Logic and Knowledge

Edited by

Carlo Cellucci, Emily Grosholz
and Emiliano Ippoliti

CAMBRIDGE
SCHOLARS
PUBLISHING

Logic and Knowledge,
Edited by Carlo Cellucci, Emily Grosholz and Emiliano Ippoliti

This book first published 2011

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-3008-9, ISBN (13): 978-1-4438-3008-9

TABLE OF CONTENTS

Foreword ix

Acknowledgements xxv

Section I: Logic and Knowledge

Chapter One 3
The Cognitive Importance of Sight and Hearing in Seventeenth-
and Eighteenth-Century Logic
Mirella Capozzi

Discussion 26
Chiara Fabbrizi

Chapter Two 33
Nominalistic Content
Jody Azzouni

Discussion 52
Silvia De Bianchi

Chapter Three 57
A Garden of Grounding Trees
Göran Sundholm

Discussion 75
Luca Incurvati

Chapter Four 81
Logics and Metalogics
Timothy Williamson

Discussion 101
Cesare Cozzo

Chapter Five 109
Is Knowledge the Most General Factive Stative Attitude?
Cesare Cozzo

Discussion 117
Timothy Williamson

Chapter Six 123
Classifying and Justifying Inference Rules
Carlo Cellucci

Discussion 143
Norma B. Goethe

Section II: Logic and Science

Chapter Seven 151
The Universal Generalization Problem and the Epistemic Status of Ancient Medicine: Aristotle and Galen
Riccardo Chiaradonna

Discussion 168
Diana Quarantotto

Chapter Eight 175
The Empiricist View of Logic
Donald Gillies

Discussion 191
Paolo Pecere

Chapter Nine 197
Artificial Intelligence and Evolutionary Theory: Herbert Simon's Unifying Framework
Roberto Cordeschi

Discussion 216
Francesca Ervas

Chapter Ten 221
Evolutionary Psychology and Morality: The Renaissance of Emotivism?
Mario De Caro

Discussion 232
Annalisa Paese

Chapter Eleven 237
Between Data and Hypotheses
Emiliano Ippoliti

Discussion 262
Fabio Sterpetti

Section III: Logic and Mathematics

Chapter Twelve 273
Dedekind Against Intuition: Rigor, Scope and the Motives of his Logicism
Michael Detlefsen

Discussion 290
Marianna Antonutti

Chapter Thirteen 297
Mathematical Intuition: Poincaré, Polya, Dewey
Reuben Hersh

Discussion 324
Claudio Bernardi

Chapter Fourteen 329
On the Finite: Kant and the Paradoxes of Knowledge
Carl Posy

Discussion 358
Silvia Di Paolo

Chapter Fifteen 363
Assimilation: Not Only Indiscernibles are Identified
Robert Thomas

Discussion 380
Diego De Simone

Chapter Sixteen 385
Proofs and Perfect Syllogisms
Dag Prawitz

Discussion 403
Julien Murzi

Chapter Seventeen 411
Logic, Mathematics, Heterogeneity
Emily Grosholz

Discussion 427
Valeria Giardino

Contributors 433

Index 437

CHAPTER NINE

ARTIFICIAL INTELLIGENCE AND EVOLUTIONARY THEORY: HERBERT SIMON'S UNIFYING FRAMEWORK

ROBERTO CORDESCHI

SUMMARY A number of contributions are been given in recent years to illustrate Herbert Simon's multidisciplinary approach to the study of behaviour. In this chapter, I give a brief picture of the origins of Simon's bounded rationality in the framework of rising AI. I show then how seminal it was Simon's insight on the unifying role of bounded rationality in different fields, from evolutionary theory to domains traditionally difficult for AI decision-making, such as those of real-life and real-world problems.

KEYWORDS bounded rationality, heuristics, ill structured problems, evolutionary theory, Artificial Intelligence, robotics

1. Introduction. Bounded rationality between human and machine decision making

As Herbert Simon once said, "Artificial Intelligence (AI) was born in the basement of the Graduate School of Industrial Administration at Carnegie Mellon University, and for the first five years after its birth, applications to business decision making (that is, operation research applications) alternated with applications to cognitive psychology" (Simon 1997, p. 5). Simon's view of the origins of AI (see Cordeschi 2007) suggests the role of disciplines, such as AI and psychology, not usually involved in the study of decision-making processes before the early 1950s.

Briefly, the prevailing model of the decision maker, in economics and classical game theory, was that of an agent who makes the best choice using an evaluation function based on the minimax procedure.[1] Chess game was the common metaphor for this, but it was also a kind of *Drosophila* for the rising AI, as Simon was used to saying. Within the field of machine intelligence of the early 1950s, it was evident that the use of an exhaustive or brute-force strategy for playing chess on computers would have encountered an insurmountable obstacle in the combinatorial explosion of possible moves, which Claude Shannon calculated to be of the order of $10.^{120}$ In cases such as chess, minimax procedure cannot be used in practice. Thus Shannon raised the issue of how "to develop a tolerably good strategy for selecting the move to be made," so that a machine could play "a skilful game, perhaps comparable to that of a good human player" (Shannon 1950, p. 260). "We might be satisfied—he concluded—with a machine that designed good filters even though they were not always the best possible" (p. 256).

Notice that Shannon's conclusion seems to touch on the *limits* of both a human and a computer decision maker dealing with problems of such computational complexity as that of chess. Which such a "tolerably good strategy" might that be, a *selective* strategy good enough albeit "not always the best possible"? Shannon suggested embodying in a chess program selection strategies such as those analyzed by the Dutch psychologist Adrian de Groot in chess masters, who made their analyses by "thinking aloud" during the game. His reference to de Groot seems to call attention to the *choice processes of an actual problem solver* in order to improve the performance of a computer program.

Shannon did not develop this point further. The choice processes of the actual problem solver were instead a primary interest for Simon in the 1950s. He had already rejected the normative approach of classical game theory, i.e. the analysis of choices or procedures that a rational agent *should* adopt in order to gain the optimal solution to a given problem (Simon 1946). He introduced psychology into the study of choice in management sciences and economics, and his concern centered on the decision-making behavior that characterizes an *actual* agent in his relationship with the environment. Such an agent is conditioned both by his own internal cognitive limits (concerning, for example, memory and information-

---

[1] See Sent (2004) on the development of game theory in relation to Simon's legacy. Gintis (2007) proposes game theory, in its recent *evolutionary* kind, as a ground for unifying behavioral sciences. I don't touch on this issue here, but it seems to me that Gintis' main claim that current (evolutionary) game theory "has shed its static and hyperrationalistic character" (p. 8, and see p. 9) might be in agreement with Simon's viewpoint.

processing ability) and by the complexity of the task environment. The chess player's behavior was still the metaphor for the rational agent's behavior, but in Simon's view such a metaphor should be referred to the "bounded rationality" of an actual problem solver or "administrative man," as he put it, rather than to the perfect-rationality of the idealized, omniscient *Homo oeconomicus* of classical economics. Internal limits and the complexity of the external environment, vividly exemplified by chess, force an actual agent to use not the ideally best strategy in the choice of moves, but different strategies, or *heuristics*, which are fairly "satisficing," to use Simon's well-known term (see Simon 2000, for a recent statement).

Heuristics are thus shared by humans and computers, to the extent they both display *efficient* problem solving behavior and can be considered as Information Processing Systems (IPSs), or, equivalently, as Physical Symbol Systems (PSSs)—see Newell and Simon (1972; 1976). Briefly, such systems are endowed with a receptor/effector apparatus, and with a fairly limited capacity of both long-term and short-term memory and of information processing.[2]

Newell and Simon always described those systems as being *adaptive*—a primary feature of a problem solving system. For Simon (1980, p. 36), there are "three different forms of adaptation." The first one is characterised by changes in the system's behavior over a short timescale; i.e. the behavior constantly changes as the system moves towards the solution to a problem in the external (or task) environment. This is *adaptation* in a strict sense. Many systems of this kind are also able to adapt over a somewhat longer timescale, meaning they are able *to learn*—in that learning consists of keeping and successfully reusing on different occasions certain strategies in adaptation or problem solving, i.e. certain heuristics. Finally, over the longest timescale, many systems are able to *evolve* or, as biological systems, to transmit changes through mutations and natural selection, as in Darwin's theory.

The latter point introduces a topic of interest here: Simon's claims on bounded rationality might be dealt with also within the framework of Darwinian evolution. A number of contributions have been given in recent years to illustrate Simon's multidisciplinary approach to the study of

---

[2] See Cordeschi (2006) on the double-faceted feature of heuristics in AI and Cognitive Science, and see Cordeschi (2008) for an investigation on Newell and Simon's IPS/PSS.

behavior.[3] In the sequel of this chapter, I will attempt to show how seminal Simon's insight was on the unifying role of bounded rationality in different fields, from evolutionary theory to domains traditionally difficult for AI decision-making, such as those of real-life and real-world problems.

2. A world with its history: bounded rationality and evolutionary theory

Simon's aforementioned adaptation taxonomy based on different timescales has not always been presented in the same terms. For example, Newell and Simon (1972, p. 3) considered *development,* not evolution, the third form of adaptation. Furthermore, the question of different timescales has been left out in many cases. For example, Simon soon established his analogy between Darwinian processes of natural selection and the processes of adaptive problem solving—based on bounded rationality—in one of his classic articles, "Rational choice and the structure of the environment" (Simon 1956). Later, Simon would recall how he proposed to elaborate on a kind of Darwinian model of bounded rationality in that article. He said: "Bracketing *satisficing* with *Darwinism* may appear contradictory, for evolutionists sometimes talk about survival of the fittest. But, in fact, natural selection only predicts that survivors will be "fit enough," that is, fitter than their losing competitors; it postulates satisficing, not optimizing." (Simon 1991, p. 166)

At the end of this section I shall return to the question of "survival of the fittest." Meanwhile we can ask: what exactly does bracketing satisficing, or bounded rationality, with Darwinian evolution mean? How is it that adaptive systems able *to evolve* are organized? Surely the best answer to these questions can be found by reference to another classic article by Simon published in 1962, "The architecture of complexity" (now reprinted in Simon 1996).

In this case Simon obviously takes the timescale into consideration, but here to justify the speed of evolution. For Simon, complex systems such as biological organisms evolve more quickly toward steady states if they are organized on different levels, i.e. in a way that facilitates the formation of intermediate steady states. This means that complex systems share a peculiar hierarchical structure in which the interactions of subsystems are weak but not negligible, i.e. those systems have the distinctive feature of

[3] See at least Augier and March (2004), and several of Mie Augier's later contributions.

being "nearly-decomposable."[4] In such systems "the short-run behavior of each of the component subsystems is approximately independent of the short-run behavior of the other components; […] in the long run the behavior of any one of the components depends in only an aggregate way on the behavior of the other components" (p. 198). To sum up, a complex, hierarchically organized system is (in the most interesting cases) a system which interacts with the environment and is composed of a number of parts or subsystems, so-called "intermediate stable forms" or "configurations." Biological systems are precisely of this type. It must be noted that it is near-decomposability which allows such systems to place themselves within all three timescales mentioned above. A rigid, non nearly-decomposable, hierarchy would not allow for any of these.

This particular internal organization of systems able to evolve in line with Darwinian theory is exemplified by Simon through human problem solving—actually neglecting different timescales in two cases, fairly slow in the case of evolution, and fairly fast in that of human problem solving. Consider the task of discovering the proof for a theorem. As Simon puts it, this task can be described as a search-process through a maze. Such a process involves much trial-and-error, but this is not completely random or "blind"—actually, it is selective. The formulae one obtains by applying rules are cues which direct further search. Thus,

> Indications of progress spur further search in the same direction; lack of progress signals the abandonment of a line of search. Problem solving requires *selective* trial and error. A little reflection reveals that cues signaling progress play the same role in the problem-solving process that stable intermediate forms in the biological evolutionary process. (Simon 1996, p. 194)

Beyond the already notable ambiguity regarding different timescales in natural selection and in problem solving (and others to which we shall return in the next section) the message seems clear. In the previous section we saw that it is heuristic procedures which reduce computational complexity to a reasonable size. Now we see that something similar also applies to Nature. Nature, like the administrative agent, does not optimize. Neither of them follows anything like a strategy of optimization. Different historical contingencies influence both. As Simon pointed out:

[4] See Callebaut and Rasskin-Gutman (2005), a collection of essays on near-decomposability and related topics.

> Evolutionary theory […] resembles most closely the […] model [of bounded rationality]. In both theories, searching a large space of possibilities and evaluating the products of that search are the central mechanism of adaptation. Both theories are myopic. Such optimization as they achieve is only local. (Simon 1983a, p. 73)

In conclusion, both evolutionary theory and the bounded-rationality claim are best described "not as optimization processes, but as mechanisms capable of discovering new possibilities that are 'improvements' over those attained earlier" (p. 74). This particularly explicit statement was in full agreement with certain now well-known claims by evolutionary biologists such as Stephen J. Gould and Richard C. Lewontin. Suffice it to mention the following passage by Gould, where the similarities with Simon are clear:

> Our world is not an optimal place. […] It is a quirky mass of imperfections, working well enough (often admirably); a jury-rigged set of adaptations built of curious parts made available by past histories in different contexts. […] A world optimally adapted to current environments is a world without history, and a world without history might have been created as we find it. History matters; it confounds perfection and proves that current life transformed its own past. (Gould 1985, p. 54)

This criticism of naive versions of adaptationism is reminiscent of Simon's criticism of optimizing classical rationality. It is not surprising that Gould, in a letter to Simon of October 10, 1990 ("you have been one of my intellectual heroes," he said), maintained that "the general ideas in the paper with Lewontin, and the concept of exaptation in particular (in the paper with Vrba) will indeed relate strongly to your 'docility' proposal."[5] This allows us to introduce another important topic.

The papers mentioned by Gould in his letter are now well known, and were immediately at the centre of heated discussions among biologists as well as philosophers (see Gould and Lewontin 1979; Gould and Vrba 1982). Briefly, and without entering into the merits of these discussions, the authors proposed (and would continue to propose in numerous later publications) a view of evolution which contrasted with that based on selective optimization, which they attributed to the naive, strictly adaptationist, version of Darwinism. They insisted on the presence and role

---

[5] This letter can be found in the Herbert Simon Collection at Carnegie-Mellon, URL: http://diva.library.cmu.edu/Simon/

of characters which are not immediately recognizable as adaptive, and considered the organisms as subjects which are not passive nor pliable in an optimal way by selection, as is maintained by the view they attributed to so-called panselectionism.

Beside adaptation, which is present when a character is selected for the function which it currently carries out, they introduced the notion of "exaptation" (which, besides, was already known to Darwin at least in certain forms). In a first case, exaptation acts when a characteristic selected previously by natural selection for a certain function is co-opted or "exapted" to a new function or use (as is the case with birds' feathers, which originally served for heat regulation and not flight). In a second case, exaptation acts when the characteristic is selected for reasons of development, architecture or more generally for historical contingencies, and is later co-opted or "exapted" to the function which we currently see (this is the case with the spandrels in the cupola of San Marco in Venice in the article Gould referred to in his letter to Simon). In both the cases, the role of natural selection remains crucial in fixing the (new) function within the species.

The general picture that Gould and others proposed was consistent with Simon's claim, whereby a world in which an optimal adaptation would prevail would be a world without history, and the great varieties of strategies for survival in complex environments used by organisms would have no explanation. But what relationship is there between exaptation and Simon's "docility," to which Gould refers in his letter?

The notion of docility is for Simon linked to that of altruism. Altruism has been a difficulty in social sciences and classical economics as well as in evolutionary biology at least since Darwin's time. In both cases, it is difficult to explain how certain individuals manifest behaviors which are of no immediate profit (neither of social or economic gain nor survival). As Simon put it, we speak of altruism "when an individual sacrifices fitness in the short run but receives indirect long-run rewards that more than compensate for the immediate sacrifice" (Simon 1983a, p. 58). Now, on the one hand altruism raises difficulties only if we assume a naively adaptationist view of natural selection, with the corollary of survival of the fittest. On the other hand, altruistic behaviors can be observed both in the social sphere and in biological evolution. At times together with opportunistic behaviors, forms of altruism can be observed even in very simple organisms, such as bacteria (see, for a critical review, West et al. 2007).

Docility is also connected to bounded rationality and, particularly in humans, has strong social consequences linked to the fact that it controls and guides individual behaviors through socially arranged channels. Competition is also abated here and survival could thereby be advantaged:

> The theory accounts for altruism on the basis of the human tendency (here called docility) […] to accept social influence—which is itself a product of natural selection. Because of the limits of human rationality, fitness can be enhanced by docility that induces individuals often to adopt culturally transmitted behaviors without independent evaluation of their contribution to personal fitness. (Simon 1990a, p. 1665)[6]

As with altruism, docility would not find place in a world without history, a simple and predictable world where a short-term advantage always becomes a long-term advantage. This certainly agrees not only with Gould and Lewontin's view, but also with every non-simplistic view of Darwinian evolution. There might be something more: in his letter to Simon, Gould might have seen a form of social exaptation in docility. Docility, as conceived by Simon, actually could be seen as a behavior exapted by constraints later triggered by the opportunity for social advancement. The notion of near-decomposability fits in with this possibility. If this is true, the very idea of adaptive systems in the world of economic competition (as in a company or a firm) could be broadened—which currently is justified in proposals to consider entrepreneurial firms as *exaptive* more than adaptive systems.[7]

Within this framework of a complex and unpredictable world where the survival strategies adopted by organisms multiply, Simon elaborated on a

---

[6] "Bounded rationality with docility produces altruism", as Simon briefly summed up (2004, p. 96). As he pointed out, his concern was not "the extension of an evolutionary metaphor to economics", but "the direct influence of the processes of neo-Darwinian biological evolution upon the characteristics of the individual human actors in the economy, and, through these characteristics, the influence of biological evolution upon the operation of the economy" (p. 89).

[7] See Dew et al. (2008). More generally, exaptation has been introduced in regard to the evolution of technology in recent time: "A strong focus on the adaptation of technology products and processes to user needs and efficiency criteria has generally obscured the phenomenon of exaptation, which points to the *non*-adaptive origins of many technologies, and the process by which they are later co-opted for other roles" (Dew et al. 2004, p. 70). An example by the authors is CD-ROM, which was patented in 1970 as a digital-to-optical recording and playback system, and subsequently it was co-opted for another use—a data storage medium for computers. A kind of exaptation in nervous systems could be found in the "neural reuse" introduced by Anderson (2010).

view of the notion of ecological niche which, as Callebaut (2007, p. 79) well noted, competes with that developed by Lewontin in the same years. They both criticize the concept of an ecological niche where an organism would easily adapt, but which is too simple (or simplistic) to represent a model for actual adaptation. It must be added that even before this, Simon (1980), whilst rejecting the claim of "survival of the fittest" *qua* optimization criterion, was already referring to the niche. If that claim refers to the outcome of competition between species in seeking to occupy a particular ecological niche, then there should be as many niches as there are species—an obviously absurd consequence. Instead,

> The niches themselves *are not determined by some inflexible, invariant environment, but are defined in considerable measure by the whole constellation of organisms themselves*. [...] Hence, it is not obvious what optimization problem, if any, is being solved by the process of evolution. At most, the occupancy of each niche is being locally 'optimized' relative to the entire configuration of niches. (Simon 1980, p. 44, my italics)

In this case, too, organisms are considered as subjects which are not passive nor pliable in an optimal way by natural selection. Organisms, as systems endowed with bounded rationality, change the external environment and are changed by it reciprocally. This justifies Simon's famous metaphor of bounded rationality theory as a theory which has two blades, like a pair of scissors (Simon 1990b, p. 7).

3. Ill-structured domains as an issue of bounded rationality

In the previous section we saw how Simon, in interpreting natural selection processes as problem-solving processes, gave the example of theorem proving. He was referring to LOGIC THEORIST (LT), the automatic logic theorem proving which began to run on a computer in 1956 (see Cordeschi 2002, ch. 5). LT proved a number of theorems of sentential logic, and did so using selective heuristics similar to those used by students who were given the same type of task. These heuristics were inferred via the thinking-aloud-protocols method (see section 1). Now, if we look at how LT works, we see how Simon's analogy between problem solving and natural selection (or, rather, between formulae generated during a proof and intermediate stable forms) risks being misleading.

One of the heuristics originally used successfully by LT was the "similarity test." Beginning with the formula given as a theorem, the proof consisted of generating a succession of formulae which *resembled* more and more closely the formula which was known to be the solution. LT was programmed to do what verbal protocols suggested that students were doing when making certain proofs: from amongst all the legal rules of sentential logic, they were trying to apply only those that could seem useful in *eliminating the differences* between what had been given as a theorem, or starting point of the proof, and what they were seeking, the problem solution. This example does not seem to be a basis for establishing a well-founded analogy between natural selection and problem solving within the framework of bounded rationality. I shall attempt to explain why.

Problems like those faced by the LT are typical "end of chapter" problems—they are textbook exercises for students. These problems are "well-structured," i.e. their definition leaves no room for ambiguity, the existing state and desired state are clearly identifiable, the rules firmly established, and there are no exceptions. These problems are usually set against "ill-structured" problems which concern real-life or real-world domains. Ill-structured problems have different features: the problem itself may not be easy to identify, the starting data uncertain, the goal not well defined, there are hardly ever any explicit fixed rules, there may be different viewpoints to evaluate and maybe even different criteria.[8]

Now it is true that defining a logical problem (and other games, puzzles, and toy-problems of early AI) as well-structured could relegate to second place the *difficulty* such a problem raises in any case to the bounded-rationality problem solver. Simon always insisted on this point, underlining that there is a continuum of problem types and not a strict dichotomy between well- and ill-structured problems. In domains like logic or chess, there is a risk of confusing the idealized (limitlessly powerful) problem solver with the actual problem solver, who has limited computational capacities (Simon 1973, pp. 185–186). From the *actual* problem solver's point of view, the problem is always fairly ill-structured, given the combinatorial explosion of alternative paths which characterizes tasks like logic.

This is true, but not what I am discussing here. The point is that it is one thing to find a path from among the many possible paths between the starting state and end state, i.e. the solution, when these are both known and well-defined, and it is another thing to find the path when they are not well-defined, and particularly when the solution is literally *not* known. In the

---

[8] For a review and a case study, see Shin et al. (2003), who touch upon the issue of the different skills shown by students in solving the two different kinds of problem above.

first case, which is the case of Simon's example, as the solution is known *a priori*, it strongly guides the search, i.e. it leads towards the generation of *certain* well-formed strings of symbols or formulae rather than others. This happens through continual pattern-matching between what with time is obtained through the similarity-test heuristic and the problem solution. As regards Darwinian selection, here one has a guided generation of *certain* intermediate stable forms rather than others with reference to a pattern *known as the final outcome* of the evolutionary process. Nothing less Darwinian than this, one could say.

Simon insists time and again on defining the process of generating intermediate stable forms as "random" and "non-teleological" (Simon 1996, pp. 191–192). Yet he recognizes that this generation, to the extent that it is *heuristic*, is due to "not completely random or blind" trial and error (pp. 193–194). How do we resolve this apparent contradiction between heuristic problem-solving and random problem-solving in Darwinian selection?

Examples like those of LT ought to be abandoned, and we should look to considering heuristics used in ill-structured domains. Scientific research could be a case in point. According to Simon (see, e.g., Langley et al. 1987, p. 3, and pp. 15–16) the processes of creative discovery, like scientific discovery, are not completely "random" and, what is more, are not qualitatively different from those of ordinary problem solving, usually used in solving more or less complex toy-problems. Yet scientific discovery is a problem which cannot in any way be defined as well-structured, and certainly not in the sense of LT's end-of-chapter problems. On the contrary, scientific discovery is a typically ill-structured problem in terms of the definition above. In this case, we can do away with the strict teleological component of knowledge-of-the-end-state that we see in end-of-chapter tasks. But the problem of generating intermediate stable forms remains: is it entirely random? What precisely is the role of heuristics if one wants to keep the analogy with Darwinian evolution?

In classical AI, I believe that these problems were tackled best by Douglas Lenat.[9] His theory of heuristics suggests a hypothesis concerning

[9] An excellent survey of machine-discovery programs of the time was carried out by Rowe and Partridge (1993). A harsh criticism of Lenat's programs comes from Koza (1992). He insisted on greater plausibility of the analogy with Darwinian selection based on genetic programming than one based on Lenat's heuristic programming—a issue that I shall not deal with here (I am uniquely interested in the analogy within classical AI). Pennock (2000) looked at a similar question to the one I have posed above regarding Simon ("can we show that Darwinian

the difficulties in the analogy above, and indicates a possible solution within the field of heuristic programming of the time. The starting point is the well-known one from the pioneering time of AI, also considered by Simon: a completely random *generator* of potential solutions together with a rigorous *tester* is not an efficient procedure. Applying only this "random generate and test" procedure would not produce solutions in a reasonable time in the case of automatic problem solving nor anything like survival of the fittest in biological evolution. Lenat's heuristic theory and the EURISKO program that implemented it foresaw that the heuristics could themselves *evolve*, from the weakest to the most domain-specific. As far as natural selection is concerned, these heuristics evolved by random mutations, whose effects are in primitive organisms and then extended to higher organisms. For Lenat, natural selection begins with primitive organisms and a weak method (random generation, followed by more rigid tests) for improving them. It is in this way that the first primitive heuristics accidentally came into being, and then overcame the less efficient random-mutation mechanism. Thus,

> By now the evolution of most higher animals and plants may be under the guidance of a large corpus of heuristics, judgmental rules abstracting a billion years of experience into prescriptions and (much more rarely) proscriptions regulating and coordinating clusters of simultaneous mutations. Random mutation would be still present, but in higher organisms its effect might be mere background noise. (Lenat 1983, p. 288)

Lenat's theory of heuristics seems to be dealing with phenomena connected to forms of exaptation. Lenat mentions Gould on the possibility of considering the " ontogeny recapitulating phylogeny" claim. Moreover, the above quotation suggests that the increase of efficient heuristics in higher animals and humans can increment their ability in solving problems which caused major difficulties for their ancestors. Thus, the process which began with Darwinian "random generate and test" could evolve into a more efficacious process, a kind of "plausible generate and test" which at this point *channels* the evolutionary process. The intertwining of random mutation and plausible adjustment allows evolution to be channeled along one of the many possible paths. From a certain point on, there are certain

---

processes—that is to say, natural selection acting upon *undirected* heritable variations—really are *capable* of making non-trivial novel discoveries?" (p. 231). On this point, also Pennock holds a point of view sympathetic to evolutionary computation.

heuristic constraints according to which certain changes can have increasing success and so make evolution, *in this sense*, guided. Heuristics sum up and incorporate past history and different kinds of contingency: “there is no inherent ‘direction’ […]; rather, it is simply a mechanism for avoiding what seems, empirically or historically, to be deleterious, and for seeking what seems empirically to be advantageous” (p. 288).[10]

Simon’s near-decomposability also has its role in this case, but here within the framework of the evolution of heuristics, which is now the key point for maintaining the analogy with the problem solver on a Darwinian basis. Heuristics evolve like species: “incorrect heuristics die out with the organisms that contain them. […] Otherwise, as animals get more and more sophisticated, they would begin to evolve more and more slowly. Random mutations, or those guided by a fixed set of heuristics, would become less and less frequently beneficial to the complex organism, less frequently able even to form part of a new stable subassembly, as Simon suggests” (p. 294).

The case of scientific discovery is certainly not the only unequivocal case of ill-structured domain for the bounded-rationality problem solver. Scientific discovery is among cases which, differently from problems like games, puzzles and logic, require the problem solver to have a more or less ample quantity of specialized knowledge. This is the case with medical diagnosis, design, management and so on. In AI, these problems were initially considered within the field of expert systems.[11]

---

[10] I wish to point out that this claim is different from others, which aim at setting the two “tests” above against each other, rather than integrating them. For example, Johnson-Laird (1983) set a completely random “Neo-Darwinian” procedure against a “Neo-Lamarckian” procedure, guided by certain constraints, adding a third, or “multi-stage” procedure, which is in a sense an intermediate procedure. Lenat’s claim, stemming from Simon’s idea of heuristics which *guide* evolution, could be seen as the appreciation of “Baldwin effect” on Darwinian evolution—that is to say, James M. Baldwin’s 1896 claim that offspring tends to have an increased ability in learning skills, and this would suggest the *right direction*, as it were, of evolution. Baldwin effect has had a revival in a context different from Lenat’s, i.e. in Artificial Life (see, e.g., Mitchell and Forrest 1994).

[11] Lenat and Feigenbaum (1991) critically dealt with the difficult issues regarding expert systems of the time. At least two topics pointed out by Lenat in EURISKO have been developed in expert systems (see Cordeschi 2006) and in the more recent field of data mining (Brazdil et. al. 2009)—that of the trade-off between universality and performance and that of metaheuristics.

An AI ill-structured domain which is different, in turn, from the expert system domain is that of real-world problems. Here the notion of bounded rationality must be further extended, since it concerns artificial agents like robots which are able to interact with the world. I want to highlight something that in my opinion has been overlooked here: Simon considered this issue regarding robotics from a novel viewpoint *before* the birth of the so-called "new" or "behavioral" robotics of the mid-1980s (see Simon 1973; 1983b).

As Simon explicitly recognizes, traditional AI systems, such as his own UNDERSTAND system or ISAAC system, not to mention much simpler puzzle-solving systems of early AI (such as the missionaries-and-cannibals puzzle), are simulated systems: all lack a feedback mechanism which receives sensory information from the world. Nevertheless, it is just such a mechanism that enables a system such as a robot "continually to adjust its expectations to the unfolding reality" (Simon 1983b, p. 28). Furthermore, "what distinguishes robotics […] from other areas of AI […] is that the *robot is embedded in an external environment that it can sense and act upon*" (p. 26, my italics). And notice that the robot's performance is evaluated by its actual behavior in the real world, "not by its self-imagined behavior in a toy problem space" (*Ibid*.), i.e. in a simulated and simplified world.

> The robot can adapt the complexity and other properties of its internal problem space to its own computational capabilities, depending upon feedback to eliminate the discrepancies between expectations and reality. […] For this reason, *robotics appears to be the most promising area in AI in which to study everyday problem solving*. (Simon 1983b, p. 27, my italics)

In this framework one important question—which would often be a cause for debate in AI up to present time—concerns the role of representations or "models" of external world. These are built by the agent in this feedback-based relationship with the world in which it is "embedded." As Simon puts it, these representations can never be meant as *complete* models of reality, according to the unrealistic assumptions of the planning. I wish to point out two consequences of this claim.

First, accurate information, and so a complete model of reality, is possible only within microworlds, "represented inside the computer" and populated by *simulated* robots (Simon mentioned Terry Winograd's then famous SHRDLU). These simulated robots "do not face the issue that is critical to a robot when dealing with a real external environment—the issue of *continually revising its internal representation* of the problem situation to conform to the facts of the world" (Simon 1973, p. 195n., my italics). Second, "the unrealism of planning models need not always be remedied by making them more realistic, hence more complete" (Simon 1983b, p. 26). When the real world is the case in point, in many cases it is better to start with "*an over-simple model*, and to depend on feedback to provide second-order approximation" (*Ibid*., my italics). To conclude, a feedback mechanism is essential in producing adequate representation of the environment—and this is a matter of empirical knowledge about the world itself. Such knowledge is the "critical component of human bounded rationality" (p. 27).

If I have stressed these quotes from Simon, it is because they come from the decade immediately before Rodney Brooks's proposal of agents which are "embodied," or "embedded" (Simon's term) in the world with which they interact by means of different feedback systems (Brooks 1991). Simon's claims brought to the fore the need for a robotics *different* from that which was prevalent at his own time—a robotics based on pure planning or PLANNER-like systems. On the one hand, his claims show how situated cognition and embodiment, issues often *raised contra* Simon by various critics of so-called classical or disembodied AI, are in fact problems *raised by* Simon, within the framework of bounded rationality. On the other hand, few AI researchers now doubts that artificial agents that interact with the external world like robots must be endowed not only with on-going feedback from the environment, but also with off-line mechanisms for planning, reasoning, anticipation, and so on, i.e. for abilities which greatly increase agent's *autonomy*.[12]

Therefore, it was Simon who raised the problem of how real agents like robots could develop representations or models of the environment in which they are embedded which are more and more adequate. They do this notwithstanding their internal *limits*, usually starting from simple representational systems. What Simon thus rejected was the false image of a world given once and for all, i.e. a world with *ad hoc* representations. This is the world without history, without exceptions and without failed expectations which, as we have seen, is the illusory model of reality Simon always rejected.

When Simon many years later revisited the problem of what a PSS is in the light of important developments in behavior-based robotics, what he proposed was not an opportunistic adaptation of the "old" notion of PSS to

[12] Presently, autonomy is a central notion as regards agent's behavior, also beyond the robotics domain: see Bibel (2010).

the context of the (then) “new” robotics. On the contrary, Simon again took up issues that he himself had raised years before the development of such robotics. Here is the conclusion:

> A complex task is much more difficult to accomplish when assisted only by direct feedback from the environment. […] Of course, pure planning, with no situational feedback, is equally ineffective, but it is unfortunate that failures of pure planning schemes have motivated researchers to argue for the opposite extreme instead of a more sophisticated intermediate strategy. (Vera and Simon 1993, pp. 15–16)

If one considers the evolution of robotics from the 1990s to the present time, one concludes that things have gone largely in the direction that Simon had hoped for. Something like an “intermediate strategy” is what has prevailed in much robotic research.[13]

4. Conclusion

In this chapter I have discussed the reason why I believe bounded rationality has become a key notion for unifying the study of both natural and artificial systems when dealing with complex problem-solving tasks via adaptation and evolution strategies. Presently, bounded rationality seems to be the theoretical ground for various behavioral and evolutionary disciplines. Different researchers in such disciplines shared it as a basic assumption, albeit from different viewpoints—e.g., Daniel Kahneman on the one side and Gerhard Gigerenzen on the other: they both share such an assumption, while maintaining different viewpoints on the nature and origins of human mistakes in reasoning and choice (see at least their contributions in Augier and March 2004).

---

[13] See Carlucci Aiello et al. (2001) on “hybrid” (i.e. *intermediate* reactive/deliberative) robotics. Siciliano and Khatib (2008) includes chapters relevant for the issues touched on here: see at least the chapters on behavioral robotics by M.J. Mataric and F. Michaud, on biologically inspired robotics by J.-A. Meyer and A. Guillot, and on evolutionary robotics by D. Floreano, P. Husbands and S. Nolfi. More generally, a well balanced view on the issue of “old” and “new” AI has been stated by Paul Verschure (see, e.g.: “One motivation for trying to unify these two views is that they both seem to capture different aspects of intelligence. […] Traditional AI failed to ground its solutions in the real-world, while new AI faces the challenge to scale up to non-trivial cognitive processes”, Verschure and Althaus 2003, p. 562).

In such cases one should always remember, as Simon concluded (1985, p. 297), that the departure from the classical view of rationality "should not be mistaken for a claim that people are generally 'irrational.' […] They usually have reasons for what they do."

References

Anderson M.L. (2010). Neural Reuse: A Fundamental Organizational Principle of the Brain. *Behavioral and Brain Sciences*, 33: 245–313

Augier M. and March J.G., eds. (2004). *Models of a Man*. Cambridge, MA: MIT Press.

Bibel W. (2010). General Aspects of Intelligent Autonomous Systems. *Intelligent Autonomous Systems*, 275: 5–27.

Brazdil P., Giraud-Carrier C., Soares C. and Vilalta R. (2009). *Metalearning: Applications to Data Mining*. Berlin-Heidelberg: Springer.

Brooks R.A. (1991). Intelligence Without Representation. *Artificial Intelligence*, 47: 139–159.

Callebaut W. (2007). Herbert Simon's Silent Revolution. *Biological Theory*, 2: 76–86.

Callebaut W. and Rasskin-Gutman D., eds. (2005). *Modularity: Understanding the Development and Evolution of Natural Complex Systems*. Cambridge, MA: MIT Press.

Carlucci Aiello L., Nardi D. and Pirri F. (2001). Case Studies in Cognitive Robotics. In: Cantoni V., Di Gesù V., Setti A. and Tegolo D., eds. *Human and Machine Perception 3: Thinking, Deciding, and Acting*. Dordrecht: Kluwer.

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

—. (2006). Searching in a Maze, in Search of Knowledge, *Lecture Notes in Computer Science*, 4155: 1–23.

—. (2007). AI Turns Fifty: Revisiting Its Origins. *Applied Artificial Intelligence*, 21: 259–279.

—. (2008). Steps Toward the Synthetic Method. In: Husbands P., Holland O. and Wheeler M., eds. *The Mechanical Mind in History*. Cambridge, MA: MIT Press.

Dew N., Read S., Sarasvathy S.D. and Wiltbank R. (2008). Outlines of a Behavioral Theory of the Entrepreneurial Firm. *Journal of Economic Behavior and Organization*, 66: 37–59.

Dew N. and Sarasvathy S.D. and Venkataraman S. (2004). The Economic Implications of Exaptation. *Journal of Evolutionary Economics*, 14: 69–84.

Gintis H. (2007). A Framework for the Unification of the Behavioral Sciences. *Behavioral and Brain Sciences*, 30: 1–61.

Gould S.J. (1985). *The Flamingo's Smile: Reflections in Natural History*. New York: Norton.

Gould S.J. and Lewontin R.C. (1979). The Spandrels of San Marco and the Panglossion Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London B*, 205: 581–598.

Gould S.J. and Vrba E.S. (1982). Exaptation—A Missing Term in the Science of Form. *Paleobiology*, 8: 4–15.

Johnson-Laird P.N. (1983). *Human and Machine Thinking*. Hillsdale, NJ: Erlbaum.

Koza J.R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

Langley P., Simon H.A., Bradshaw G.L. and Zytkow J.M. (1987). *Scientific Discovery: Computational Exploration of the Creative Processes*. Cambridge, MA: MIT Press.

Lenat D.B. (1983). Learning by Discovery: Three Case Studies. In: Michalski R.S., Carbonell J.G. and Mitchell T.M., eds. *Machine Learning: An Artificial Intelligence Approach*. Palo Alto, CA: Tioga Press.

Lenat D.B. and Feigenbaum E.A. (1991). On the Thresholds of Knowledge. *Artificial Intelligence*, 47: 185–250.

Mitchell M. and Forrest S. (1994). Genetic Algorithms and Artificial Life. *Artificial Life*, 1: 267–289.

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

Newell A. and Simon H.A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM*, 19: 113–126.

Pennock R.T. (2000). Can Darwinian Mechanisms Make Novel Discoveries? Learning from Discoveries Made by Evolving Neural Networks. *Foundation of Science*, 5: 225–238.

Rowe J. and Partridge D. (1993). Creativity: A Survey of AI Approaches. *Artificial Intelligence Review*, 7: 43–70.

Sent E.M. (2004). The Legacy of Herbert Simon in Game Theory. *Journal of Economic Behavior and Organization*, 53: 303–317.

Shannon C.E. (1950). Programming a Computer for Playing Chess. *Philosophical Magazine*, 41: 256–275.

Shin N., Jonassen D.H. and McGee S. (2003). Predictors of Well-Structured and Ill-Structured Problem Solving in an Astronomy Simulation. *Journal of Research in Science Teaching*, 40: 6–33.

Siciliano B. and Khatib O., eds. (2008). *Handbook of Robotics*. Berlin-Heidelberg: Springer.

Simon H.A. (1946). *Administrative Behavior*. New York: Macmillan.

—. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63: 129–138.

—. (1973). The Structure of Ill Structured Problems. *Artificial Intelligence*, 4: 181–201.

—. (1980). Cognitive Science: The Newest Science of the Artificial. *Cognitive Science*, 4: 33–46.

—. (1983a). *Reason in Human Affairs*. Stanford, CA: Stanford University Press.

—. (1983b). Search and Reasoning in Problem Solving. *Artificial Intelligence*, 21: 7–29.

—. (1985). Human Nature in Politics: The Dialogue of Psychology and Political Science. *The American Political Science Review*, 79: 293–304.

—. (1990a). A Mechanism for Social Selection and Successful Altruism. *Science*, 250(4988): 1665–1668.

—. (1990b). Invariants of Human Behavior. *Annual Review of Psychology*, 41: 1–19.

—. (1991). *Models of My Life*. New York: Basic Books.

—. (1996). *The Sciences of the Artificial*. Cambridge, MA: MIT Press (3rd edition).

—. (1997). The Future of Information Systems. *Annals of Operation Research*, 71: 3–14.

—. (2000). Barriers and Bounds to Rationality. *Structural Change and Economic Dynamics*, 11: 243–253.

—. (2005). Darwinism, Altruism and Economics. In: Dopfer K., ed. *The Evolutionary Foundations of Economics.* Cambridge: Cambridge University Press.

Vera A.H. and Simon H.A. (1993). Situated Action: A Symbolic Interpretation. *Cognitive Science*, 17: 7–48.

Verschure P.F.M.J. and Althaus P. (2003). A Real-World Rational Agent: Unifying Old and New AI. *Cognitive Science*, 27: 561–590.

West S.A., Griffin A.S and Gardner A. (2007). Social Semantics: Altruism, Cooperation, Mutualism, Strong Reciprocity and Group Selection. *Journal of Evolutionary Biology*, 20: 415–432.