

INTRODUCTION

The neurones concerned in the behavior of a single man probably exceed in number by a thousand-fold all the telephone lines in the world [...] Even if we knew the exact arrangement of each neuron in a man's brain it would take a model as large as St. Paul's Cathedral to make them visible to the naked eye (E.L. Thorndike, 1914; 1919).

To make a model which would reproduce all the behavior of a rat would require a mechanism probably as large as the Capitol in Washington (C.L. Hull, 1935).

A large building could not house a vacuum tube computer with as many relays as a man has in his head, and it would take Niagara Falls to supply the power and Niagara River to cool it. ENIAC with some 10000 tubes has no more relays than a flat worm (W.S. McCulloch, 1949).

The 'discovery' in the title of this book refers to the coming of age of a new concept of the machine. This concept is at the core of a methodology that profoundly influenced the sciences of the mind and behavior in the twentieth century. The main ambition of this methodology was to overcome, through that new concept of the machine, traditional oppositions between the inorganic and organic worlds, between the laws that govern the behavior of physical systems and those that govern the behavior of organisms, and between causal and teleological explanation.

The origins of this methodology are usually traced back to the middle of the 1940s, with the advent of cybernetics, which Norbert Wiener described, in his 1948 book, as the study of "control and communication in the animal and the machine." That is when the new machines, those equipped with automatic control and with forms, albeit primitive, of self-organization, seemed to suggest a way to escape those oppositions. Mechanical models were built to simulate adaptive behavior as well as various forms of learning of living organisms. The concepts of machine and teleology no longer seemed incompatible.

The fundamental insight of cybernetics, i.e. the proposal of a unified study of organisms and machines, was inherited, starting in the mid-1950s, by AI (Artificial Intelligence). However, AI proposed a different simulative methodology. To put it quite generally, this methodology used computer programs to reproduce performances which, if observed in human beings, would be regarded as intelligent. In the course of the quarter-century that followed, many cybernetics research programmes were cut back, if not altogether abandoned, in the face of the early successes of AI. This is recent history, and still more recent is the history of the revival of some cybernetic projects in the changed context of the 1980s. This revival brought about renewed interest in self-organizing systems, neural nets and connectionism, and influenced later developments in ALife (Artificial Life) and behavior-based and evolutionary robotics.

Some of the main points in this recent history are dealt with in Chapters 5 and 6 of the present book. One of my central claims is that certain basic features of the simulative methodology whose origins are usually put no further back than cybernetics, actually go back in significant ways to the early decades of the twentieth century. In the first four chapters of the book, while examining little-known, if not forgotten or unpublished, texts of the pre-cybernetic age, I investigate various projects involving artifacts that were

considered to be mechanical models of organisms' behavior. I identify in those projects what Marvin Minsky called the "intellectual ancestors" of cybernetics, when he dated the latter's public emergence to the nearly simultaneous appearance, in 1943, of Rosenblueth, Wiener and Bigelow's paper on the teleological features of feedback machines, of McCulloch and Pitts' paper on neural nets, and of Craik's modeling approach in *The Nature of Explanation*. As a result of my investigation, the simulative approaches of the second half of the twentieth century should appear, in Chapters 5 and 6, in a light that should make it easier to grasp those elements, both of continuity and discontinuity, that they share with earlier stages in the discovery of the artificial.

In Chapter 1, taking my cue from the disagreement between Jacques Loeb and Herbert Jennings in the study of animal behavior at the beginning of the twentieth century, I bring to light some features of the discovery of the artificial. We shall see how, alongside the traditional opposition between reductionist and vitalist approaches in the study of organisms' behavior, another approach is adumbrated, in which the behavior of both "inorganic machines" and living organisms can be studied by referring to their functional organization. Furthermore, with the development of new technologies, it is suggested that a positive test for a behavioral theory is provided by building a machine that functions as the theory predicts. According to Loeb, the phototropic automaton (or "electric dog") built by John Hammond Jr. and Benjamin Miessner was precisely an "artificial machine" of this kind: it supported his own theory of phototropism in living organisms or "natural machines." As we shall see, Loeb's theory prompted the design or actual building of several phototropic robots: from Edward Tolman's "schematic sowbug" up to Walter Grey Walter's electronic "tortoises" and Valentino Braitenberg's "vehicles."

The heated disputes of Loeb's days about the opposition, in the study of animal behavior, between the world of intelligence and purpose on the one hand, and that of instincts and automatism on the other, now seem very distant. The same might be said of other more radical disputes, like that between vitalism and mechanism. And yet, various issues that were raised then with regard to both lower and higher animals were later raised with regard to machines. (After all, this is not by chance: had animals not been formerly considered as the true *automata*, lacking purposes and intentionality?) Think, for example, of issue of the legitimacy of describing animals in mentalistic or teleological terms or, if you like, of assuming towards them what Daniel Dennett calls the "intentional stance." As Jennings put it in 1906, "this [kind of description] is useful; it enables us practically to appreciate, foresee, and control [the] actions [of the animals] much more readily than we could otherwise do so."

Various issues recurring in those disputes are dealt with in Chapter 2. Early neurological and psychological associationism and connectionism (a term, it should be remembered, introduced by Edward Thorndike) is the framework for my analysis of early attempts to test hypotheses about the nature of memory, adaptation and learning through mechanical models. This kind of theory testing in the behavioral sciences appeared to the psychologist J.M. Stephens as a genuine *discovery*, when in 1929 he wrote about his "surprise" at finding the "startling possibilities" of building a working artifact embodying the hypotheses of his theory of learning. And the engineer S. Bent Russell,

in discussing his mechanical model of simple forms of learning, had already pointed out that “as the cooperation of workers in different fields of knowledge is necessary in these days of specialists it may be argued that engineers can consistently join [with neurologists] in the consideration of a subject of such importance to man” as it is the study of nervous system. Bent Russell made this suggestion in 1913, about thirty years before the advent of cybernetics, and Wiener would have appreciated his words.

The “robot approach,” promoted by Clark Hull in the context of behaviorism in the 1920s and 1930s, develops along the lines suggested by Bent Russell. The robot approach is the subject of Chapter 3. In this case, one is for the first time in the presence of a coherent research programme with an openly interdisciplinary calling, and of a fully aware simulative or modeling methodology, which raised the hope of progressively building ‘intelligent’ machines. Nicolas Rashevsky, who initially showed interest in Hull’s approach, wrote in 1931 that a machine might be made that could not only express itself in a natural language and carry on a conversation but could also learn a language and even lie intentionally—even if the actual construction of such a machine would have taken “tremendous expense and labor.”

In Chapter 4, I examine Kenneth Craik’s position between Hull’s robot approach and the new-born cybernetics, and discuss Craik’s symbolic theory of thought. I also emphasize how various proposals for ‘reconciling’ natural science and teleology were discussed before the advent of cybernetics. These proposals appear to have been influenced by their author’s acquaintance with some self-regulating artifacts, which were explicitly defined as *teleological* in the above-mentioned 1943 paper by Rosenblueth, Wiener and Bigelow.

I have already alluded to the content of Chapters 5 and 6, in which the notion of artificial as *computational* or *algorithmic*, not merely as *inorganic*, comes into focus with the advent of AI. Finally, in Chapter 7, while bringing out the main thread running through my entire investigation, I state some theses that I believe have characterized the different stages in the discovery of the artificial over the course of the twentieth century. I use these theses to identify the points of convergence and divergence between the different approaches and research trends discussed in my investigation, with the aim of understanding major critical points better. Furthermore, I try to identify in those theses the roots of certain conflicting positions that are still being debated today in much cognitive science and philosophy of mind, concerning, for example, functionalism, the role of representations, and intentional vocabulary.

The present book is not concerned with the history of those automata or robots that were built or imagined before the twentieth century. On the one hand, that history has already been told several times and from various points of view; on the other hand, the ‘artificial’ as understood in the present book is quite remote from the clockwork automata which were especially popular during the eighteenth century. Notwithstanding certain important intuitions of people like Jaques de Vaucanson, such automata could hardly be seen as pre-cybernetic machines endowed with actual sense organs, self-controlling or feedback devices and motor organs. They could not ‘adjust’ their responses to incoming stimuli. As for self-controlling automata, their early history was first carefully reconstructed

by Otto Mayr. These machines are linked to the discovery of the artificial in so far as they exhibit some form of reactive behavior and, albeit in fairly elementary forms, some ability to change their internal organization when interacting with the environment on the basis of their previous 'history.' Hence, these machine exhibit different forms of adaptation, memory, and learning. However, machines of this kind, as well as other artifacts such as computer programs or contemporary mobile robots, are here chiefly analyzed as tools for theoretical investigations into intelligent behavior.

In fact, there are three stories, as it were, that are told in the present book. First, there is the story of building artifacts of different kinds and complexity whose behaviors, if exhibited by organisms, would be regarded as 'adapted,' 'learned,' 'intelligent,' and so forth. Second, there is the story of the attempts to build artifacts that could be considered as tools for building and testing theories about those behaviors. Third, there is the story of the gradually developing discussion about the implications of the first two stories for classical epistemological questions, the mind-body problem, the possibility of scientific theorizing about cognition, and the like. These three stories proceed in parallel, and I have tried to bring out salient conceptual connections between them.

Progress in electronics and computer science and technology allow current modeling approaches to obtain results that inevitably make pre-cybernetic machines, which play a leading role in much of the present book, seem rough and extravagant. These machines almost always are electrical circuits or electro-mechanical robots with fairly simple performances. Nonetheless, the wealth of interesting issues surrounding early stages of the discovery of the artificial is surprising. The present book shows how, with regard to different notions of the machine, many problems were raised which are still being dealt with by scientists and philosophers of mind currently working on organism's behavior and the human mind.

In fact, those pre-cybernetic machines, just like those of the cybernetic age and those in several current areas of research, were aimed at simulating *functions*, rather than at reproducing the *external appearances* of living organisms. The flight analogy argument ("if material is organized in a certain way, it will fly like an eagle; if it is organized in another way, it will fly like an airplane") was already stated by Hull, and would later recur in AI. Since many of those pre-cybernetic machines were built with the ambitious aim of testing hypotheses on organism behavior, they were designed to be working models embodying *theories*, whether psychological or neurological, about behavior. They neither surprise nor seduce us because of the fairly realistic appearance that characterized the automata of former centuries, which chiefly imitated the outward appearance of animals and human beings. On the contrary, their goal is an avowedly non-mimetic one. A behaviorist model of learning by conditioning does not have to salivate like Pavlov's dog; rather, it has to grasp some of the *essential features* of the learning phenomenon. As we shall see, this was the goal underlying the building of simulative models of behavior, a goal clearly stated from the outset: from the days of Hull and his early attempts to simulate conditioned reflex by means of electro-mechanical artifacts, continuing through the days of Grey Walter and his electronic tortoises, and up to much current AI and robotics.

Some of the pre-cybernetic machines described in the present book were rather care-

fully designed, as can be seen in the neat drawings by Bent Russell, or in the circuit diagrams by Norman Krim. Other machines are known to us through rough drawings, like those by Rashevsky, or through rather involved descriptions, like that of Hugh Bradner's mobile robot, or through informal descriptions, like that of Tolman's schematic sowbug. Still other machines were actually built, including those by Krim himself, by Hammond Jr. and Miessner, by H.D. Baernstein, and by Thomas Ross. Exhibited to the public, they were immediately dubbed "intelligent" or "thinking" machines.

The use of these mental terms in the case of machines that today seem so rudimentary stimulates reflections on their public impact and the exaggeration of their actual intelligent abilities—a phenomenon that occurred again and again until more recent times. A "thinking machine" was how William Ross Ashby's homeostat was called in the 1940s, and the very expression used for the first digital computers—"electronic brains"—speaks volumes in this regard. We can well imagine that, when Baernstein's electrical model of conditioned reflex was exhibited at the 1929 Convention of the Midwest Psychological Association, the impression it created was no less than the one created when Grey Walter's electronic tortoises were exhibited at the 1951 Festival of Britain, or when Arthur Samuel's computer program, which played successfully a game of checkers, was exhibited on the 1960 CBS television show *The Thinking Machine*—let alone the recent triumphs of chess-playing programs like Deep Blue, or the World Cup Robot Soccer competitions.

Quite different albeit related considerations could be made in the case of the new 'intelligent' war-machines from the very outset during the First World War. In 1915, for example, the radio-controlled torpedo embodying the simple self-orientation mechanism of Hammond Jr. and Miessner's electric dog was said "to inherit almost superhuman intelligence." The electric dog, initially a "scientific curiosity," Miessner remarked in 1916, "may within the very near future become in truth a real 'dog of war,' without fear, without heart, without the human element so often susceptible to trickery, with but one purpose: to overtake and slay whatever comes within range of its senses at the will of its master." Current development of 'intelligent' weapons make this observation fairly prophetic. And it reminds Wiener's worried judgment on those machines that, as far as their purposeful activity transcend the limitations of the human designer, "may be both effective and dangerous." As Wiener concluded in 1960, the designer could be unable to control a machine whose goal-seeking operation is so fast and irrevocable that, once he become aware that machine's final action has undesired consequences, he had not the time and the data to intervene before such an action is complete.

Various aspects of the discovery of the artificial that are emphasized in the present book have been mostly neglected in investigations on the sciences of the mind and behavior in the twentieth century. A number of authors who play leading roles in this book are not even mentioned or have only a small role in those investigations. Psychologists like Max Meyer or Stephens, let alone engineers or technicians such as Hammond Jr., Miessner, Bent Russell, or Ross, are cases in point. Other authors, who are usually presented as leading figures in the research of the period, are examined in this book in order to bring to light less-known aspects of their research, as in the case of Hull's robot approach or Jennings' views on "inorganic machines." Still other authors, who are tow-

The Discovery of the Artificial
Behavior, Mind and Machines Before and Beyond
Cybernetics

Cordeschi, R.

2002, XX, 314 p., Hardcover

ISBN: 978-1-4020-0606-7