as stored symbolic information), and not computation, provides good reasons for making choices.

Summarizing, in many circumstances agents do not act as problem solvers when they are faced by decisional problems. As a consequence, the term “heuristic” (and biases) in Khaneman’s and Tversky’s approach (including studies of scholars who refer to their teaching) denotes a different perspective with respect to Simon, who uses the same word.

For the former, heuristic is a simplifying rule which leads to “systematic errors” (Tversky and Kahneman, 1974; Khaneman, Slovic and Tversky, 1982), whereas for the latter heuristic is a simplifying rule which helps agents look for “satisficing” alternatives. Yet, these differences can be explained by considering the different perspectives which lie behind these approaches. In Kahneman’s and Tversky’s theory, heuristic generates errors from the point of view of the observer, who evaluates the departure from the normative rational theory. In Simon’s approach, the heuristic

[38] Attribute substitution determines the structure of a judgment. When the target attribute of an object is relatively inaccessible to the mind, then the search for it retrieves the value of some other attribute, which is conceptually and associatively related, and it is not refused by System 2.

[39] Aspiration levels, Selten convincingly argues, are not permanently fixed, but change in accordance with the situation (Selten, 2002). Yet, it is possible to imagine circumstances in which no aspiration level has been formulated. Cf. Simon (1979, p. 503).

[40] See Kahneman, Knetsch, Thaler (1991), where the authors explain the relations among 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). Cf. Motterlini (2006).

[41] Intuition is linked to the notion of “accessibility”, which helps us “to understand why some thoughts are accessible and others not”, and when some attributes are more accessible than others, without effort (Kahneman, 2003, p. 1452).

leads to "satisficing" outcomes from the agent's point of view (even if this issue could be considered an error by the observer).

However, heuristics which depend on System 1 do not always lead individuals into error: in fact, "there are situations in which choices are governed by (and simplified by) intuitive impressions" and exhibit a perfectly adequate heuristic as, for example, "choosing by liking" (i.e., a choice made by spontaneous affective evaluations), and "choosing by default" (i.e., choosing the option that first comes to mind) (Frederick, 2002, p. 550).

Kahneman and 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 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. In general, I think, this 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". In Kahneman's (and collaborators') view, the mind's operations, cognitive processes, and (bounded) rationality tend to be explained by considering how the two systems interact, and the outcomes, in terms of judgments and choices, are the result of this interaction (Sanfey et al. 2006). 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 a derivation from the interaction with System 1. As neurobiologists maintain, the mind is a whole, whose parts are closely integrated, and this issue probably depends on its evolution. In this perspective, (limited) rationality is not based only on overt (or instrumental) reasoning; it also works by means of unconscious and automatic mechanisms. The two systems are complementary and cannot be dealt with as independent mechanisms, although they perform different functions.[42] It is therefore unsurprising that

[42] System 1 "competes" with System 2, sometimes it produces "the initial error, and System 2 fails to correct it, although it could" (Kahneman and Frederick, 2002, p. 78), in other cases, choices are mediated by spontaneous ("natural") assessments (Tversky and Kahneman, 1983, p. 20), which do not lead to errors. All this gives rise to the formulation of unconscious assessments which are not correlated to the subsequent satisfaction (Frederick, 2002, p. 550).

Kahneman includes “the large role of System 1” among the “core ideas” of his essay (Kahneman, 2003, p. 1469), and neglecting this aspect probably means elaborating another model of mind.

Finally to be examined is 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 and 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 (see Slovic et al., 2002).[43] According to Slovic, 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.[44]

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 (System 2) (Kahneman, 2003, p. 1463). Yet, in general, the pervasive role of emotions enables Kahneman to re-think large part of his research with Tversky. For example, emotions are involved in the significance of accessibilty, and are invoked in *prospect theory* because “utility cannot be divorced from emotion, and emotions are triggered by changes” (Kahneman, 2003, 1457).[45] Therefore emotions play a wider role than in Simon’s theory, where they are described in symbolic terms as a mechanism able to help agents select and focus conscious attention on specific objects, and consequently define priorities in the search associated with problem-solving activity (Simon, 1967a; 1983b; cf. note 20).

In addition, predictions of *prospect theory* have received support from recent neurobiological research (in which Kahneman has taken part) based on functional magnetic resonance imaging (fMRI) (Breiter et al., 2001). This technique (fMRI) 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”. In fact, amygdala work provides the emotional signal which induces subjects to keep certain money and gamble instead of taking a

[43] On these themes, see Sanfey et al. (2003).
[44] “[W]e make many claims for the affect heuristic, portraying it as the centerpiece of the experiential mode of thinking, the dominant mode of survival during the evolution of the human species” (Slovic et al., 2002, p. 416).
[45] In fact, *prospect theory* affirms that perception of utility depends on changes (gains and losses) relative to a neutral reference point, instead of states of wealth or welfare.

loss (Miller, 2006). Moreover, their findings show that the orbital and medial prefrontal cortex (OMPFC) can moderate the influence of emotions on decisions (De Martino et al., 2006).[46]

The experiment confirms the hypothesis of two neural systems (and of their conflicts), performing diverse functions, but also their robust correlation.[47] In this regard, "It is noteworthy that there are *strong reciprocal connections* between the amygdala and the OMPFC, although each may contribute to distinct functional roles in decision-making" (De Martino et al., 2006, p. 687),[48] and Kahneman's comment on the experiment, is telling: "The results could hardly be more elegant" (quoted in Miller, 2006, p. 600).

In short, Kahneman's and Tversky's studies received further confirmation on neurobiological grounds, not on Simon's AI. The instruments of these disciplines are different as regards their theoretical assumptions and methodological approaches, and they sometimes give rise to opposite conclusions with respect to AI (Edelman, 1994).[49] This provides an argument in support of the thesis that Kahneman's and Simon's theories should be considered more as different models than as complementary approaches (cf. Kaufman, 1999), although they share some general assumptions.

As well-known, Antonio Damasio (1994) has provided an important account of the role of emotions in decision making. His theory on the function on somatic markers leads to the conclusion that these latter increase efficiency in decision processes. He also suggests with his collaborators that non-conscious biases guide behavior prior to conscious knowledge, and "Without the help of such biases, overt knowledge may be insufficient to ensure advantageous behavior" (Bechara et al., 1997, p. 1293; 2005). Non-conscious biases assist reasoning in "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 re-appears, now in terms of cooperation, and it highlights how BR is not only a product of a limited analytically mind, but of intuitive, emotional, non-conscious mind as well.[50]

---

[46] The orbito-frontal cortex is a structure linked to dorsolateral prefrontal regions active in reasoning and planning, and amygdala.

[47] "When subjects' choices ran counter to their general behavioral tendency, there was enhanced activity in the ACC [anterior cingulate cortex]. This suggests an opponency between two neural systems, with ACC activation consistent with the detection of conflict between predominantly 'analytic' response tendencies and a more 'emotional' amygdala-based system." (De Martino et al., 2006, p. 687). The anterior cingulate cortex is considered an area able to reveal cognitive conflicts.

[48] See Sanfey et al. (2006, pp. 111-112); Greene et al. (2004).

[49] Kahneman's and collaborators' researches, and neuroscience studies referring to their theories, show a tendency to study human decision-making resorting to neurobiology to understand how the nervous system is organized and actually works. This approach differs from the one which assumes that "since the thinking human being is also an information processor, it should be possible to study his processes and their organization independently of the details of the biological mechanism" (Simon, 1964, pp. 76-77).

[50] Emotions, Damasio and colleagues maintain, play a fundamental role in the generation of moral judgments. In particular, the findings of their research "are consistent with a model in which a combination of intuitive/affective and conscious/rational mechanisms operate to produce moral judgements" (Koenigs et al., 2007, p. 3).

In conclusion, Kahneman's "maps of bounded rationality" seem to define a notion 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 and Frederick, 2002). As is well-known, *prospect theory* was presented as a formal descriptive theory of human choices. Yet Kahneman mentions certain mechanisms which can explain decision making, referring, for example, to the notion of "accessibility", which is linked to emotions, intuitions, and perceptions. In short, the mechanism evoked to explain choices and behaviors seems to go beyond the computational limits to information processing. What is the relevant information which must be processed? It depends on complex processes involving intuition, perception and emotions as alternative elements which condition and sometimes replace computational reasoning.

7.4 Ecological rationality

Gerd Gigerenzer (a German psychologist) and Reinhard Selten (1994 Nobel Prize for economics) also start from Simon's vision of BR, and maintain that their work is an elaboration of his theory (Gigerenzer and Selten, 2002, p. 4). In fact, Simon introduced the notion of "satisficing", and pointed out two sides of BR: one cognitive and one ecological, which in turn emphasized that minds are adapted to the real world environment.[51] The elaboration of these ideas has led to the notions of "ecological rationality" and "adaptive toolbox", which have characterized the research program of Gigerenzer and his group (Gigerenzer *et al.*, 1999).

This perspective is summarized as follows:

> 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. (Gigerenzer and Selten, 2002, p. 9)

[51] Gigerenzer and collaborators often refer to the two scissors blades which, according to Simon (1990, p. 7), shape human rationality: task environments and computational capabilities. But they recognize that "[t]he neglect of one [the

Ecological rationality, as a new version of BR, exhibits some shifts with respect to the original model. This kind of rationality is not instrumental, in the sense that the focus is centered less on "*intendedly rational behavior*" (Newell and Simon, 1972, p. 55), however limited computationally, and more on a rationality which emerges as a strategy adapted to a changing environment.[52] This view also suggests that (limited) rationality is less characterized by judgments and choices of an analytical (limited) mind, and is more unintentional, evolutionary, and adaptive. In fact, *intentional* (limited) rationality implies focusing attention on conscious decisional activity, whilst adaptive rationality implies a decision-making activity largely connoted by unconscious, adaptive mechanisms. If this is so, in the former case individuals are considered intentional problem solvers whose explicit rationality is limited; in the latter, individuals are adaptive problem solvers connoted by scarce degrees of awareness as regards their responses to the environment.

Moreover, 1) heuristics are fast and frugal, because they do not involve much computation and "only search for some of the available information" (Gigerenzer and Todd, 1999, p. 4); 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 provide an 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).

These characteristics are grouped into four main classes: "ignorance-based decision making" (which is well represented by "recognition heuristic"), "one-reason decision making", "elimination heuristics", and "satisficing heuristics for sequential search". In the first heuristic, the role of "ignorance" is specifically pointed out as an essential element for making decisions;[53] the second heuristic relies "on just a single cue to make a decision"; the third too is based on little information[54], while the fourth is closely connected to Simon's approach (Todd and Gigerenzer, 2003, pp. 149-150). Limits of information (for example, ignorance) do not hamper decisional

structure of environment] of the two sides of bounded rationality can even be traced in Simon's writings", although that "side" was implicitly assumed (Gigerenzer, 2002, p. 40).

[52] In fact, "rationality is defined by its fits with reality" (Gigerenzer and Todd, 1999, p. 5).

[53] "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 and Gigerenzer, 2003, p. 150). In short, "The recognition heuristic can *only* be applied when one of the two objects is not recognized, that is, under partial ignorance" (Goldstein and Gigerenzer, 1999, p. 41). In this circumstances, ignorance is beneficial.

[54] "*Categorization by Elimination* […] assigns a stimulus to a particular category by using one cue after another in a specified order to narrow down the set of remaining possible categories until only a single one remains" (Todd and Gigerenzer, 2003, p. 150).

processes; on the contrary, they help them, especially in the recognition heuristic.[55] 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 and Gigerenzer, 1999, p. 57, emphasis added). A basic evolutionary mechanism (a "capacity" evolved over time) 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 (Khaneman and Frederick, 2002, pp. 59-60).

Cognitive limits (of information and of computation) characterize Simon's concept of BR, and do not seem advantageous for human beings. By contrast, "ecological rationality" delineates a perspective in which "benefits of cognitive limits" take shape, and this suggests "that simple heuristics had a selective advantage over more complex cognitive strategies" (Todd and Gigerenzer, 2003, p. 161).

Gigerenzer and collaborators maintain that just *two forms of BR* exist: "satisficing heuristics", and "fast and frugal heuristics" (Gigerenzer and Todd, 1999, p. 7). In this way, both continuity and the differences between these approaches are explicitly pointed out. 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 one again be remarked upon. 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). In particular, ecological rationality "suggests looking outside the mind, at the structure of environments, to understand what is inside the mind" (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 and Todd, 1999, p. 20)

[55] "There is a point where too much information and too much information processing can hurt" (Gigerenzer and Todd, 1999, p. 21).

In a certain sense, computational limits (with respect to the information which should be gathered and processed) produce a counter-intuitive effect, since "computationally simple strategies" are better than "computationally complex" ones. On the contrary, in Simon's view, given the computational limits of processing information, the more such limits are reduced, the more subjects are rational, and the more they carry out good strategies for solving problems. In short, instrumental and intentional rationality is weak (more than in Simon's view), since the success of a course of action does not essentially depend on its 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. As a consequence, on the one hand Simon's agents merely adopt heuristics connoted by "satisficing" criteria; on the other, Gigerenzer's individuals adopt adaptive heuristics which enable them to accomplish "accurate inferences". Therefore, instrumental and intentional rationality (in Simon's sense) is weak, while ecological rationality is strong.[56]

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 and 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). Moreover, Simon's project to discover and explain an universal language of thought by means of physical symbol logic (including human and computer information processing) seems to reproduce in new form "Leibniz's dream" (see sect. 4), which consisted in an alphabet of human thought, and in a calculus of reasoning to combine the elements of this alphabet. But Leibniz's dream is incompatible with Gigerenzer's approach, since the universal language of thought refers to a general-purpose tool, while the "adaptive toolbox" idea assumes that "different domains of thought require different specialized tools" (Gigerenzer and Todd, 1999, p. 30; see Gigerenzer and Selten, 2002 p. 9). Nonetheless, as in the case of Kahneman and Tversky, fast and frugal heuristics can be seen as a distinctive complement to (or development of) Simon's works.

---

[56] This view, in turn, is linked to the criticisms against Kahneman and Tversky. In fact, simple heuristics lead to reasonable decisions and accurate inferences, "[i]n contrast, the [Kahneman's and Tversky's] heuristics-and-biases approach views heuristics as unreliable aids that the limited human mind too commonly relies upon despite their inferior decision-making performances, hence […] heuristics can be blamed for poor reasoning." (Gigerenzer and Todd, 1999, p. 28). The consequences of the debate among these scholars evidently cannot be discussed here. See Gigerenzer (1991); Kahneman and Tversky (1996); Gigerenzer (1996).

Conclusions

The simple assumption from which we started is that Simon's notion of BR is closely connected to scientific research in AI**.** This was the result of an intellectual process which enabled BR to be considered as a paradigm alternative to standard economics, and to redefine the nature of decisional processes. This is not a novelty; nonetheless it has certain theoretical consequences, the most important of which is the conception of BR as a highly structured concept, characterized in many respects in terms of physical symbol processing. Given this premise, it has been argued that formal logic was an approach able to influence AI by providing tools with which to conceive both symbol processing and symbol operationality. More generally, formal logic suggested the idea of a general language of thought. This influence, although limited, 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-twentieth century. Therefore, this heritage matters, as Simon himself recognized.

All these elements have been observed by focusing on a specific aspect: Simon's aim to explain phenomena as "intuition" and "insight" (and, more generally, perception and emotions) by means of AI. Simon's unitary vision of (bounded) rationality emerges from this perspective. Rationality works in essentially the same way when dealing with both formal (*closed*), and *open* or ill-defined problems, and it exhibits a homogeneous structure when it functions by sing 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; and it is computational. All these features are consistent with the most important one: rationality is bounded.

The second part of the paper has suggested that some shifts are apparent in certain theories which share Simon's criticism of standard economic theory. These works are not homogeneous, and they refer to both theoretical approaches and different methods.

A common thread seems to be the split between BR and AI. In particular, there is no reference to the rigorous structure of AI which contributed to defining BR. As a consequence, the unitary view of rationality derived from the AI perspective disappears and, for example, intuition, emotion, and perception can be separated from "reasoning", assuming a non-symbolic dimension. Computation is not so important if individuals act largely intuitively, that is, in a rapid, parallel, and emotional way. In short, intuition is apparently not only the recognition of past experience reduced to the form of physical symbols contained in short- and long-term memories of processors. Rather, intuition, in

Kahneman's and his collaborators' view, is a cognitive tool which is alternative to (but not separated from) deliberate reasoning, since it cooperates and/or competes with the latter. In addition, computation is less important if "fast and frugal" heuristics allow more accurate predictions than computationally complex strategies. On the contrary, in Simon's view, computation matters: rationality is bounded because computational capacities are limited; without such limits the individual could be Pareto's *homo oeconomicus*. In a trivial sense, the less computation capacities are limited, the more the individual is rational. Simon often invoked empiricism in studying how individuals make decisions, and computational limits in his view explain their performances. But it is possible to ask if, for example, emotions, as psychological, neurological, and biological factors so essential for decision-making, can be encapsulated in an AI computational theory.

Limits of information and information processing also play a marginal role in recent approaches to BR. Judgments and decisions often emerge by means of mechanisms which differ from the information processing mechanism described by Simon. Emotions, intuitions, and various pre-analytic (automatic) assessments often intervene before any rational, however limited, information processing, and they influence decision-making. Moreover, Gigerenzer and 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 colleagues state, limits of information can constitute an advantage, and "ignorance makes us smart".

The four approaches discussed here are very different, and some of them have produced fierce debate. Yet, in very general terms, there is a common thread that links them. Rationality is not conceivable only as a deliberate, intentional and instrumental activity realized by problem-solvers, however cognitively limited: this is only one side of the problem. Rationality works only if a deeper level of unconscious, neurobiological, and automatic structures (sometimes explained in evolutionary terms) act, determining or influencing the performances of conscious and deliberate reasoning. In this sense, Nelson's and Winter's reference to tacit knowledge, and March's emphasis on the instability and ambiguity of the world and of the self, provide interesting insights because they evoke a more complex description of rationality.

References

Arrow, K., 1996. *Preface*. In: Arrow, K., Colombatto, E., Perlman, M., Schmidt, C. (Eds.). *The Rational Foundations of Economic Behaviour*. London: Macmillan Press, xiii-xxvi.

Arrow, K., 2004. Is Bounded Rationality Unboundedly Rational? Some Ruminations. In: Augier, M., March, J.C. (Eds.). *Models of a Man: Essays in Memory of Herbert A. Simon*. Cambridge (Mass.): The MIT Press, 47-55.

Augier, M., 2000. *Models of Herbert A. Simon*. Perspective on Science 8, 407-443.

Augier, M., March, J., 2002. *A model scholar: Herbert A. Simon (1916-2001)*. Journal of Economic Behavior & Organization 49, 1-17.

Augier, M., Kreiner, K., 2000. *Rationality, Imagination and Intelligence: Some Boundaries in Human Decision-Making*. Industrial and Corporate Change 9, 659-681.

Aumann, R.J., 1997. *Rationality and Bounded Rationality*. Games and Economic Behavior 21, 2-14.

Bartlett, R.V., 1988. *Rationality in Administrative Behavior: Simon, Science, and Public Administration*. Public Administration Quarterly 12, 275-300.

Bechara, A., Damasio, H., Tranel, D., Damasio, A.R., 1997. *Deciding Advantageously Before Knowing the Advantageous Strategy*. Science 275, 28 February, 1293-1295.

Bechara, A., Damasio, H., Tranel, D., Damasio, A.R., 2005. *The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers*. Trends in Cognitive Sciences 9, April, 159-162.

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, May, 619-639.

Brown, R., 2004. *Consideration on the origin of Herbert Simon's theory of "satisficing" (1933-1947)*. Management Decision 42, 1240-1256.

Camerer, C. 1999. *Behavioral Economics*. CSWEP Newsletter, Winter. Website: http://www.cswep.org/camerer.htlm.

Camerer, C., Loewenstein, G., Prelec D., 2005. *Neuroeconomics: How Neuroscience Can Inform Economics*. Journal of Economic Literature 43, 9-64.

Conlisk, J., 1996. *Why Bounded Rationality?*. Journal of Economic Literature 34, 669-700.

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

Cowan, R., David, P.A., Foray, D., 2000. *The Explicit Economics of Knowledge Codification and Tacitness*. Industrial and Corporate Change 9, 211-253.

Crevier, D., 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: BasicBooks.

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

Damasio, A.R., 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam's Sons.

Davis, M., 2000. *The Universal Computer. The Road from Leibniz to Turing*. New York: Norton.

De Martino, B., Kumaran, D., Seymour, B., Dolan, R.J., 2006. *Frames, Biases and Rational Decision-Making in the Human Brain*. Science 313, 4 August, 684-687.

Dreyfus, H.L., Dreyfus, S.E., 1990. Ricostruire la mente o progettare modelli del cervello? L'intelligenza artificiale torna al bivio. In: Negrotti, M., (Ed.). *Capire l'artificiale: dall'analogia all'integrazione uomo-macchina*. Torino: Bollati Boringhieri, 29-59. English edition, 1991. *Understanding the artificial. On the future shape of Artificial Intelligence*. London: Springer-Verlag.

Edelman, G.E., 1992. *Bright Air, Brillant Fire: On the Matter of the Mind*. New York: Basic Books.

Edwards, P.N., 1997. *The Closed World. Computers and the Politics of Discourse in Cold War America*, paperback edition, Cambridge (Mass.): The MIT Press.

Egidi, M., Marengo, L., 2004. Near-Decomposability, Organization, and Evolution: Some Notes on Herbert Simon's Contribution. In: Augier, M., March, J.C. (Eds.). *Models of a Man: Essays in Memory of Herbert A. Simon*. Cambridge (Mass.): The MIT Press, 335-350.

Feigenbaum, E.A., 1989. What Hath Simon Wrought?. In: Klahr, D., Kotovsky, K., (Eds.). *Complex Information Processing: The Impact of Herbert A. Simon*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 163-182.

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, 2, pp. 185-201.

Frantz, R., 2003. *Herbert Simon. Artificial intelligence as a framework for understanding intuition*. Journal of Economic Psychology 24, 265–277.

Frederick, S., 2002. Automated Choice Heuristics. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 548-558.

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, 83-115.

Gigerenzer, G., 1996*, On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)*. Psychological Review, 103, 592-596.

Gigerenzer, G., 2002. The Adaptive Toolbox. In: Gigerenzer, G., Selten, R. (Eds.). *Bounded Rationality. The Adaptive Toolbox*. Cambridge (Mass.): The MIT Press, pp. 37-50.

Gigerenzer, G., Goldstein, D.G., 1996. *Reasoning the Fast and Frugal Way: Models of Bounded Rationality*. Psychological Review, 103, 650-699.

Gigerenzer, G., Todd, P.M., 1999. Fast and Frugal Heuristics. The Adaptive Toolbox. In: Gigerenzer, G., Todd, P.M., and the ABC Research Group. *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press, 3-34.

Gigerenzer, G., Selten, R., 2002. Rethinking Rationality. In: Gigerenzer, G., Selten, R. (Eds.), *Bounded Rationality. The Adaptive Toolbox*. Cambridge (Mass.): The MIT Press, 2-11.

Gilovich, T., Griffin, D., 2002. Introduction - Heuristics and Biases: Then and Now. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 1-18.

Goldstein, D.G., Gigerenzer, D., 1999. The Recognition Heuristic: How Ignorance Makes Us Smart. In: Gigerenzer, G., Todd, P.M., and the ABC Research Group. *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press, 37-58.

Golembiewski, R.T., 1988. *Nobel Laureate Simon "Looks Back": A Low-Frequency Mode*. Public Administration Quarterly 12, 275-300.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*. Neuron 44, October 14, 389-400.

Hargreaves Heap, S., Hollis, M., Lyons, B., Sugden, R., Weale, A., 1992. *The Theory of Choice. A Critical Guide*. Cambridge (Mass.): Blackwell.

Kahneman, D., 2003. *Maps of Bounded Rationality: Psychology for Behavioral Economics*. American Economic Review 93, 1449-1475.

Kahneman, D., Frederick, S., 2002. Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 49-81.

Kahneman, D., Knetsch, J.L., Thaler, R.H. 1991. *Anomalies: The Endowment Effect, Loss Aversion and Status Quo Bias*. Journal of Economic Perspectives 5, 193-206.

Khaneman, D., Slovic, P., Tversky, A. (Eds.) 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge and New York: Cambridge University Press.

Kahneman, D., Tversky, A., 1996. *On the Reality of Cognitive Illusions*. Psychological Review. 103, 582-591.

Kaufman, B.E., 1999. *Emotional arousal as a source of bounded rationality*. Journal of Economic Behavior & Organization 38, 135-144.

Klaes, M., Sent, E.-M., 2005. *A Conceptual History of the Emergence of Bounded Rationality*. History of Political Economy 37, 27-59.

Kline, M., 1972. *Mathematical Thought From Ancient to Modern Times*. Oxford: Oxford University Press.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., Damasio, A., 2007. *Damage to the prefrontal cortex increases utilitarian moral judgements*. Nature 21, March, 1-4.

LeDoux, J., 1996. *The Emotional Brain. The Mysterious Underpinning of Emotional Life*. New York: Simon and Schuster.

March, J.G., 1978. *Bounded Rationality, Ambiguity, and the Engineering of Choice*. The Bell Journal of Economics 9, 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*. 1st edition. New York: John Wiley.

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

McCorduck, P., 1979. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. San Francisco: Freeman.

Miller, G., 2006. *The Emotional Brain Weighs Its Options*. Science 313, 4 August, 600-601.

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

Motterlini, M., 2006. *Economia Emotiva*. Milano, Rizzoli.

Motterlini, M., Guala, F., 2005. Psicologia ed esperimenti in economia. In: Motterlini, M., Guala, F. (Eds.). *Economia cognitiva e sperimentale*. Milano: Università Bocconi Editore. 1-59.

Nelson R.R. (2005), *Bounded Rationality, Cognitive Maps, and Trial and Error Learning*, LEM Working Papers Series, 28, December, 1-36.

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

Newell, A., 1989. Putting It All Together. In: Klahr, D., Kotovsky, K. (Eds.). *Complex Information Processing: The Impact of Herbert A. Simon*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 399-440.

Newell, A., Simon, H.A., 1972. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.

Newell, A., Simon, H.A., 1981. Computer Science as Empirical Inquiry: Symbols and Search. In: Haugeland, J. (Ed.). *Mind and Design. Philosophy, Psychology, Artificial Intelligence*. Cambridge (Mass.): MIT Press, 31-66.

Newell, A., Shaw, J.C., Simon, H.A, 1962. The Process of Creative Thinking. In: Simon, H.A., 1979. *Models of Thought*, Vol. I. New Haven and London: Yale University Press, 144-174.

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, 2, pp. 149-183.

O'Neill, M., 2005. The Biology of Irrationality: Crime and the Contingency of Deterrence. In: Parisi, F., Smith, V.L. (Eds.). *The Law and Economics of Irrational Behavior*. Stanford: Stanford University Press, 287-313.

Polanyi, M., 1962. *Personal Knowledge: Towards a Post-Critical Philosophy*. London: Routledge & Kegan Paul.

Polanyi, M., 1967. *The Tacit Dimension*. Garden City, N.Y.: Doubleday Anchor.

Ridley, C.E., Simon, H.A., 1943. *Measuring Municipal Activities*. Chicago: The International City Managers' Association.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. *The Neural Basis of Economic Decision-Making in the Ultimatum Game*. Science 300, 13 June, 1755-1758.

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, 108-116.

Sargent, T.J., 1993. *Bounded Rationality in Macroeconomics*. Oxford: Oxford University Press.

Selten, R., 2002. What Is Bounded Rationality?. In: Gigerenzer, G., Selten, R. (Eds.), *Bounded Rationality. The Adaptive Toolbox*. Cambridge (Mass.): The MIT Press, 13-36.

Sent, E.-M., 1997. *Sargent versus Simon: bounded rationality unbound*. Cambridge Journal of Economics 21, 323-338.

Sent, E.-M., 2000. *Herbert A. Simon as a Cyborg Scientist*. Perspective on Science" 8, 380-406.

Sent, E.-M., 2004. *Behavioral Economics: How Psychology Made Its (Limited) Way Back Into Economics*. History of Political Economy 36, 735-760.

Shafir, E., Simonson, I., Tversky, A. 1993. *Reason-based choice*. Cognition 49, 11-36.

Shafir, E., LeBoeuf, R., 2002. R*ationality*. Annual Review of Psychology 53, 491-517.

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

Simon, H.A., 1951. *A Formal Theory of the Employment Relationship*. Econometrica 19, 293-305.

Simon, H.A., 1955. A Behavioral Model of Rational Choice. In: Simon, H.A., 1982. *Models of Bounded Rationality*. Vol. 2, Cambridge (Mass.): The MIT Press, 239-258.

Simon, H.A., 1956. Rational Choice and the Structure of the Environment. In: Simon, H.A., 1982. *Models of Bounded Rationality*. Vol. 2, Cambridge (Mass.): The MIT Press, 259-274.

Simon, H.A., 1957a, Rationality and Administrative Decision Making. In: Simon, H.A. 1957. *Models of Man*. New York: Wiley, 196-206.

Simon, H.A., 1957b. *Admnistrative Behavior*, 2nd ed., repr. in 1965. New York: The Free Press.

Simon, H.A., 1964. Information Processing in Computer and Man. In: Simon, H.A., Egidi, M., Marris, R., Viale R. (Eds.) 1992. *Economics, Bounded Rationality and the Cognitive Revolutions*. Aldershot: Elgar, 76-99.

Simon, H.A., 1962. *The Architecture of Complexity*. Proceedings of the American Philosophical Society 106, 467-482.

Simon, H.A., 1967a. Motivational and Emotional Controls of Cognition. In: Simon, H.A., 1979. *Models of Thought*, Vol. I. New Haven and London: Yale University Press, 29-38.

Simon, H.A., 1967b. An Information Processing Explanation of Some Perceptual Phenomena. In: Simon, H.A., 1979. *Models of Thought*, Vol. I. New Haven and London: Yale University Press, 353-362.

Simon, H.A., 1972. Theories of Bounded Rationality. In: Simon, H.A., 1982. *Models of Bounded Rationality*. Vol. 2, Cambridge (Mass.): The MIT Press, 408-423.

Simon, H.A., 1976a. *Bradie on Polanyi on the Meno Paradox*. Philosophy of Science 1, 147-151.

Simon, H.A., 1976b. From Substantive to Procedural Rationality. In: Simon, H.A., 1982. *Models of Bounded Rationality*. Vol. 2, Cambridge (Mass.): The MIT Press, 424-443.

Simon, H.A., 1976c . *Admnistrative Behavior*, 3rd ed. New York: The Free Press.

Simon, H.A., 1978. Rationality as Process and as Product of Thought. In: Simon, H.A., 1982. *Models of Bounded Rationality*. Vol. 2, Cambridge (Mass.): The MIT Press, 444-459.

Simon, H.A., 1979. *Rational Decision Making in Business Organizations*. American Economic Review 69, 493-513.

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, No 14, 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: Simon, H.A., 1989. *Models of Thought*. Vol. II, New Haven and London: Yale University Press, 481-493.

Simon, H.A., 1990. *Invariants of Human Behavior*. Annual Review of Psychology 41, 1-19.

Simon, H.A., 1991. *Models of My Life*. Repr. in 1996. Cambridge (Mass.), The MIT Press.

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

Simon, H.A., 1997a. *Admnistrative Behavior*, 4th ed. New York: The Free Press.

Simon, H.A., 1997b. *Spiegare l'ineffabile: l'intelligenza arificiale a proposito di intuizione, insight e ispirazione*. Sistemi intelligenti 9, 169-192.

Simon, H.A., 1999. *Interview with Herbert A. Simon*. Working paper, CISTEMA, Copenhagen Business School.

Simon, H.A., 2001. *On Simulating Simon: His Monomania and its Sources in Bounded Rationality*. Studies in the History and Philosophy of Science 32, 501-505.

Simon, H.A., Shaw, J.C., 1958. Elements of a Theory of Human Problem Solving. In: Simon, H.A., 1989. *Models of Thought*. Vol. II, New Haven and London: Yale University Press, 7-19.

Sloman, S.A., 2002. Two Systems of Reasoning. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 379-396.

Slovic, P., 1995. *The Construction of Preference*. American Psychologist 50, 364-371.

Slovic, P., Finucane, M., Peters, E., MacGregor, D.G., 2002. The Affect Heuristic. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 397-420.

Stanovich, K.E., West, R.F. 2002. Individual Differences in Reasoning: Implications for the Rationality Debate?. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 421-440.

Stigler, G.J., 1961. *The economics of information*. Journal of Political Economy 69, 213-225.

Todd, P.M., 2002. Fast and Frugal Heuristics for Environmentally Bounded Minds. In: Gigerenzer, G., Selten, R. (Eds.), *Bounded Rationality. The Adaptive Toolbox*. Cambridge (Mass.): The MIT Press, 51-70.

Todd, P.M., Gigerenzer, G., 2003. *Bounding Rationality to the world*. Journal of Economic Psychology 24, 143-165.

Tversky, A., Kahneman, D., 1974. *Judgment under Uncertainty: Heuristics and Biases*. Science 185, 1124-1131.

Tversky, A., Kahneman. D., 1983. Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.) 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 19-48.

Tversky, A., Kahneman, D., 1987. Rational Choice and the Framing of Decisions. In: Hogarth, R.M., Reder, M.W. (Eds.). *Rational Choice. The Contrast between Economics and Psychology*. Chicago and London: The University of Chicago Press, 67-94.

Tversky, A., 1996. Rational Theory and Constructive Choice. In: Arrow, K., Colombatto, E., Perlman, M., Schmidt, C. (Eds.). *The Rational Foundations of Economic Behaviour*, London, Macmillan Press, 185-197.

Winograd, T., Flores, F., 1986. *Understanding Computers and Cognition. A New Foundation for Design*. Norwood – New Jersey: Ablex Publishing Corporation.

Zajonc, R.B., 1980. *Feeling and Thinking: Preferences Need No Inferences*. American Psychologist 35, 151-175.