CHAPTER 3

THE ROBOT APPROACH

1. Introduction. Forms of learning

More than once we have had occasion to note that a form of learning which might be described in terms of Pavlov's conditioned reflex was quite familiar to modern psychological and neurological associationism. When the West, especially the United States, learned of Pavlov's work, psychologists posed the question of how conditioning might fit into the framework of the different forms of learning that had already been studied.[1] In Watson's version, conditioning, understood as stimulus substitution, could be explained by the temporal contiguity of the conditioned stimulus to the unconditioned stimulus and to the response to be conditioned, i.e. the response that was evoked initially by the unconditioned stimulus. Watson pointed to temporal contiguity of stimulus and response to explain other forms of learning as well, including selective or trial-and-error learning, which according to Thorndike was based on the law of effect. Thorndike did not ascribe much importance to conditioning as stimulus substitution (or "associative shifting," as he put it). He considered conditioning a very limited form in comparison with selective learning and, in any case, a form based on other principles. From his point of view, if learning by conditioning could result from simple contiguity, more complex forms such as selective learning could be explained only by the strengthening of certain S-R connections as set forth by the law of effect.

The dichotomy that Thorndike saw between his principles of learning and those of Pavlov led to various classifications based on the differences between the two forms of learning, not to mention attempts to proceed in a direction other than the reduction of both to simple temporal contiguity, as Watson proposed. According to Stephens, for example, learning by conditioning could be fully explained by, and hence reduced to, the law of effect, which remained the sole law of learning. Stephens found confirmation in the functioning of his mechanical model, where every response of Pavlov's dog to a stimulus to be conditioned that did not have the appropriate effect, namely greater adaptive value (such as the response that prepared for the intake of food, namely salivation), did not become a conditioned response, and the corresponding S-R connection was stamped out.

Clark Hull suggested that the two forms of learning, selective and by conditioning, were particular cases of a single principle, that of primary reinforcement, by which it is

[1] Classic accounts of the development of different learning theories, before and after Pavlov's ideas came to the West, are given in Hilgard and Marquis (1940), Postman (1947), and Spence (1951). Western psychologists first became familiar with Pavlov's work in 1906, when *Science* published the text of one of his lectures, and R.M. Yerkes and S. Morgulis described his experiments in an article in 1909. But it was Watson who familiarized the West with the principle of the conditioned reflex in the study of learning, which he wrote about in *Behavior* in 1914. When Pavlov's book was translated (Pavlov, 1927), his work became well known to American behaviorists. For an introduction to the subject, see Mecacci (1979).

always and only the S-R connection that is strengthened, and this leads to the reduction or elimination of an organism's drive. The hungry cat in the Thorndike box selects and fixes one response among several evoked at random: the one that is followed by primary reinforcement in the form of food intake. Such a response leads to a reduction in the drive and has the maximum adaptive value for the animal. The case of Pavlov's dog is more complicated but does not entail a different principle, because in conditioning too the learned response leads to the reduction of a pre-existing drive. What changes is the situation in which the principle is applied. Actually, in conditioning it is not one of the animal's several possible responses that is followed by primary reinforcement but a single response; and what is more, this response is not one the animal selects but one that is induced by the experimenter, namely salivation. Since salivation has always been in close temporal contiguity with food intake (or primary reinforcement), it acquires its own so-called "secondary" reinforcement capacity.

Hull, like Stephens, wanted to interpret the law of effect in objective terms: reformulating the law in terms of the organism's drive reduction meant freeing such a law from Thorndike's controversial references to pleasure and discomfort. When reviewing Thorndike's *Fundamentals of Learning*, Hull said that he referred initially to such an objective law in his own investigations and interpreted it as the law of conditioning (Hull, 1935a: 821). What in this case was central to the explanation of learning was reinforcement in the presence of a drive and not the subjective component of motivation. It is sufficient to reconsider the formulation of the principle of primary reinforcement to see that Hull shared Thorndike's neurological hypotheses of connectionism, namely, that learning always occurs through the strengthening of some neural connections or bonds and the weakening of others. But Hull was never particularly interested in the neurological details of connectionism. His writings only occasionally make reference to results from neurology and do not contain the frequent drawings of the nervous system taken from neurological texts of the time which Thorndike usually included in his writings. Hull picked up a distinction that Edward Tolman had made and maintained the study of the laws of learning at a markedly "molar" level, i.e. he thought that the laws should be drawn chiefly from the study of the overt behavior of organisms, given our scant understanding of the laws at the neurophysiological or "molecular" level.

Thus, Hull claimed that he had started out by objectively interpreting the law of effect as conditioning with reinforcement, and that in this way he was able to eliminate Thorndike's motivational references from the explanation of learning and adaptation. Hull also thought he had made explicit an aspect neglected by Pavlov, namely that reinforcement is always connected to the reduction of a drive of the organism. For Hull, though reinforcement was the basis of adaptation and learning, it was a wholly "automatic" mechanism: it regulates the organism's dynamic relations with the external environment, and it is solely because of its greater complexity that it differs from simple reflexes or the many mechanisms that regulate the organism's internal environment (the ones Walter Cannon termed "homeostatic"). Hull's made continuing efforts in his theoretical and experimental investigations to explain this complexity on the basis of the principles of associationism and connectionism. In one of the first articles he wrote on the subject, he described the conditioned reflex as "an automatic trial-and-error mechanism which mediates, blindly but beautifully, the adjustment of the organism to a complex environment" (Hull, 1929: 498). This marked the first step towards integrating two forms of learning in a single set of assumptions. At the same time, Hull insisted that in so far as it could be explained or "deduced" (the term he preferred) from the principles of associationism and connectionism, *every* form of learning, however complex, was a purely physical process that involved increasingly complex degrees of automatism.

Hull explained that he used "blind" in the sense that where learning or reinforcement of neural connections is concerned, "it is not assumed that there is available for its guidance and control any disembodied soul or spirit," not in the sense that these processes work without "recognized principles" (Hull 1930a: 250). No Ghost Theory could take the place of the search for the molar principles of learning. Hull dealt with the problem that we have been looking at since Chapter 1 in terms of the contrast between automatic and plastic control of behavior. He did so by opposing the Gestalt psychologists, who claimed that the different forms of learning could not be explained by the principles of associationism and connectionism. Nonetheless, Hull did not base his defense of those principles on the bare hypotheses advanced by such behaviorists as Watson, for whom the automatism of the nervous system was reduced to the possibility of activating different pre-existing chains of reflex arches. As we shall see, Hull tried, instead, to do justice to certain ideas of the Gestalt psychologists.

In his later work too, Hull continued to defend the hypothesis of the automatic nature of all forms of learning. But in the late 1920s and mid-1930s, he urged that the hypothesis could be corroborated by results from a new kind of investigation conducted alongside experimental psychology. The point was to start building machines that could simulate the simplest learning processes in order ultimately to build others, which Hull called "ultra-automatic" or "psychic" machines, with a hitherto unexplored degree of automatism that could simulate the most complex kinds of learning typical of higher organisms and humans. In the course of our investigation into the discovery of the artificial we have already seen some proposals along the lines of this simulative methodology, including the machine as a test of a psychological theory, the functional comparison of machine and organism, the idea that successful simulation of the most *elementary* forms of learning would ultimately make possible the simulation of *complex* forms, and the interpretation of simulation itself in an anti-vitalist vein. In the case of Hull and some of his followers, this method took the form of a particularly explicit and consistent proposal, even in light of the evolution it underwent beginning a decade later, in the cybernetic age.

The present chapter will describe the development of the simulative methodology of Hull and other researchers from a different background, which hark back to the idea of the automatic and mechanical nature of mental processes. The electromechanical devices and robots they built, primitive forebears of the synthetic animals of cybernetics, were meant to embody some of the principles that inspired Hull's theoretical and experimental investigations of learning. We shall be looking at least at those aspects of such investigations which help to clarify the aims of those researchers. We shall conclude the chapter with a consideration of the reasons why Hull abandoned his simulative approach.

2. Simulating learning processes.

Hull was a psychologist who had studied logic and engineering. He also designed a machine for automatically producing syllogistic conclusions. But according to his *Idea Books*, the diary he kept regularly all his life, at least as early as 1927 he was thinking of an altogether different machine that could simulate the processes of learning by conditioning in living organisms.[2] It seems that Hull publicly launched the idea of this kind of machine in 1928, almost as a challenge to those attending one of his weekly University seminars, at which theoretical problems in psychology and the laws of learning were discussed (Hull later said that "my suggestion [was] that a model might be made to test the theory"). The following week, three devices were brought to him that reproduced the conditioned reflex: two were arrangements of very crude wooden levers, but one, designed by H. D. Baernstein, a chemist who had been at the earlier seminar, was particularly ingenious. It was displayed in May 1929 at the Midwest Psychological Association Conference in Urbana, Illinois. Baernstein's device attracted wide interest, and the press referred to it as a "mysterious mechanical brain," the forefather of a whole generation of "thinking machines."[3] The aim of Hull's challenge was explained in an article he and Baernstein published in July of the same year in *Science*. They situated their project within the "mechanistic tendency of modern psychology" and described it in the following terms:

> If it were possible to construct non-living devices—perhaps even of inorganic materials—which would perform the essential functions of the conditioned reflex, we should be able to organize these units into systems which would show true trial-and-error learning with intelligent selection and the elimination of errors, as well as other behavior ordinarily classed as psychic. [...] Learning and thought are here conceived as by no means necessarily a function of living protoplasm any more than is aerial locomotion (Hull and Baernstein, 1929: 15).

Hull later remarked that "there was a time when the properties of aerial locomotion were associated only with organic life," but now it was clear that "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." In general, "it is only a question how the material is organized that determines how it will behave" (quoted by Gray, 1935-36b: 413).

[2] The description of the syllogistic conclusion machine, which was also mentioned by Martin Gardner (Gardner, 1982) was never published, and Hull referred to it only occasionally. In Hull's autobiography he said that it was made of "concentric sheet-metal plates" (Hull, 1952: 146)—a kind of Lullian machine. But Hull had already distinguished between a logical machine and a "psychic" one (as he called it) and made it clear that the former had nothing to do with his hopes for the latter, which were to simulate the versatile and adaptive behavior of a living organism. "When, and if, this takes place the thinking mechanism will surely be a far more subtle and complete character than a mere logic machine consisting of sliding disks. [...] Fertility, originality, invention, insight, the spontaneous use of implements or tools—these things, clearly, do not lie in the syllogism" (Hull, 1935b: 220 and 219 n.). Another of Hull's automatic machines, this one for calculating statistics (a newspaper report was headlined "Machine does year's work in a day"), is in the Smithsonian Institution and dates to 1925 (C. Eames and R. Eames, 1990: 89). Hull died in 1952 and his *Idea Books* (Hull, 1962) were published posthumously, selected and edited by his secretary, Ruth Hays. The present chapter refers to Hull's work from the late 1920s to around the mid-1930s. For the evolution of Hull's thinking, see the excellent analysis of Smith (1986).

[3] This information comes from "Thinking Machines," an article that appeared in the popular journal *Harper's Magazine* in 1936 (Gray, 1935-36b).

Certain ideas were clearly in the air, and Hull and Baernstein's proposal was fully accordant with the approach Stephens independently proposed that very year, i.e. the "synthetic approach" to the study of learning apart from the features of living protoplasm.[4] Both Hull and Stephens were thinking about "non-living devices" that had a specific capacity which traditional machines did not have, that of changing their internal organization in response to an outside stimulus and varying behavior as predicted by a learning theory. "In designing a learning machine," Hull wrote in his *Idea Books* in 1927, "I felt the necessity of providing a device to vary the reactions so that trial-and-error could take place effectively" (Hull, 1962: 823). Hull's project looked like the first realization of the hope Bent Russell expressed in 1913 that psychologists, neurologists, and engineers might work together in the study of the principles of learning—a genuine "robot approach," to use Hull's own expression.[5] Hull was referring to Russell when, in introducing another device—that built by Robert G. Krueger, an electro-technical engineer interested in psychology—he wrote that this was "not the first time that a model intended to parallel adaptive behavior has been designed" (Krueger and Hull, 1931: 262).

Baernstein's and Kreuger's devices were intended to simulate several features of conditioned reflex learning. These devices might be considered as the "units" about which Hull and Baernstein spoke in the *Science* article: when organized in more complex systems, these would display several forms of behavior typical of higher organisms. Actually, what is typical of the robot approach is the idea that successful simulation of elementary learning processes is the premise for simulating higher mental processes.

To get an idea of how these devices worked, it suffices to look at the simplest one, built by Krueger. As the figure in Plate 3.1 shows, the device consisted of a set of electric circuits. A charged battery E is inserted in the first circuit and uncharged batteries E_1, ..., E_5 in the others. When switch S_u is pressed, lamp L lights because battery E is charged. This reproduces the simple or unconditioned reflex, where S_u is the unconditioned stimulus (food in the case of Pavlov's dog) which always evokes the same reflex response (salivation at the sight of food), and this corresponds to the lamp going on. Since the other batteries are uncharged, lamp L does not light if any other switch S_1, ..., S_5 is pressed. These may be considered as neutral stimuli to be conditioned (a sound, a light and so on). The simulation of conditioning in its simplest form takes place if one imagines pressing switch S_u and one of the others, say S_1, simultaneously. In this case the current from battery E flows through the lamp and battery E_1 as well, and charges the latter. This operation simulates double stimulation during conditioning and gives sufficient charge to battery E_1 that when switch S_1 alone is pressed, the lamp glows, although faintly. In other

[4] The description of another electric machine simulating some features of conditioning was published in 1930 but "for purposes of demonstration in introductory psychology," as the author put it (Walton, 1930: 110), rather than to test a psychological theory, which was the explicit aim of Hull and Stephens. It should be noted, however, that some researchers' altogether *pedagogical* interest in machines (Meyer was considered in the previous chapter in this connection) sometimes rears its head in other authors who seem more interested in using machines in connection with a *theory* of behavior.

[5] As we shall see (in section 7 below), Hull adopted the expression "robot approach" several years later in his *Principles of Behavior* (Hull, 1943: 29-30). Notwithstanding, we use this expression here because it captures the spirit of the simulative approach Hull adopted from the late 1920s to the mid-1930s.

http://www.springer.com/978-1-4020-0606-7

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