HERBERT SIMON, ARTIFICIAL INTELIGENCE AND BOUNDED RATIONALITY

Soim Horatiu
Western University "Vasile Goldis" Arad, horatiu_soim@uvvg.ro

ABSTRACT:

The aim of this paper is to present the contribution Herbert Simon made in social science, particularly in what was called "the cognitive revolution". Being a complex and controversial personality, Simon challenged the scientific community. His major contributions were acknowledged by awarding him the Nobel Prize for economics in 1978, and by the impressive number of disciples who are aware of his work and recognize him as a pioneer in Artificial Intelligence and Cognitive Psychology. Artificial intelligence can be used as a model and verifying tool of human rationality. In economics, modeling of rationality through artificial intelligence led to the concept of "bounded rationality" that had and have a major impact on the economic theory.

I. INTRODUCTION

The aim of this paper is to present the contribution Herbert Simon made in social science, particularly in what was called "the cognitive revolution". Being a complex and controversial personality, Simon challenged the ideas of the orthodox scientific community. His major contributions were acknowledged by awarding him the Nobel Prize for economics in 1978, and by the impressive number of disciples who are aware of his work and recognize him as a pioneer in Artificial Intelligence and Cognitive Psychology. The foundation of his path breaking contribution is his generalized view of an intelligent system as any physical device - including the human mind - that could manipulate symbols. Thus he used AI methods in cognitive science research in order to model subsystems of human cognition. His work made important contributions in many fields, namely psychology, Artificial Intelligence, economics.
In the second part is presented the Simon's scientific path, both the personal career and the scientific contributions. In the third part we address his vision regarding Artificial Intelligence and how it can be used to model and verify human rationality. In the fourth part we present the concept of "bounded rationality" and its impact on the economic theory. Conclusions are drawn in the fifth part.

II. HERBERT SIMON – THE MAN

Herbert Simon is one of the most controversial economists of the XX century. Born in 1916, he studied in Chicago, the place of one of the best university at that time and where he returned to teach at 26 years. His teachers and fellows names tell us a lot about the quality of the scientific environment where he studied: Henry Schultz, Rudolf Carnap, Keneth Arrow, Don Patinkin, Oscar Lange, Milton Friedman, Franco Modigliani.

At the end of the 50' and beginning of 60' he wrote about the decision making in firms and administration, and took place a slightly shift toward psychology. He was particularly interested in the process of decision making and reached to the conclusion that the best way to study it is the computer simulation of human thinking. At the beginning of '70 he transferred to the Psychology department of Carnegie Mellon University, where he promoted his ideas

regarding cognitive psychology. In 1978 he was awarded with Nobel Prize in economics "for his pioneering research into the decision-making process within economic organizations." (Nobel-prize, press) Thus its not surprising that Herbert Simon won the A. M. Turing Award for his work on computer science in 1975 and the National Medal of Science in 1986. Finally, he was awarded the American Psychological Association's award for outstanding lifetime contributions to psychology, in 1993.

This diversity of Herbert Simon interests could seem unusual, but all these domains have a common point – In his entire career he was "[…] concerned with laying foundations for a science of man that will comfortably accommodate his dual nature as a social and as a rational animal." (Simon Herbert, 1957: p. vii)

His path in scientific area is surprising and educative. Starting with political science, he approached management theory, economics, artificial intelligence and cognitive psychology. But we will se that there is logic of this path because difficulties he faced in an area led him to research in other areas that provided responses to those questions. Thus the shortcomings of the foundations of economics determined him to study cognitive psychology and artificial intelligence.

He was, along with Allen Newell, one of the pioneers of Artificial intelligence starting from an experimental point of view, founding the first artificial intelligence laboratory at Carnegie Mellon University.

Newell and Simon elaborated the information processing conception of human cognition in the 1950s (Newell, Shaw and Simon, 1958). In their opinion the knowledge consists in representations or collections of abstract symbols which are stored in the mind. These representations arrive in the mind via 'transducers' that transform external stimuli into internal representations. Human problem-solving and reasoning, consists in computations performed on these symbols, basically, a search processes designed to find a match between external stimuli and the existing repertory of stored representations

As we will present in the next part Newell and Simon developed this view and argue that the basic architecture of human cognition is a physical symbol system: "A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called expression (or symbol structure)… Besides these structures the system also contains a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction, and destruction. […] A Physical Symbol System has the necessary and sufficient means for general intelligent action" (Newell and Simon, 1976, p. 116).[1]

III. INFORMATION PROCESSING AND ARTIFICIAL INTELIGENCE

A definition of Artificial intelligence (AI) states that it is intelligence exhibited by any manufactured (i.e. artificial) system. It is also known as machine intelligence and is often applied to general purpose computers. Much of the initial interest in AI research derived from the experimental approach to psychology, studying what may be called linguistic intelligence. This issue is best exemplified by the Turing test.[2] Other approaches to artificial intelligence

1 Newell et.al (1989, p. 103) caution that the architecture should not be confused with a representation of the external world. The architecture supports such a representation but does not itself provide it.

2 The Turing test is a test of a machine's capability to perform human-like conversation. Described by Alan Turing in the 1950 paper On Computing Machinery and Intelligence (1950)"Computing machinery and intelligence", it proceeds as follows: a human judge engages in a natural language conversation with two other parties, one a human and the other a machine; if the judge cannot reliably tell which is which, then the machine is said to pass the test. It is assumed that both the human and the machine try to appear human. In order to

are those that don't study linguistic intelligence. They include robotics and collective intelligence approaches, and study the active manipulation of an environment, or consensus decision making, and use issues from biology and political science for modeling the organizing of "intelligent" behavior. During the time, there were two broad branches of AI research:

– classical or symbolic AI approach. It was interested, generally, by the symbolic manipulation of abstract concepts, and its methodology was employed in the expert systems.

- "connectionist" approach, based on artificial neural networks, which try to create a learning capable machine, leading to an "evolving" intelligence. The man creates the building systems capable to self improving them through some automatic process rather than systematically designing something to complete the task.

They had priorities at different moments but it has become clear that the methods using both broad approaches have severe limitations. The initial enthusiasm left place to disappointment. The promises were grandiose and the failure of the projects to produce immediate results led to large cutbacks in funding in the late 1980's.[3] The next years many AI researchers moved into related areas such as robotics, machine learning and computer vision.[4] Today we can say that modern AI research is concerned with producing useful machines to automate human tasks requiring intelligent behavior.

Herbert Simon is one of the pioneers of the artificial intelligence. We can say that the year artificial intelligence[5] was born is 1955, when Simon developed, along with his fellow Allen Newell, the first computer program that "thought" – "Logic Theorist".[6] It was the first computer program capable of "thinking" and represented the beginning of Artificial Intelligence. Using this program and another - "General Problem Solver", Simon proved that a computer can discover proofs of geometric theorems, can play chess or "discover" laws of Physics, Chemistry, Astronomy (The third law of Kepler). This is the reason why he said that computers can "think".

The next step was to study human decision making by trying to simulate it on computers.[7] A new theory of decision making emerged which led to the idea that computers can have artificial intelligence that resemble human thinking. They were not "not comparing computer structures with brains, nor electrical relays with synapses" (Newell, Shaw and Simon, [1958] 1989, p. 8). According to them the computer and the human brain are just two different physical supports for executing the same type of information processes: "Digital

keep the test setting simple and universal (to explicitly test the linguistic capability of some machine), the conversation is usually limited to a text-only channel such as a teletype machine as Turing suggested.

3 The consequence of reducing the funds led to a general downturn in AI activity known as AI winter.

[4] Modern AI research focusses on practical engineering tasks. . Examples of this are pattern recognition, image processing, neural networks, natural language processing, robotics and game theory. While the roots of each of these disciplines is firmly established as having been part of artificial intelligence, they are now thought of as somewhat separate.

[5] The most basic A.I. program is a single If-Then statement, such as "If A, then B." If you type an 'A' letter, the computer will show you a 'B' letter. Basically, you are teaching a computer to do a task. You input one thing, and the computer responds with something you told it to do or say. All programs have If-Then logic.

[6] In December 1955 Herbert Simon and Allen Newell succeeded in writing a computer program that could prove mathematical theorems taken from the Bertrand Russell and Alfred North Whitehead classic on mathematical logic, "Principia Mathematica."

[7] Some of these programs displaying some degree of intelligence are:

- Brainboost (*http://www.brainboost.com*) - a question-answering system
- SHRDLU - an early natural language understanding computer program developed in 1968-1970.
- Creatures, a computer game with breeding, evolving creatures coded from the genetic level upwards using a sophisticated biochemistry and neural network brains.
- Talk to William Shakespeare (*http://www.shakespearebot.com*) - William Shakespeare chatbot

computers… can, by appropriate programming, be induced to execute the same sequences of information processes that humans execute when they are solving problems. Hence, as we shall se, these programs describe both human and machine problem solving at the level of information processes." (Newell, Shaw and Simon, [1958] 1989, p. 8). Simon developed, along with his colleagues, a theory that the human mind manipulates symbols in a similar way with a computer. Human beings use symbolic processes to solve problems, reason, speak and write, learn and invent. Since then cognitive psychology has built and tested empirical models of these processes. The models take the form of computer programs that simulate human behavior.

Simon said that thinking can be analyzed both at the symbolic and physical, chemical, neurological levels, and that the gap between them has to be filled. He approaches the symbolic level of thinking, stating that it can be understood a serial processing of information at the symbolic level.

How is processed information at the elementary level? The elementary information is expressed through symbols, which are put into lists. Between symbols there are links that describe other symbols. In a very simplistic description we can say that a process get symbols, process them and deliver symbols. The elementary processes that can be applied to these symbols are: discrimination (comparing) structures, branching, creating, memorizing or copying, inserting or erasing a symbol, modifying a term of an association. For this the process needs initial inputs – axioms and inference rules. These are not the result of reason but are induced from observations. "[…] the processes that produce the transformations of inputs into outputs (rules of inference) are also introduced by fiat and are not the products of reason." (Simon, 1983: p. 5) The attempt to justify them would lead to an infinite regression. These inductions are based on observation, perception, inferences, that are influenced by preconceptions and presuppositions. We cannot obtain axioms through reason from facts and values. They have to come from other side. Descriptive inputs cannot give normative outputs. Reason is instrumental. It cannot say us where to go, at most it can say how to get there. This thing drives us to limitations of rationality. We cannot obtain through induction infallible laws no matter the amount of facts - "No number of viewings of white swans can guarantee that a black one will not be seen next." (Simon, 1983: p. 6)

How is thinking process seen? The thinking process is viewed as a computer program. Both human mind and computer, whatever the physical structures that support it, operate with symbols, and finally this is the thing that matter. Thinking process uses some basic procedures: means-end analysis, planning process, problem solving organization, productions, etc. In fact these procedures were used in computer programmes – like General Problem Solver – that simulate the human thinking. Making an analogy between human thinking and Artificial Intelligence, Simon concluded that computers can reach finally, through technological development, at human performances. This belief was enforced by the performances those software programmes that allow learning from experience. Simon was very enthusiastic for artificial intelligence and predicted in 1957 that within 10 years, "a digital computer will be the world's chess champion unless the rules bar it from competition"[8] and within the visible future will exist "machines that think, that learn and that create".(Simon Herbert, 1957).

[8] We have to mention that this prophecy fulfilled about three decades after Herbert Simon had predicted - the I.B.M computer Deep Blue did finally beat the world chess champion Gary Kasparov in 1997.

Melanie Mitchell analyzed what are the usually arguments raised against the possibility of replicating the life mechanism in computers. Most of the people believe that computers cannot mimic the key requisites for life (Mitchell, 2000):

"*Autonomy*: - a computer can't do anything on its own; it can do only what humans program it to do."

Metabolism: - computers can't create or gather their own energy from their environment like living organisms do; they have to be fed energy (e.g., electricity) by humans."

Self-reproduction: - a computer can't reproduce itself; to do so it would have to contain a description of itself, and that description would have to contain a description of itself, and so on ad infinitum."

Survival instinct: - computers don't care whether they survive or not." (For example, from an editorial in the Boston Globe: - Deep Blue may have beat Kasparov, but it didn't get any joy out of it.")

Evolution and adaptation: - a computer can't evolve or adapt on its own; it is restricted to change only in ways specified ahead of time by its programmer."

However the researches created life-like computing systems that cover these characteristics. For example self-copying programs reflect self-reproduction and genetic algorithms reflects evolution and adaptation. On the other hand „[...] the discipline of artificial intelligence is now able, in many domains, to build expert systems capable of matching the decision-making performance of human professionals." (Simon, 1992: p. 4)

Cognitive psychology is the science which studies the mental processes the behavior is grounded on. It is radically different from previous psychological approaches in two key ways. First, it accepts the use of the scientific method, and rejects introspection as a valid method of investigation. Second it posits the existence of internal mental states (such as beliefs, desires and motivations) unlike behaviorist psychology. Cognitive psychology covers a broad range of research domains. Major research areas are presented next:

Perception: attention and filter theories[9], pattern recognition[10]

Memory: short term memory[11] and long term memory[12], episodic memory[13], semantic memory[14], encoding, storing and retrieving memory-based information

Knowledge representation: mental imagery[15], propositional encoding, mental models[16]

Language: grammar[17] and linguistics[18], phonetics[19] and phonology, language acquisition[20]

[9] The ability to focus mental effort on specific stimuli while excluding other stimuli from consideration

[10] The ability to correctly interpret ambiguous sensory information

[11] That part of memory which stores a limited amount of information for a limited amount of time (roughly 15-30 seconds).

[12] The memory that can last as little as 30 seconds or as long as decades. It differs structurally and functionally from short-term memory. Short-term memory is a temporary potentiation of neural connections that can become long-term memory through the process of rehearsal and meaningful association, which leads to a physical change in the structure of neurons.

[13] Episodic memory, or autobiographical memory, is the explicit memory of events. It includes time, place, and associated emotions (which affect the quality of the memorization).

[14] Semantic memory refers to the memory of meanings, understandings, and other factual knowledge; it is the memory of facts and concepts.

[15] A mental image is the representation of an idea in a person's mind. The ability to form and recall mental images, to learn about the world from them, and to communicate to others about them is a defining characteristic of the human species.

[16] A mental model is an explanation in someone's thought process for how something works in the real world. It is a kind of internal representation of external reality, hypothesized to play a major part in cognition.

[17] Grammar is the study of rules governing the use of language.

[18] Linguistics is the scientific study of human language,

[19] Phonetics is the study of sounds (voice).

Thinking: logic[21], formal and natural reasoning[22], problem solving, decision making

This is a way of thinking and reasoning about mental processes, envisaging them like software running on the computer that is the brain.[23] Because of the use of computational metaphors and terminology, cognitive psychology was able to benefit greatly from the research in artificial intelligence.

IV. BOUNDED RATIONALITY AND ECONOMICS

Herbert Simon was interested to find out the way people decide, the decision making process and not the decisions results. He was interested by the limits of rationality, the way organizations, institutions, norms, rules drive the decision process. By decision he means all kind of decisions, from the elementary to the most complex ones. The simplest action is based on a decision. The manner an individual decides and behave is of uttermost importance in economics, and that's why the controversy around this issue is so vividly.

The foundation of neoclassical theory is – Expected Subjective Utility. The individual decides in order to maximize his subjective utility. This means that he is rational and purposeful. The presuppositions of neoclassical economics are: the individual has a well defined utility function and a well defined set of alternative decisional strategies. The decision maker will choose the alternative that maximizes the expected utility given his utility function. This is an elegant theory, but often doesn't correspond to reality. This is the observation made by Herbert Simon. He challenged the classical economic theory that economic behavior was essentially rational behavior in which decisions were made on the basis of all information needed, and where the decision maker attains optimum result. His critique started from the observation of the limited cognitive capability of the individual: (1) there are huge difficulties to gather the necessary information for decision making, and for generating all the possible alternatives; (2) it is difficult to analyze and compare different alternative because of limited computing power of the human brain; (3) it is difficult for the individual to be aware of its own preferences. "It is that human beings have neither the facts nor the consistent structure of values nor the reasoning power at their disposal that would be required, even in these relatively simple situations, to apply SEU principles." (Simon, 1983: p. 17) Herbert Simon said that in today's complex world individuals cannot possibly process or even obtain all the information they need to make fully rational decisions. Rather, they try to make decisions that are good enough and that represent reasonable or acceptable outcomes. He called "bounded rationality" this view of human decision making and described its results "satisficing." The concept of "bounded rationality" reflects the fact that there are limitations in decision making. First, the individual lacks a complete and consistent utility function; second, he cannot generate all the possible alternatives; third, the consequences of choosing an alternative can be evaluated only in a very approximate manner.[24] But is "bounded rationality" a feasible concept? What is necessary for an individual to have "bounded rationality"? The

[20] Language acquisition is the process by which language develops in humans. First language acquisition concerns the development of language in children, while second language acquisition focuses on language development in adults as well.
[21] Logic, is most often said to be the study of criteria for the evaluation of arguments,
[22] Reasoning is the act of using reason to derive a conclusion from certain premises, using a given methodology.
[23] Theories commonly refer to forms of input, representation, computation or processing, and outputs.
[24] The uncertainty about the environment an the cognitive limits of the individuals make impossible to evaluate the impact of his behavior on long term.

minimal requirements are: (1) a mechanism for focusing attention, an emotional mechanism that set the priorities and allow the serial decision making process; (2) a mechanism for generating alternative, because an important part of the problem solving consists in searching and improving alternatives; (3) a mechanism for gathering information about environment and making inferences. (Simon Herbert, 1983: p.21-22)

The consequences of "bounded rationality" for neoclassical theory are devastating. First, this rationality doesn't optimize. The decision process stops when the individual reaches a satisficing result. Second, there aren't any guaranties and its not necessary for individuals' decisions to be consistent. The choice result may depend on the order the alternatives are presented (if, for example, both are satisficing).

AI has its utility in Business. There are four common techniques of Artificial Intelligence used in businesses:

- expert systems – they apply reasoning methods and capabilities, to a large amounts of known information, to provide a conclusion based on them.
- neural networks - AI that are capable of finding and differentiating between patterns.
- genetic algorithms – they generate increasingly better solutions to different problems by applying the survival of the fittest approach.
- an Intelligence Agent - software that assists and help people, in performing repetitive tasks. (Haag et all, 2004)

These results were announced by Simon in 1957 in his address, along with Newell to the Operations Research Association of America: "We are now poised for a great advance that will bring the digital computer and the tools of mathematics and the behavioral sciences to bear on the very core of managerial activity – on the exercise of judgment and intuition; on the processes of making complex decisions" (Simon and Newell, [1957] 1982, p. 382).

Generally, Artificial Intelligence use heuristics for processing information. These are techniques that can solve a problem not necessarily in a correct way, but which usually produces a good solution or solves a simpler problem than the original one. They obtain computational performance or simplicity at the cost of accuracy or precision.

V. CONCLUSIONS

In conclusion we can say that many scientific areas benefited by the work of Herbert Simon. We can mention management, political science, economics, cognitive psychology, artificial intelligence. One of his ultimate goals was to use mathematics to give the social sciences the same rigor like hard sciences - physics and chemistry, but found that a frustrating experience - it seemed that something was always missing when human factors were translated into numbers. The main interests of Herbert Simon's in computer science were in artificial intelligence, principles of the organization of humans and machines is information processing systems, the use of computers to study (by modeling) philosophical problems of the nature of reason and intelligence, and the impact of computer technology in social and economic matters. He admitted that „[...] research in cognitive psychology and artificial intelligence has taught us a lot about how alternatives are generated." (Simon, 1983: p. 22)

His starting point was that computer hardware and software design and the design of human organizations and other social institutions have much in common. He studied the structure of the problems they face and the methods they use to solve these problems. Simon had been particularly interested in the similarities and the limits of parallelism.

He was concerned with building computer simulation models of human processes for handling increasing complex and ill-structured cognitive tasks, including the processes of scientific discovery and the use of visual imagery in thinking, for over forty years.

In economics, Simon dismissed the foundation of neoclassical economics, the perfectly rational and maximizing individual, the later efforts to save these presuppositions made by great economists like Milton Friedman, Armen Alchian, Gary Becker or Robert Lucas being unable to "repair the damage" done. The concept of "bounded rationality" became more and more accepted by economists and constitutes the foundation for a more realistic economics.

REFERENCES

1. Boudon Raymond, (1992). - *Subjective rationality and the explanation of social behavior*, in Economics, Bounded rationality and the Cognitive revolution, Edwards Elgar, Northampton, MA,
2. Egidi Massimo, (1996): - 'Routines, Hierarchies of Problems, Procedural Behavior: Some evidence from experiments', in Keneth Arrow et al. (eds.), *The Rational Foundations of Economic Behavior*, MacMillan, London.
3. Haag Stephen, Cummings Maeve, McCubbrey Donald, (2004): - *Management Information Systems for the Information Age*, McGraw-Hill Irwin, (4th Edition)
4. Holden Constance, (1986). - The Rational Optimist: Will Computers Ever Think Like People?, *Psychology Today*, Oct
5. McCorduck, Pamela. (1979). – *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence.* San Francisco: W. H. Freeman,
6. Mitchell Melanie, (2000): - *Life and Evolution in Computers*, Biophysics Group, Los Alamos National Laboratory, http://www.cs.pdx.edu/~mm/life-and-evolution.pdf, p. 1-21
7. Newell, A., Shaw J.C. and Simon H. ; ([1958] 1989) 'Elements of a Theory of Human Problem Solving', *Models of Thought*, Vol. II, Yale University Press, New Haven, 6-19.
8. Newell, A. and Simon, H. (1976). - "Computer Science as Empirical Enquiry: Symbol and Search," *Communications of the AMC*, 19 (3).
9. Newell, A. and Simon, H. (1972). - *Human Problem-Solving*, Prentice-Hall, New Jersey.
10. Sent Esther-Mirjam, (1997). - *Sargent versus Simon: Bounded Rationality Unbound,* Cambridge Journal of Economics, 21 (3), 323 - 338
11. Simon, H.A. (1996). The patterned matter that is mind. In D.M. Steier & T.M. Mitchell (Eds.), Mind matters: A tribute to Allen Newell (Chapter 11). Mahwah, NJ: Erlbaum.
12. Simon, H.A. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge, MA: The MIT Press.
13. Simon, H.A. (1995). Artificial intelligence: an empirical science. *Artificial Intelligence,* 77(1), 95-127.
14. Simon, H.A. (1995). *Machine as mind*. In K.M. Ford, C. Glymour, & P.J. Hayes (Eds.), Android epistemology (pp. 23-40). Menlo Park, CA: AAAI/The MIT Press.
15. Simon Herbert, (1992), Introductory comment, in *Economics, Bounded rationality and the Cognitive revolution*, Ed. Edwars Elgar, Northampton, MA,
16. Simon H. (1991), "Organizations and markets," Journal of Economic Perspectives, vol. 5, no. 2
17. Simon Herbert, ([1988], 1992). - Scientific discovery as problem solving Economics, in *Economics, Bounded rationality and the Cognitive revolution*, Edwards Elgar, Northampton, MA,
18. Simon Herbert, (1986). - "*Rationality in Psychology and Economics*," in Robin M. Hogarth and Melvin W. Reder, eds., *Rational Choice: The Contrast Between Economics and Psychology*, Chicago: University of Chicago Press
19. Simon Herbert, (1983). - *Reason in human affairs*, Stanford University Press, Stanford, California
20. Simon Herbert, ([1966], 1992). - Thinking by computers, in *Economics, Bounded rationality and the Cognitive revolution*, Edwards Elgar, Northampton, MA,
21. Simon Herbert, ([1964], 1992). - Information processing in computer and man, in *Economics, Bounded rationality and the Cognitive revolution*, Edwards Elgar, Northampton, MA,
22. Simon Herbert, (1957). - *Models of Man: Social and Rational*, New York: John Wiley and Sons, Inc., 279 pp. , http://www.psy.cmu.edu/psy/faculty/hsimon/

23. Soim Horatiu, (1999). - Comportamentul individului. Rationalitate perfecta sau limitata, Studia Universitatis Vasile Goldis, vol 9, seria B, p 257 - 261, Editura Universitatii Vasile Goldis, Arad
24. Viale Ricardo, (1992). - *Cognitive constraints of economic rationality*, in "Economics, bounded rationality and the cognitive revolution", Edward Elgar, Northampton, MA,
25. *** Nobel-prize, press, (1978): - http://nobelprize.org/economics/laureates/1978/press.html