REVIEW

Herbert A. SIMON : **Models of Thought.** New Haven/London, Yale University Press, 1979, xiii + 524 pp.

1. Nutshell characterization. - The book is an annotated collection of thirty-two of Simon's research papers in cognitive psychology originally published between 1955 and 1977. Twenty-one of the chapters were written with one of several coauthors, many of whom are associated, as Simon himself, with the Psychology Department of Carnegie-Mellon University, Pittsburgh, USA. The co-authors are, in alphabetical order : Michael Barenfeld, George W. Baylor, William G. Chase, Edward A. Feigenbaum, Kevin J. Gilmartin, Lee W. Gregg, John R. Hayes, Kenneth Kotovsky, Glenn Lea, Allen Newell, Jeffery M. Paige, Stephen K. Reed, J. C. Shaw, Dorothea P. Simon and Peter A. Simon. The book is presented as a companion volume to Simon and Allen Newell's **Human Problem Solving** (Englewood Cliffs, New Jersey, Prentice-Hall, 1972). Yet, it is self-contained in the sense that it may be read fairly profitably without prior acquaintance with any other works by the author.

Models of Thought documents Simon's long-time involvement with the construction of models/theories for the main components of human thinking and cognition by means of computer programming languages, and with the experimental testing of models/theories of the mental by means of computer simulations of predicted human behavior. The **information processing approach** to cognitive psychology represented by Simon's school is congenial to, but not identical with artificial intelligence. After considering basic elements of the human cognitive apparatus – selective attention, serial and "satisficing", i.e. non-optimizing search, motivational and emotional controls, and storage in memory – the book deals with the issues of learning, problem solving, rule induction and concept attainment, perception, and understanding, in that order. Together, the essays describe a model of "thinking man" (an androgynous being) of awesome simplicity and unity which contrasts sharply with the casual empiricism and the micro-theorizing which characterize most work in mainstream psychology.

Although some chapters are rather technical and/or specialized, the whole book is clearly written and the argument is always interesting. It is accessible to motivated readers with an elementary background in psychology and mathematics. As in previous collections of papers by the author,there is a considerable amount of overlapping between the various chapters and sections, an inevitable consequence of the format chosen (juxtaposition without revision of previous publications). This is not necessarily a drawback, though : some redundancy may help readers – especially those unfamiliar with the subject – to better digest Simon's message.

2. About the author. - Herbert Alexander Simon is one of the few academic intellectuals of our day who has always managed to combine rigor with relevance, and vice versa. It is no overstatement to say that he has succeeded, throughout his career, to fully live up to the old-fashioned, yet often unvalued creed that "philosophy should be done scientifically, and science philosophically." He is also an uncommonly prolific and, by contemporary academic standards, an outrageously versatile author. It will be useful to recall briefly the main stages of his intellectual journey.

Simon is rightly considered the founder of modern Anglo-Saxon organizational theory and public administration theory. His intriguing views on the limited or "bounded" rationality of individual and collective action – roughly, that real human beings, faced with **complexity** and **uncertainty**, lack the wits to optimize, and must be content to **satisfice**, i.e. to find "good enough" solutions to their problems and "good enough" courses of action – and, more generally, his economic behavioralism – not to be confused with psychological behaviorism ! – have earned him the Nobel Prize for Economics in 1978. Simon has been instrumental in bringing the potential of computer science and artificial intelligence to bear on cognitive psychology, the subject of the book reviewed here. We also owe to Simon some highly original and influential work in methodology (e.g., statistical causal analysis), in epistemology and the philosophy of science (discovery, identifiability), and in general systems theory (complexity and hierarchy theory). Finally, he has contributed to the philosophical reflection on the influence of digital computers on culture and society (automation of repetitive as well as non-repetitive actions; the centralization vs. decentralization issue).

3. The information processing approach. - The information processing approach which **Models of Thought** is all about gradually took shape during the postwar years. It is not exaggerated to say that by now, it has changed the very face of cognitive psychology. Information processing may be characterized as a way of dealing with human mental processes, both theoretically and experimentally, which is distinguished from other approaches to cognition by the extensive usage (i) of computer programming languages as tools for expressing theories of the mental, and (ii) of actual computer simulations as devices for testing such theories experimentally. Simon describes the twofold function of computer software in studying psychological phenomena as follows :

> "On the one hand, computer languages have proved to be the most powerful and appropriate languages for stating psychological theories formally. Computer languages have vocabularies and syntaxes specifically designed to describe symbols and symbol structures, elementary information processes, and the organizations of those processes called programs. At the same time, whenever a theory is stated in a computer language, the computer becomes available as a powerful tool for predicting the behavior of the system described by the theory, and thereby permitting a comparison of theoretical predictions with empirical data" (p. 65).

it is good to keep these two aspects conceptually apart, as they do not always go hand in hand. Thus, some rather important insights about the strategies used in chess playing, i.e. a particular kind of problem solving, were gained by using a tentative "program" – a set of completely operational specifications of behavior – whose more immediate characteristics were explored by repeated **"hand simulation"**, long before the program could actually be checked out on the computer (p. 146). Other examples abound. On the other hand, not any arbitrary computer program will be relevant here. For example, a chess playing program which performs at the skill level of a master by making extensive and essential use of the "arithmetic brute force of the computer", as contradistinguished from the much more modest calculating power of the human brain, may well be interesting as an "exercise in artificial intelligence". But is not usually very helpful to those studying the skills of human problem solvers (p. 139).

As a logical consequence, critics of the information processing approach may question either solely the interest for psychology of specific simulations claiming to describe

and to explain one or several human cognitive skills (the second aspect); or they may more drastically challenge the claim that computer languages are appropriate tools and, in fact, the best tools we have to date for expressing psychological theories (the first aspect). Both positions do indeed occur. This brings us to my next remark.

4. Simon's case vs. skeptical arguments. - I said earlier that the information processing approach has changed the face of cognitive psychology completely. This certainly being the case, it must be added immediately that the information processing "revolution" continues to meet with serious (and not so serious) resistance from various quarters : experimental psychology, neurophysiology, linguistics, philosophy, and what-have-you. Even within the larger computer science community, many informed people – very often, not the least renowned ones – are unwilling to accept the optimistic claim that present-day computer simulations already explain a good deal of genuinely human cognitive behavior. Readers interested in a balanced account of the pros and cons of the information processing approach as well as those looking for straight Weizenbaum style polemics[1] will be almost certainly disappointed by the book. For with the exception of a rather short chapter (1.3) on the "Motivational and Emotional Controls of Cognition" (a 1967 reply to a paper skeptic of computer simulation dating back to 1963, when cognitive simulation was still in its infancy) and some occasional polemical remarks scattered over the book (e.g., on the unwarranted application of significance tests to extreme hypotheses as used in Simon's models for the acquisition of concepts, and against behaviorism in general), Simon is pleading his own cause rather monotonically without any more ado. Of course, such a procedure is common – and probably sound – scientific practice. It is efficacious to persuade both sympathetic specialist and the candid layman. But it is frustrating for those readers (no small minority, I bet) who are eager to learn the whole story. The latter should consult other sources as well[2].

5. Some remarks concerning Simon's methodology. - The appeal of Simon's general theory of human thinking depends crucially on its ability "to sum up and explain a bewildering mass of experimental results in terms of a few basic mechanisms" (cf. p. 97). Its challenge derives in part from the successful application of the underlying ideas of seriality (cf. infra), selective search and satisficing to domains which, in the compartmentalized world of science that's ours, are readily perceived as "alien" to the endeavor of cognitive scientists : human organizations, microeconomics (e.g., consumer behavior), evolutionary adaptation, etc.

This observation is of paramount importance to those who put special emphasis on **simplicity** or **parsimony** as methodological criteria of merit for judging concepts, hypotheses, and theories. And this Simon himself patently does. Thus, describing his own research strategy, he writes that "it is a disciplined cumulative strategy, parsimonious in its use of mechanisms and inhospitable to ad hoc solutions" (p. xi). "Adhocery" being a very tricky notion, judgments as to the extent to which Simon has sticked to his word will inevitably differ considerably from one person to another. But it is certainly true that his theory is both coherent and very general, aiming at "a unified explanation of human cognition in all its manifestations" (ibid.) Unfortunately, Simon does not tell us much about the necessary **trade-off** between the generality of a method – the range of tasks to which it may be applied properly – and its power – how well it does on problems to which it is applicable –; a problem that seems to lie at the heart of the debate between the advocates of the information processing approach and those cognitive scientists skeptical of it. Thus at times it looks like Simon wants to have his cake **and** eat it, claiming that his approach

is both the most encompassing one (in terms of applicability) and the most powerful one (in terms of explanatory detail). Maybe this is only a pseudoproblem : by calling on the multi-layered structure of the information processing theory, it might be argued that it is superior to rival theories for any relevant couple of values of the arguments of the power/generality trade-off function. But this the reader can only guess; the explicit structure of Simon's argument certainly does not allow to infer it.

One consequence of Simon's emphasis on parsimony is that he tries to see how far some specific research strategy, say, seriality (i.e., in his case, the claim that any kind of mental activity requiring some attention is essentially serial) can be "pushed" before it has to be modified in important respects. Part of the controversy about the serial vs. parallel functioning of the human information processing system derives from disregarding this methodological stance, e.g. when reference is made to the manifest and important parallel activity of the sensory organs – which nobody, including Simon, has ever wanted to deny.

6. The conceptual foundation of the information processing view. - According to the author, the theoretical and empirical research reported in **Models of Thought** was carried on "within a guiding scheme which, though modified and expanded over the years, still expresses the goals that were sketched out in the early 1950s" (p. x). This conceptual foundation was laid out in detail for the first time in his **Models of Man** (New York, Wiley, 1957), a collection of essays dealing with fundamental problems in virtually all of the behavioral and social sciences (except anthropology) which together constituted an attempt to set forth a consistent general theory of the rational and nonrational aspects of human behavior. (It is precisely in ecological approaches to the study of comparative administration (which rely heavily on anthropology and sociology) that Simon's theory of rationality has been most radically challenged.) In view of what we have said in the preceding section, it was a happy decision to reprint two of these pivotal essays, "A Behavioral Model of Rational Choice" (1955) and "Rational Choice and the Structure of the Environment" (1956), as the opening chapters of **Models of Thought.** In fact, they provide the reader with an indispensable clue for appreciating what Simon and his associates have been able to accomplish hitherto, as well as what they haven't.

The first section of **Models of Thought** introduces specifications which, according to the author, any model of thinking processes must satisfy. First, there is the requirement that the model incorporate mechanisms for coping with complexity even when it is unable to digest it completely. This is the **satisficing** idea mentioned earlier. (In humans, the information processing limitations are the result of their genetic make-up, their phylo- and ontogeny, circumstantial factors, etc.) Satisficing systems include mechanisms that set aspirations and adjust these aspirations upward or downward in the face of benign or harsh circumstances, respectively. Since the alternatives of action are not given but must be sought out, usually in an infinite space, a stop rule ("search ends when a good-enough alternative is found") must be imposed. Secondly, the satisficing rule stipulates that search stops when a solution has been found that is good enough along **all** dimensions, thus avoiding the unrealistic marginalist solution to the problem of trading off incommensurables of traditional optimization theory. Dynamically adjustable aspiration levels guarantee the termination of search without prior knowledge of how rich an environment is being explored. Thirdly, thinking man is **a motivated and emotional creature** as well. Emotion may be captured in information processing terms as an interrupting mechanism that allows an information processor to respond to urgent needs in real time; and motivation as that mechanism which controls attention at any given time. Simon's theory explains how an in-

formation processor endowed with multiple needs behaves adaptively and survives in an environment that presents unpredictable threats and opportunities. Finally, the theory stipulates that the central nervous system of humans is basically organized to operate in **serial** fashion. This most controversial claim is supported by a number of arguments and observations. I can only mention here Simon's reflection that in general, if the components of a parallel system are to operate with a high degree of interdependence, there must be a correspondingly adequate system of **coordination** among them; and that the latter will itself be of necessity a serially organized system. When two attention-demanding tasks are performed simultaneously (e.g., carrying on a conversation while driving a car), this is explained, convincingly, as **time sharing.** Simon also sets out how he conceives the seriality thesis can be held up in the face of such contrary evidence (or supposed evidence) as incubation, or the familiar "tip-of-the tongue" phenomenon (i.e. that the recall of a familiar name or word often occurs only after attention has been turned to other matters), or other instances of parallel "ruminating" activities in the mind. Here, his arguments may be felt to be far less conclusive. Specifically, it seems imperative that simulation models try to account for the recent neurophysiological evidence – which points to **parallelism**[3] –, a strategy that would seem to be more in line with Simon's own in principle reductionism (see chapter 2.3).

It is impossible to discuss these basic claims in any detail in the limited space available here. Let me just make two remarks. First, whether the satisficing idea will be adopted or not would seem to depend, ultimately, on the esthetic standards one happens to hold. The hallmark of satisficing is the non-uniqueness of solutions that is also typical of most real-life problems, including many of those **Models of Thought** attempts to deal with (albeit in rather simplified versions). Now many people – let me call them the "pragmatists" for brevity's sake – will highly value satisficing for the fit between theory and "data" (practice) it quasi-automatically ensures. It is only a small step from the observation that a particular decision has been made to the conclusion that this was a "good enough" decision for the actor to make. This remains true even when it is assumed that aspiration levels are adjusted dynamically. On the other hand, there are "theoricists" (also a great many, to wit) who prefer to stick to an ephemerical theoretical "one best solution" as a kind of Weberian "ideal type", because it allows them – so they assume – to compare the relative merit of different 'good enough" solutions of the same problem, or performances on the same task, etc. The "theoricists" abhor relativism. The "pragmatists" claim that non-uniqueness is a basic fact of life, and that we better face it. Both views are as falsifiable, or as immune against falsification, as one wants them to be. Both are valuable in their own right. The case seems to me undecided, and think it will probably remain undecidable for a long time to come.

My second remark concerns Simon's seriality thesis. The main rationale he offers for it – parallel operation requires coordination or synchronization, which in turn requires seriality – reminds one of the "unity of command" principle dear to an older generation of organization theorists. At a time when computer scientists are turning to organization theory in order to accomplish a "technology transfer" from this field to aid in the analysis and design of so-called **distributed (computing) systems** (i.e. computing devices in which the overall task is decomposed among a collection of separate, essentially parallel processors, such as the **Hearsay-II** speech understanding system)[4], it is worthwhile to look at the fate of the idea of unity of command. Holy to the classical organization theorists of the pre-war era, Simon in his earlier work on organizations had shown it to be a logically unnecessary condition for efficient operation in certain circumstances. Yet he eventually returned to

hierarchy as a basic fact of human organizations[5]. In the 1970s, Mesarovic and his collaborators[6] have demonstrated that a whole fan of "coordination modes" are conceivable to "synchronize" the operations of the separate decision makers which together make for the implementation of the organizational goal(s), and that such a diversity of coordination modes does indeed exist. The moral of their story (and of other work inspired by their example) was that unity of command may be either desirable or dispensed with, according to a number of parameters that ought not bother us here. To the extent that ideas derived from Simon's organizational hierarchy theory are incorporated in the theory of distributed problem solving that is presently emerging, his seriality thesis may well become "true" by the tricky bias of a self-fulfilling prophecy !

7. Implementation of the information processing approach : memory structures, learning processes, problem solving, rule induction and concept formation, perception, and understanding. - The format of the book, which presents in-depth analyses of all these topics, does not allow to reconstruct the author's various lines of reasoning in a few pages. To quarrel about specific, isolated points which the reviewer — a philosopher not specialized in cognitive science — finds particularly important would be utterly arbitrary. Therefore, I shall only enumerate, without discussion, the basic ingredients of the book.

In Section 2 ("Memory Structures"), the **chunk** is introduced as the basic unit for measuring memory capacity. The chunk ("The Magical Number Seven Plus or Minus Two") was proposed by George A. Miller as the relevant unit for expressing the capacity of **short-term memory** (STM), which is characterized by access times in the 200msec range. Simon uses the same unit for measuring fixation in **long-term memory** (LTM), which has much slower access and store times (of the order of seconds) than STM, but a far greater (in a sense, unlimited) storage capacity. The chunk corresponds to anything that has become a familiar unit of experience. LTM as described here is a structure consisting of two components, the **text** and the **index** (cf. an encyclopedia). The contents of STM are symbols that point to corresponding nodes in LTM. Information can be retrieved from LTM either by using the index **(recognition)** or by using sequences of links in the text **(association)**.

Section 3 ("Learning Processes") describes EPAM, the Elementary Perceiver and Memorizer, a theory of human rote learning. It is used to explain the phenomena derived from the verbal learning experiments all psychologists are familiar with. It also includes a warning against confusing psychological parameters with parameters reflecting the experimenter's construction of the stimuli.

Section 4 ("Problem Solving") discusses creativity, the "hindsight" aspect of learning, trial-and-error search in solving difficult problems, algebra world problems, the Tower of Hanoi puzzle, a chess mating combinations program, and related topics. It elaborates on the selective search mechanisms of the Logic Theorist and the means-ends analysis of the General Problem Solver introduced earlier in Newell and Simon's **Human Problem Solving**, and on the heuristics of planning.

Section 5 ("Rule Induction and Concept Formation") proposes a theory of serial pattern induction based on the observation that most complex patterns are generated from a few very simple "components", of which two are predominant : repetition of elements and orderly progression down a sequence of elements. It also discusses process models and stochastic theories of simple concept formation, as well as a unified theory of problem solving and rule induction.

Section 6 ("Perception") discusses such topics as illusions and visual imagery, perception in chess – the MATER, PERCEIVER and MAPP (Memory-Aided Pattern Perceiver) programs –, and spelling skills. The six chapters in this section concern only stimuli that have already been encoded symbolically (e.g., line drawings, strings of letters, etc.); they bypass the problem of "scene" composition.

The final section 7 ("Understanding") mainly describes the UNDERSTAND model for understanding written instructions for abstract puzzles, i.e. puzzles not calling for real-world knowledge. Research of the Carnegie-Mellon group on domains with substantial informational content is not reported in the book because of "limits of space".

8. Concluding remark. - By way of conclusion, I would like to notice that Simon uses the notions of "theory" and "models" as synonyms throughout the book, as was already the case in his earlier work. For the purposes at hand, blurring the conventional distinction between the meanings of these two words seems allowable to the present reviewer. Yet some people might claim that the shift from "models" of cognition (e.g., in the title of the book) to the more compelling notion of "theories" or even "theory" is a **petitio principi** and should therefore at least be justified explicitly. The circumstance that Simon has not done this in no way diminishes his impressive achievement, however : **Models of Thought** is an important and beautiful book.

Werner Callebaut
Aspirant NFWO

NOTES

[1]See Joe Weizenbaum's highly critical **Computer Power and Human Reason** (San Francisco, Freeman, 1976) and his "Where are We Going ? : Questions for Simon" (in T. Forester (ed.), **The Microelectronics Revolution,** Oxford, Blackwell, 1980, pp. 434-438). The latter is a reply to Simon's "What Computers Mean for Man and Society" (originally published in **Science 145,** 1977, and reprinted in Forester, op. cit., pp. 419-433), an exposition of his scientistic views on this matter. Cf. also Simon's most recent statement of his technological optimism in "Prometheus or Pandora : The Influence of Automation on Society", **I.E.E.E. Computer 14** (1981).

[2]For a more balanced discussion, see e.g. Z. W. Pylyshyn's "Computational models and empirical constraints" and the open peer commentary on his position paper in **The Behavioral and Brain Sciences 1** (1978), pp. 93–127.

[3]See, e.g., P. M. Lavorel and H. Gigley, "How we name or misname objects – the simulation of cooperative computation in the human brain", and H. Gigley, "Artificial Intelligence Meets Brain Theory", abstracts, Sixth European Meeting on Cybernetics and Systems Research, Vienna, 1982.

[4]M. S. Fox, "An Organizational View of Distributed Systems", **IEEE Transactions on Systems, Man, and Cybernetics 11** : 1 (1981), pp. 70–80.

[5]See, e.g., V. Ostrom, **The Intellectual Crisis in American Public Administration,** University of Alabama Press, 1974.

[6]M. S. MESAROVIC, M. MACKO, Y. TAKAHARA, **Theory of Multi-level, Hierarchical Systems,** New York, Wiley, 1970.