The Discovery of the Artificial. Some Protocybernetic Developments 1930–1940

Roberto Cordeschi

Istituto di Filosofia, Università di Salerno, Italy

Abstract. In this paper I start from a definition of "culture of the artificial" which might be stated by referring to the background of philosophical, methodological, pragmatical assumptions which characterizes the development of the information processing analysis of mental processes and of some trends in contemporary cognitive science: in a word, the development of AI as a candidate science of mind. The aim of this paper is to show how (with which plausibility and limitations) the discovery of the mentioned background might be dated back to a period preceding the cybernetic era, the decade 1930–1940 at least. Therefore a somewhat detailed analysis of Hull's "robot approach" is given, as well as of some of its independent and future developments.

Keywords: Cybernetics; Artificial Intelligence (history of); Mechanistic psychology

Protocybernetics and the Culture of the Artificial

If one were to attempt to define the notion of the "culture of the artificial" which is prevailing in this century, it would perhaps be possible to define it on the basis of some dominant traits which characterized AI, as a candidate science of mind, during its brief history. I would be tempted to summarize those traits point by point as follows:

(1)   Mental processes are independent of the features of organic structure; more precisely, there is a *functional equivalence* between their instantiation in an organic structure and their instantiation in an inorganic or artificial structure.

(2)   The hypothesis that mental processes are mechanical is verified (or tested) by the building of models that reproduce them. This "synthetic" or model method gives the proof of the *sufficiency* of the processes postulated by the hypothesis to explain intelligence, intentionality and so on, and of the futility of introducing explanatory "non-naturalistic" entities.

(3)   The principles behind explanations of "simple" processes of intelligence are the same as those at work in the "higher" processes. The successful artificial

reproduction of the former opens up the way to the ("empirical") hypothesis that from this starting point it will be possible to simulate complex processes which are more commonly known as "psychic", by means of a procedure of accumulation and generalization of the positive results reached which is comparable in principle with the "mature" sciences. It is thus not impossible to deal with the *complexity* of psychic phenomena using ordinary analytical criteria.

(4) knowledge of the external world of an organism occurs through *representations*. These representations are variously viewed – as identical to the functional states or as their semantic content. In any case they concern physical structures, which as such can in principle be implemented in particular artefacts.

(5) The observer can legitimately describe the processes of the artefact (seen as an "intentional system") using a *mentalistic* language (like the one commonly used in the higher organisms), at least in order to control and predict its behavior.

When I said I intended to make a point by point summary of the culture of the artificial, I emphasized that I was referring in particular to what may be inferred from AI as a science of mind.[1] If one considers AI exclusively as a branch of engineering, concerned with emulating and extending certain human capacities, then almost none of points (1)–(5) proposed above would be relevant. But let me be more explicit on this point. In the present paper my aim is to show that the discovery of the notion of the artificial as is defined by (1)–(5) goes back to an era preceding AI, indeed preceding cybernetics itself: the era that we would define here of protocybernetics. Are all the above points relevant in order to define the "culture of the artificial" in the protocybernetics era? Obviously, one should be clear on the nature of the artificial: in the protocybernetic era – but also in the cybernetic one – the "artificial" is essentially and generically the "inorganic"; in the more recent times of information processing psychology and cognitive AI the "artificial" is specifically the "computational". This should be kept in mind, for example, when considering point (4). The physical structures referred to in point (4) are *symbolic codes*: as such they are specific to classical (symbolic) AI, and not proper to any preceding mechanistic trends in the science of mind. In fact it is on this point that computational cognitivists of the so-called classical view distinguish themselves both from old cybernetics and from new connectionistic AI (see for example, Pylyshyn, 1989). And not by chance do they oppose to neoconnectionists using the same argument which they assume to be conclusive, and use against connectionism from Hull to Hebb to Osgood, i.e. what I would like to call the protocybernetic connectionism of the thirties (i.e. of the era of the "discovery" of the artificial as defined above).[2] It is thus understandable that Fodor and Pylyshyn find themselves "with a gnawing sense of déjá vu" when concluding their well-known critical analysis of contemporary connectionism (Fodor and Pylyshyn, 1988). Therefore, we should be careful to avoid confusion when extending the notion of "culture of the artificial" to the period preceding classical cognitive AI.

Discovering the Artificial: Clarke L. Hull's Robot Approach

Let us start from the so-called "robot approach", to which Clark Hull dedicated himself between the late twenties and 1935, and some of its independent and later

developments.[3] A psychologist with a background of studies in logic and engineering, Hull had designed and constructed a machine which found syllogistic conclusions automatically. However, from at least 1926 onwards he was already thinking of a design for a different kind of machine which could stimulate higher functions.[4] It appears that Hull managed to realize his project for the first time after a chance meeting with a young chemist, H. D. Baernstein. Their model claimed to reproduce numerous characteristics of Pavlov's conditioned reflex. Many of the main points (which I shall return to frequently) of my definition of the culture of the artificial are summarized in the brief preview of the project which Hull published with Baernstein in *Science* in 1929. Placing their project against the background of the so-called "mechanistic tendency of modern psychology", the authors presented it as a "synthetic verification" of the behavioristic hypothesis of mental functions through the "inorganic" simulation of the characteristics of the conditioned reflex. The methodological inspiration of the project was summarized as follows in *Science*:

"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. Thus emerges in a perfectly natural manner a direct implication of the mechanistic tendency of modern psychology. 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, pp. 14–15).

These ideas must have been around to a greater extent than Hull himself thought. In the same year another psychologist, J. M. Stephens, proposed a similar "synthetic test" of a psychological theory, this time related to a learning theory based on Edward Thordike's law of effect.[5] Also according to Stephens it was possible to make a "synthetic" approach to the subject, an approach which is comparatively independent of any "analysis" of the living protoplasm. It is worthwhile rereading the passage below. In his pioneering work, Stephens was surprised to discover that the mechanistic metaphor suggested a possible new conception of the relationship between psychology and neurology and a fresh possibility of testing a psychological theory:

"The above conception occurred to me as a means of evaluating those expositions of learning couched in neurological terms. This test of possible synthesis had, at first, only theoretical interest. To my surprise, however, I found that some of the analyses presented startling possibilities of mechanical synthesis. . . . I have tried to use no explanation of animate learning which could not be considered to work in a machine. I have considered protoplasmic organisms as very complicated machines and nothing further" (Stephens, 1929, p. 423).

"Very complicated machines" therefore. But what *kind* of machines? Both Hull and Stephens were obviously thinking of machines capable of *modifying their own internal organization* and therefore of showing a certain degree of *variation in their behavior*, a variation that originated in learning. Obviously it was not a question of simulating a simple reflex action, but of reproducing the mechanisms on the basis of which reflexes are *modified* in the way suggested by Pavlov and Thorndike's connectionism.[6]

The main point, therefore, was *memory* – how could a machine have been given any. You will recall that the objection in principle by the antimechanist *par excellence* of the time, William McDougall, lay in the impossibility of building

machines that could modify their behavior on the basis of memory.[7] To give you an idea of how the contemporary mechanists were beginning to approach this problem, let us return to Hull, in order to get a closer look at the function of the model projected in particular by Baernstein. Their device could change its internal organization thanks to the presence of mercury toluene thermoregulators in the form of "U"-shaped tubes, where the toluene, by heating or cooling, caused the mercury level to rise or fall, and thus close or open the electric circuit respectively (Fig. 1a). In practice, they actually represented time–lag relays included in an electric circuit of which I am providing a simplified illustration in Fig. 1b. In the circuit we can see push buttons S_u, S'_c, and S''_c. If one presses push button S_u, lamp R lights up via *first* pair of contacts. This produces the simple *un*conditioned reflex (Pavlov's classic food-saliva case). Vice versa, the lamp does not light up if one presses one of the push buttons S'_c or S''_c (which correspond to neutral stimuli, or stimuli to be conditioned: a bell, light and such like are typical examples). But look what happens when we press, for example, S'_c for a couple of seconds: the *first* pair of contacts allow the current to pass through the heating wire surrounding the "U"-relay. This is marked C_1 in the diagram, and the wire passes alongside it. In this way the toluene expands and the mercury rises and closes the circuit. Next we press S_u: its *first* pair of contacts close the main circuit and cause R to light up; at the same time the *second* pair of contacts send a current through C_1 to the heating wire surrounding C_2. Two wires are connected to C_2, a longer one from S'_c, and a shorter one from S''_c; the heating up of the toluene causes the mercury to rise in C_2 until it reaches the first of the two wires inside it (the longer one from S'_c) and closes the circuit. At this point, if we press push button S'_c, its *second* pair of contacts send a current to R, which now lights up.
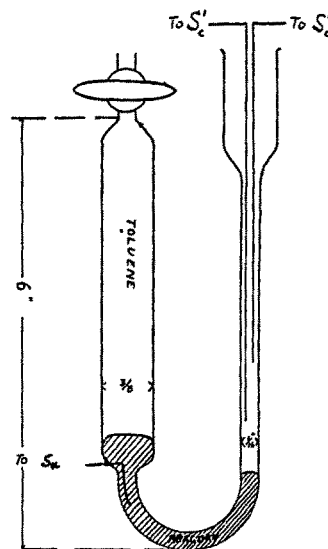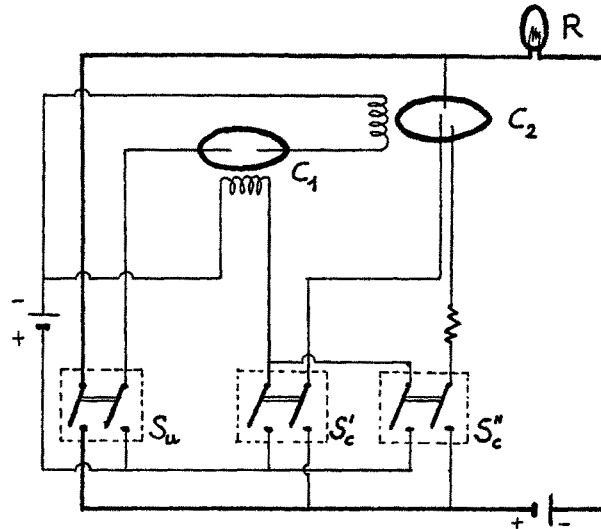

Fig. 1a. From Baernstein and Hull (1931) (Modified).

Fig. 1b. From Baernstein and Hull (1931) (simplified).

In this way a *substitution of the stimulus* is created (S_u and S'_c, in the example), which is typical of simple conditioning.

As I said, the original circuit involves several complications leading to "inorganic" reproduction of various aspects of the theory of conditioned reflexes. However, not all these aspects are simulated adequately by the model. For example, an *identical* behavior of the model can be interpreted on the basis of completely *different* principles of the theory. This is the case when the model's behavior seems to reproduce the principle of redintegration. As I mentioned earlier, conditioning in its classic formulation can be seen as a case of associative learning, consisting of the "association" of an (unconditioned) stimulus with another (first neutral, then conditioned) one, occurring in strict temporal contiguity. In the case of redintegration the unconditioned stimulus is associated with a group or "complex" of conditioned stimuli, each of which is then able to replace the former separately when evoking the response.[8] If now, to simulate redintegration in Baernstein and Hull's circuit, we press *simultaneously* S'_c and S''_c, as a conditioned stimulus complex, and then we press S_u in the end the mercury will rise in C_2 until contact occurs between both wires in C_2; first with the longer one from S'_c, and then with the shorter one from S''_c (see again Figs. 1a and b). From that moment (for a certain period) in the conditioned response (the lighting up of R) will follow the pressing of S'_c or S''_c *separately*, following, so it appears, the principle of redintegration. In reality, as can be easily seen, this latter phenomenon (the lighting up of R by pressing S'_c and S''_c *separately*) could be obtained in any case, even *without* including S''_c initially in the conditioning process. But this behavior of the model can be explained, in Pavlov's theory, on the basis of a *different* principle, the principle of irradiation (Pavlov 1927, pp. 152 ff.). According to it some unconditioned stimuli can evoke the same (although weaker) response through a process of stimulus generalization (and precisely the

resistance that in Fig. 1b is included in the circuit of the second pair of contacts of S''_c gives place to the weakening of the "irradiated" response).

It is not surprising that for a certain period Hull and those who were inspired by this approach dedicated themselves to designing circuits that would reproduce the various features of Pavlov's theory in as realistic a way as possible: the basic principles that explained the change of performance of an organism were recognized in these features. A little later Hull himself collaborated with Robert Krueger, an electrical engineer, designing another circuit that, to keep within the bounds of the above example, provided a better simulation of redintegration. Conceptually Krueger's model was not any different to Baernstein's; the novelty lay in the fact that the learning process did not occur via a modification of the "strategic points" (as Krueger and Hull called them) of the connections (the relays in Baernstein's model) but via storage cells that, initially uncharged, were charged when the conditioned stimuli buttons were pressed (analogue of learning), were discharged (analogue of extinction without reinforcement) and were spontaneously reactivated to a greater or lesser extent (analogue of recovering). An analysis of Fig. 2 (which reproduces the original wiring diagram of Krueger and Hull's model) suggests as, in a sense, my conclusion that we have here a better parallel of redintegration is justified.[9]

It is not by chance that I have been emphasizing the "inorganic" simulation of redintegration and its possible refinements. At that time, it appears that Hull maintained it was essential for the purpose of investigating a mechanism underlying several types of trial-and-error learning. This was what he defined as the "divergent excitatory mechanism" in which a single stimulus was connected with varying strength to a group of mutually incompatible responses. Normally one of these would correspond to the extinction of a need and be followed by (since it is an adaptive response) a reinforcement which strengthens the connection between the response and the stimulus. These incompatible responses could be single simple acts, and this would then be a form of trial-and-error learning (called "simple") in which a single act from among those possible is enough to bring success or extinction of the need. It is that single act that the organism ends up learning after a process of random selection; or (and this is
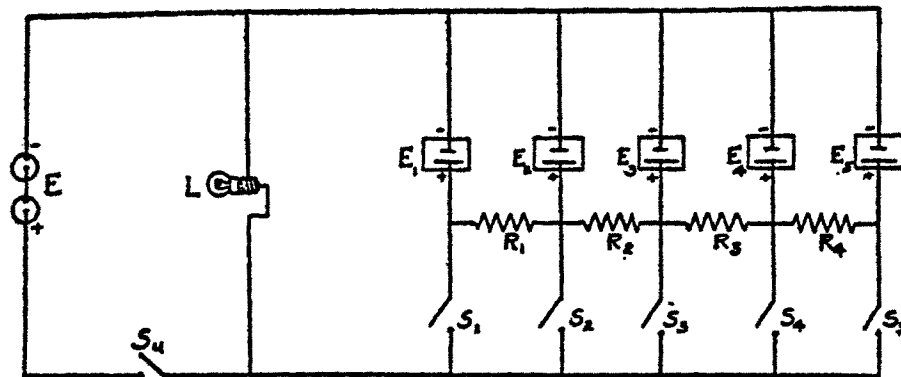

Fig. 2. (From Krueger and Hull (1931).)

more frequent) the responses are not simple acts, but particular sequences or chains of simple acts: and it is one of these (or rather, to use Hull's words, a "family" of these) that leads to success or solves the problem. This second type of trial-and-error learning was later to be named "compound" by Hull (1942).

In spite of a certain air of artificiality about this distinction (how does one define a "simple" act?) one could perhaps render it heuristically useful by referring the first type of learning (the "simple" kind) to lower organisms or to elementary type problem-box situations and the second ("compound") type to higher organisms or maze learning situations. Nevertheless both are based on the same divergent excitatory mechanisms which, as I said, Hull believed could be "derived" (his expression) from the redintegration mechanism. In fact Hull is not always consistent in his presentation of his "deductions" (and in later years he also used other expository criteria). Nevertheless, a way of reconstructing them (at least for my present purposes) could be as follows.

As we have seen in redintegration, a single stimulus belonging to a complex of stimuli simultaneously conditioned to give a certain response, can elicit that response even when it occurs separately. Let us assume we have a sequence of stimulus complexes which elicit a corresponding sequence of responses. For instance, let us take the case of a behavior sequence consisting of single alternative actions, e.g. in simple trial-and-error (Fig. 3). Let S_1, S_2, S_3 be the stimuli from the external environment, and R_1, R_2, R_3 the organism's responses. For the sake of completeness we shall also indicate the (internal) proprioceptive stimuli, s_1, s_2, caused by the responses. Let us assume that in any redintegrative stimulus complex (dotted rectangles in Fig. 3) there are: (1) stimuli like S_i (external) and s_i (internal) that vary from one complex to the next (thus proving conditioned by one single response in the sequence); and (2) a stimulus persisting in each complex. Let us take it to be, for example, the (internal) drive stimulus S_D triggered by an animal's need for food. It persists while the other stimuli vary, and is thus conditioned to *all* the responses in the sequence. Let one single response, e.g. R_3, be followed by reinforcement (food reducing the need). Even assuming that the S_D–R_1 and S_D–R_2 conections initially possess greater excitatory potential, subsequent failure of reinforcement will, according to the principle of experimental extinction, lead to the weakening of these connections, while the presence of reinforcement in S_D–R_3 will lead to the strengthening of the latter (possibly including strengthening through the spontaneous recovery phases of the two other responses, due to the time intervals between the successive cycles of trials).
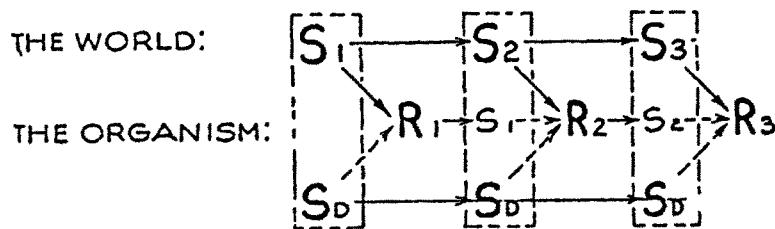

Fig. 3. (From Hull (1931), modified.)

Thus, to the extent that S_D proves conditioned by every response, it is also able to elicit them all according to the various strengths of the respective connections. The divergent excitatory mechanism consists precisely in this. In practice we may see it emerging from a "persisting core" (Hull's expression), i.e. S_D, in the redintegrative stimulus complexes present in a behavior sequence.

In a well-known theoretical article published in those years, Hull (1930a) presented a detailed analysis of the divergent responses mechanism present in an organism which we might define as redintegrative. Analysis was made of the theoretical strengths of S-R connections and the theoretical values of reinforcement, experimental extinction and spontaneous recovery at various intervals of time and in relation to the organism's so-called "previous history" (in practice, the strengths of the various S-R connections prior to learning). The behavior of this hypothetical, redintegrative organism, Hull observed, showed "the persistence of effort at solution by means of varied response". As he pointed out, this is a type of behavior that is generally defined as "intelligent" or "purposive" in the case of organisms, in order to contrast this "psychic or mental" behavior (recall McDougall) with that of "ordinary automatic machines" (Hull, 1930a, p. 255). However, for Hull the interest appears to have lain elsewhere: if his description, or "deduction" as he put it, of simple trial-and-error learning was indeed correct, it followed that this "purposive" or "intelligent" form of behavior could be deduced from the basic principles of associationism and connectionism.[10] Significantly enough, the first simple models we saw above were created to embody precisely these principles. It was therefore to be expected that the construction of a model embodying simple trial-and-error principles would come up against the problem of how best to "combine", to use his terminology, the more elementary models created (a "combination", like the "deduction" process, entailing no qualitative differences).

Concluding his description of Krueger's model, Hull gave a list of open problems that future research would have to deal with (Krueger and Hull, 1931, p. 268). Among these was the simulation of the persisting stimulus (the type exemplified with S_D above), which lay at the basis of the simple trial-and-error divergent response mechanism. It must, indeed, have been a satisfaction for Hull to see two Yale University researchers, George Bennett and Lewis Ward, construct a device embodying the divergent response principle a few years later (Bennett and Ward, 1935). Two lamps are included in the wiring diagram of their model: they never light up simultaneously when, during the process of "conditioning", one turns on the various switches (as usual corresponding to the various stimuli – unconditioned and conditioned). Briefly, we have a machine which, once "conditioned", is capable of choosing between *two* antagonistic responses: as a simple organism confronted by a problem-box situation. Logically developing Hull's methodology of the artificial, Bennett and Ward held that if it was possible to create a device reproducing such an extremely simplified problem-box situation (with only two alternatives), it might well then be possible to use similar criteria to simulate artificially the divergent response phenomenon in more complex cases like maze learning (what I have described as "compound" trial-and-error learning).

Two year later, in 1935, psychologist Douglas Ellson published his description

of a second simple trial-and-error model conceived along totally different lines to reflect Hull's theoretical assumptions more closely (Ellson, 1935). The model consisted of three electromagnets of varying initial strength linked to devices that could subsequently increase or reduce their strength (basically by automatically varying the number of turns in the respective windings). These were radially arranged around a central, spring-suspended soft-iron pendulum. On closing a circuit activating the three magnets the pendulum was attracted by the initially strongest magnet, which we shall call X. In order to reproduce the situation described by Hull, the pendulum should be "conditioned" to move in the direction of the magnet chosen to represent the goal, let us call it Z, which had the least initial strength. In short, this was achieved by simulating the reinforcement and experimental extinction envisaged by Hull with an automatic device that increased or reduced the strength of the respective magnets as goal was achieved after the pendulum had come into contact with each of them.

Rather than go into greater detail, it will be more enlightening to look at Ellson's table giving the changes in performance (and time taken by performance) shown by the model in successive trial-and-error learning cycles (Fig. 4). Clearly, Ellson's explicit aim was to achieve detailed simulation of the changes shown in the table in which Hull had, a few years earlier, outlined the trial-and-error performance of a hypothetical organism (Hull, 1930a). From the two tables it is possible to see how the machine and the organism "learn" to respond with the reaction R_z, is seen to be progressively reinforced during the successive Problem (or Behavior) cycles. The machine reproduces various quantitative aspects of conditioning (experimental extinction, spontaneous recovery, etc.) predicted by Hull's theoretical assumptions. This is possibly the first case of detailed matching between intelligent performance in an organism and in an artificial model: a key-point of the future methodology of the artificial in psychological research.

The building of inorganic mechanisms simulating simple trial-and-error learning could be considered a success for Hull's theory: "one more step", as Ellson put it while reviewing the recent history of the mechanistic approach in psychology since Stephens (Ellson, 1935, p. 216). Nevertheless, on looking back to the 1930 paper mentioned above where Hull analyzes the features of simple trial-and-error learning, we got the impression that he would not have attached excessively great significance to such mechanisms. Here he came to the conclusion that although psychologists observing an animal learning the correct answer through simple trial-and-error might judge this "purposive" behavior, it could by no means be described as the "type of purpose involving a plan". The construction of inorganic mechanisms underlying this type of purposive behavior and other more complex, "truly psychic" phenomena like planning and insight, pointed to "a radically new order of automaticity" (Hull, 1930a, p.256), although in principle by no means impossible to construct. The theoretical and experimental research Hull carried out between 1930 and 1935 aimed at defining the terms within which such a set of problems could be tackled.[11] Here I shall limit reference to the theoretical points relating directly to this alleged "radically new" automaticity, as a candidate explanation of "truly psychic" behavior. One of these points touches on the particularly representational theory of mind outlined by Hull.

Trial No.	Problem cycle I — Status of excitatory tendencies preceding reaction	Resulting response	Problem cycle II — Status of excitatory tendencies preceding reaction	Resulting response	Problem cycle III — Status of excitatory tendencies preceding reaction	Resulting response	Problem cycle IV — Status of excitatory tendencies preceding reaction	Resulting response
1	$x = 100^*$ $y = 70$ $z = 30'$	R_x	$x = 70^*$ $y = 50$ $z = 40'$	R_x	$x = 50^*$ $y = 40$ $z = 50'$	R_x	$x = 40^*$ $y = 40$ $z = 60'$	R_z
2	$x = 70^*$ $y = 70$ $z = 30'$	R_x	$x = 40^*$ $y = 50$ $z = 40'$	R_y	$x = 20^*$ $y = 40$ $z = 50'$	R_z		
3	$x = 40^*$ $y = 70$ $z = 30'$	R_y	$x = 40^*$ $y = 20$ $z = 40'$	R_x				
4	$x = 40^*$ $y = 40$ $z = 30'$	R_x	$x = 10^*$ $y = 20$ $z = 40'$	R_z				
5	$x = 10^*$ $y = 40$ $z = 30'$	R_y						
6	$x = 10^*$ $y = 10$ $z = 30'$	R_z						

Trial No.	Behavior cycle I — Status of excitatory tendencies preceding reaction	Resulting reaction	Behavior cycle II — Status of excitatory tendencies preceding reaction	Resulting reaction	Behavior cycle III — Status of excitatory tendencies preceding reaction	Resulting reaction	Behavior cycle IV — Status of excitatory tendencies preceding reaction	Resulting reaction
1	$x = 3.0$ $y = 2.0$ $z = 1.0$	R_x	$x = 2.2$ $y = 1.5$ $z = 1.3$	R_x	$x = 1.7$ $y = 1.2$ $z = 1.6$	R_x	$x = 1.4$ $y = 1.2$ $z = 1.9$	R_z
2	$x = 2.1$ $y = 2.0$ $z = 1.0$	R_x	$x = 1.3$ $y = 1.5$ $z = 1.3$	R_y	$x = .8$ $y = 1.2$ $z = 1.6$	R_z		
3	$x = 1.2$ $y = 2.0$ $z = 1.0$	R_y	$x = 1.6$ $y = .6$ $z = 1.3$	R_x				
4	$x = 1.5$ $y = 1.1$ $z = 1.0$	R_x	$x = .7$ $y = .9$ $z = 1.3$	R_z				
5	$x = .6$ $y = 1.4$ $z = 1.1$	R_y						
6	$x = .9$ $y = .5$ $z = 1.0$	R_z						

Fig. 4. (From Ellson (1935) and Hull (1930a).)

Ultra-Automaticity and the "Psychic Machine"

A redintegrative organism has one basic feature: that of acquiring the capacity to give rise to the response sequence even indepentently of a corresponding sequence of stimuli coming from the external world. In order to do this the organism must create its own sort of "subjective parallel" to the external world, as Hull described it (1930b). When the organism interacts with the world it is as if "the world in a very important sense has stamped the pattern of its action upon a physical object", i.e. the organism. Thus the organism is able to reproduce the pattern even when the corresponding situation is absent:

"The imprint has been made in such a way that a functional parallel of this action segment of the physical world has become a part of the organism. Henceforth the organism will carry about continuously a kind of replica of this world segment. In this very intimate and biologically significant sense the organism may be said to know the world. No spiritual or supernatural forces need be assumed to understand the acquisition of this knowledge. The process is an entirely naturalistic one throughout." (Hull, 1930b, p. 514)

In particular, with this type of "replica" or "pattern" the organism can foresee and anticipate events in the external world. These replicas may be considered the knowledge the organism exploits to implement its own subjective sequence of actions before the world sequence constituting the source of external stimuli occurs in real time. Thus the end-response in the subjective sequence is characterized as an action that, once conditioned, can anticipate the external stimulus that had been indispensable to elicit the action before conditioning took place. This end-response is the only one to retain an instrumental function, the other actions or responses in the subjective sequence acquiring at this point exclusively symbolic value: in other words, they become responses that serve exclusively to trigger other actions in the sequence by means of the (internal) proprioceptive stimuli they excite in the organism. As Hull put it, they are "pure stimulus acts" or "symbolic" acts. Since they have ceased to be instrumental acts, they can conveniently be eliminated from the sequence by means of a mechanism Hull defines as "short-circuiting" (think, for example, of the anticipatory short-circuiting in an animal's sequence of actions defending it from a noxious stimulus, once the sequence has been learnt by the animal).

Thus, as Hull concluded, "the organism is no longer a passive reactor to stimuli from without, but becomes relatively free and dynamic". The role of the symbolic mechanism appears, in fact, to be the "transformation of mere act into thought", to quote Hull again: in other words, its role is to draw from the purely instrumental sequences the symbolic aspect of a thought that may, in some cases, be "rudimentary", but that already has the capacity for anticipatory behavior (Hull, 1930b, pp. 516–517). The fact that this organic mechanism is essentially *automatic* challenges us to attempt the customary synthetic verification from inorganic materials":

"It is altogether probable that a "psychic" machine, with ample provision in its design for the evolution of pure stimulus acts, could attain a degree of freedom, spontaneity, and power to dominate its environment, inconceivable alike to individuals unfamiliar with the possibilities of automatic mechanisms and to the professional designers of ordinary rigid-type machines". (Hull, 1930b, p. 517)

We might describe Hull's "psychic machine" as a machine whose behavior is not the simple effect of a simple stimulus from without: the cause of its action is

internal and can be accounted for with its symbolic model of reality. Ultimately, in order to conceive of a machine able to "dominate its environment" we should be prepared to imagine a machine possessed of that type of "subjective parallel" to the world possessed by organisms. Hull sees the building of such a machine as a real bet with the future. As usual, success achieved in the simulation of the simplest (although also most fundamental) principles of automatic functioning seemed to hold out hopes of similar success in the simulation of principles explaining the most complex, "truly psychic" behavior of organisms. *In the meantime*, as it were, this position of Hull's appears to have been established on the basis of his empirical, experimental psychological research, which seems to confirm in his mind the equally *automatic* nature of the more complex principles underlying reasoning, planning and insight.[12] Indeed, a few years after he had formulated his "psychic machine" Hull accounted for insight with a sophisticated version of the principles of associationism, concluding (in disagreement with mentalists and Gestaltists) with these words:

"To state the same thing in other words, we appear to have before us here a deduction of insight in terms such that it might conceivably be constructed by a clever engineer as a non-living – even an inorganic – mechanism". (Hull, 1935, p. 268)

In another context Hull's definition of the effective features of this automatic *but* "psychic" machine is less generic but expressed in terms of negation: the *"ultra-automaticity"* (as he defined it) of the machine is *not* that of recent automatic machines known for their practical applications, such as the steam engine or the electric motor.[13] However (and with some relevance to the field of the culture of the artificial I feel), Hull hazarded the prediction that developments in this no better defined "ultra-automaticity" might also depend on its future practical applications:

"It is not inconceivable . . . that in the demands for higher and higher degree of automaticity in machines constantly being made by modern industry, the ultra-automaticity of the type of mechanism here considered may have an important place. In that event the exploration of the potentialities lying in this radically new order of automaticity would be comparatively rapid." (Krueger and Hull, 1931, p. 268)

Nevertheless, there is no real case for far-fetched speculation on the premonitions implicit in the evocative definition of an "ultra-automatic" psychic machine ("non-rigid", possessed of a "degree of freedom", able to "dominate its environment") cited above and in its possible industrial fall-out. As pointed out above, all these expressions refer to the (automatic) response anticipatory mechanism, which is in turn based on the divergent excitatory mechanism typical of a redintegrative organism.

In its simplest form, the response anticipatory mechanism should be implicitly present in simple trial-and-error learning (and thus in its inorganic simulation). Here a "short-circuiting" occurs in the alternative response sequence resulting from a stimulus like S_D (internal and non-variable, as we have seen). As a result of this short-circuiting, one or more actions in the behavior sequence are, as it were, *skipped* – possibly even all but one of them, as in the case of simple trial-and-error mentioned above where, after a certain period of learning through

random trials, the organism produces response R_3 immediately and exclusively. In this case it is the only response having instrumental value in the achievement of success.

As we have seen, although this type of "short-circuiting" is based on forms of anticipatory behavior (and thus on a certain "subjective parallel" of the world), it would have been considered by Hull as a very "rudimentary" form of thought. Underlying the "truly psychic" behavior shown by organisms and thus, presumably, that no better defined "radically new" automaticity or "ultra-automaticity" of future machines, there must apparently have been a *second* divergent excitatory mechanism, also grounded in the organism's redintegrative capacities (the former, as we have seen, is the mechanism that triggers off short-circuiting in symbolic action sequences through S_D). In fact this second divergent mechanism was to be thought of as physically simulated in terms of notions that, as we shall see, went beyond Hull's robot approach. Anyway, it was taken to be triggered by an (internal) proprioceptive stimulus, which became the object of great consideration both among Hull's critics and successors (e.g. Mowrer and Osgood). Here I am referring to the most exemplary form of "mediational" constructs (between stimulus and response) known as the "anticipatory goal response". Without necessarily short-circuiting the symbolic action sequences (Hull, 1931, p.290), this is an internal preparatory response of the organism to a need stimulus, while at the same time it is a stimulus controlling the sequence of instrumental actions that produce the plan to eliminate the need itself.[14] Indeed, this was the authentic, dynamic mechanism that, as Hull later put it, made the organism "hormic and purposive" (Hull, 1942, p. 90). Suffice it here to recall that Hull saw in this the physical symbolic mechanism that could account for the old paradox of *backward* causation typical of intentional anticipatory actions – another anti-mechanistic war horse of Hull's time (actually of the time up to cybernetics, as we shall see).

All this brings us back to one of the key points of the culture of the artificial, as I defined it in the first section: the role of representational (physical) mechanisms in the explanation of knowledge and intentional action. However, in order to avoid anachronistic confusion over the nature of Hull's physical symbolic mechanism, we must at least distinguish between representations with a causal role, i.e. Hull's, and representations as contents of physical symbol structures. In the latter case, the causal role is *not* attributed to the representations as such, but to the physical structures (as symbolic codes), representations being their semantic content. For computationalist cognitivists, e.g. Pylyshyn, these physical symbolic structures are the key to their *different* representational solution to the *same* problem tackled by Hull: a naturalistic account of the semantic of the functional states (as explicitly recognized by Pylyshyn, 1984, pp. 24–28 and 41–42.[15]

As might have been expected, inorganic reproduction of the goal anticipatory response occupied an important position on Hull's list (mentioned above) of problems to be tackled in future research on the inorganic simulation of "truly psychic" behavior (Krueger and Hull, 1931). Another problem he tackled was the design of a device that might be "conditioned" not simply to non-organized complexes of stimuli, but also to certain particular combinations of stimuli in a

given complex: in other words, a machine capable of distinguishing between different patterns of stimuli.

An attempt to solve this problem was made the very same year by Nicolas Rashevsky, a biophysicist who was very interested in the physicalistic implications of Hull's robot approach. He interpreted learning as a property of physical systems, which could in essence be referred to specific forms of hysteresis (his research was carried out in the general field of the thermodynamics of living systems, which Lotka had already developed in that period: see Rashevsky, 1931a). Rashevsky's hypothetical machine for the recognition of stimuli patterns is shown in Fig. 5 (Rashevsky, 1931b). The machine is interesting as an "assembly" of basic units corresponding to elementary circuits of the Baernstein-Hull type. His machine consists of nine of these units (each represented by an S in the figure). The bulb (as usual representing the response: e in the figure) can be lit either directly by a_u (unconditioned stimulus) or by a_c (conditioned stimulus) through the conditioned process previously described (as we have seen, consisting in closing a_c immediately before a_u). In this case, however, once the entire system has been conditioned a_c is then linked to e via three different basic elements linked in series and showing one of the possible configurations. *Which* configuration is produced depends on the previous shifts in the dials of three manometers, M_1, M_2 and M_3, each of which can activate one of the three basic elements and thus form a particular circuit. Shifts in the manometer dials depend in turn on the particular wave pressure passing through pipe P, which is full of air, once vibration has been produced in the microphone membrane, MF. The figure shows a closed circuit between a_c and e via 1', S'_1, 3", S''_3, 2''', S'''_2. A different wave pressure will activate three different basic units linked up in a series that does not complete the circuit and does not trigger a response (the lighting of e). In
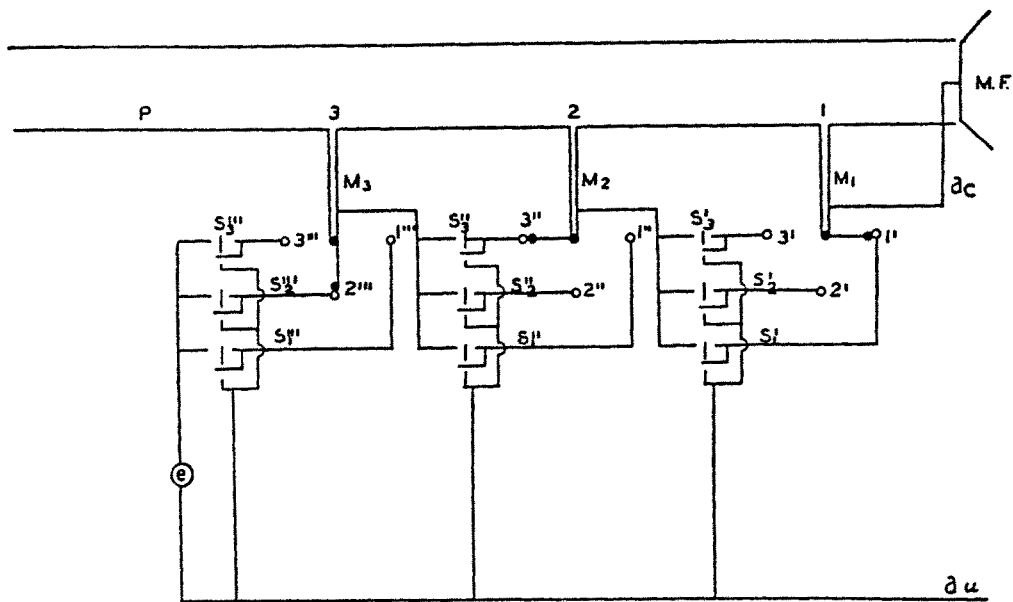

Fig. 5. (From Rashevsky 1931b; corrected).

other words, the machine discriminates the stimulus pattern it has been conditioned to and responds to that and no other pattern.

Rashevsky (in disagreement with Driesch) held that it was not in principle impossible to construct more complex machines that might even be able to learn a natural language although, he pointed out, it would require "extraordinary expense and work". He also pointed out that descriptions "in psychological terms" of systems that "know", "desire", "infer" and "try" are "justified by the circumstance that the above process of trials occurred on a different level from the ordinary muscular responses". Moreover, since these machines show behavior patterns of a certain complexity and unpredictability, the observer interacting from without can only use this language, having no apriori knowledge of the "purpose" for which a given system acts (Rashevsky, 1931b, pp. 393–403). This, too, is a (much debated) key point of the culture of the artificial as I have defined it: the legitimacy of a mentalistic language in interacting with "intentional (natural or artificial) systems".[16]

Analytic and Synthetic Methods: Towards the Cybernetics Era

Hull's methodology of the artificial has undergone further refinements through various other contributions taking their lead from his robot approach. The expert in electrical engineering Thomas Ross, for example, designed a number of automata which were subsequently to be mentioned by Ashby and Grey Walter, who looked on Ross as his precursor in the construction of synthetic animals (Walter, 1953, p.82). Here I should like to spend a few words on what was possibly the founding member of this family of machines. Completed in 1935 thanks to collaboration between Ross and the psychologist Stevenson Smith, it was a motor-run device running on three wheels along a maze consisting of twelve "Y"s, each of which had one "blind" turn (to the right). When first placed at the start of the maze, the automaton explored it thoroughly entering every blind path and "retracing its steps" before going on towards the goal. When subsequently placed once again at the start of the maze, the automaton went through avoiding all the blind paths (like its better-known descendants in the cybernetics era). The crude mechanism underlying this automatic learning process can be seen in Fig. 6. Suffice it to point out that it consists of a rotating disk on which unsuccessful (blind paths) and successful attempts during exploration of the maze are recorded with a pattern of twelve depressed or raised tabs, corresponding respectively to the choices of turning right or left. By the time the automaton had gone through the twelve sections of the maze the disk has performed one complete rotation; on the second run the automaton finds the correct route imprinted in its memory, as it were, and will make no further errors. Ross referred somewhat polemically to what I have mentioned as the vitalist's in principle objection concerning the impossibility for a machine to learn: "*That* is intelligent behavior – a machine could never do that." *His* machine could (Ross, 1938, p. 187).

Of course, precisely the "non-animal" perfection, as it were, with which Ross' automaton demonstrated its learning capacities might well have fuelled arguments for further objections, if not in principle at least by "redefinition", as
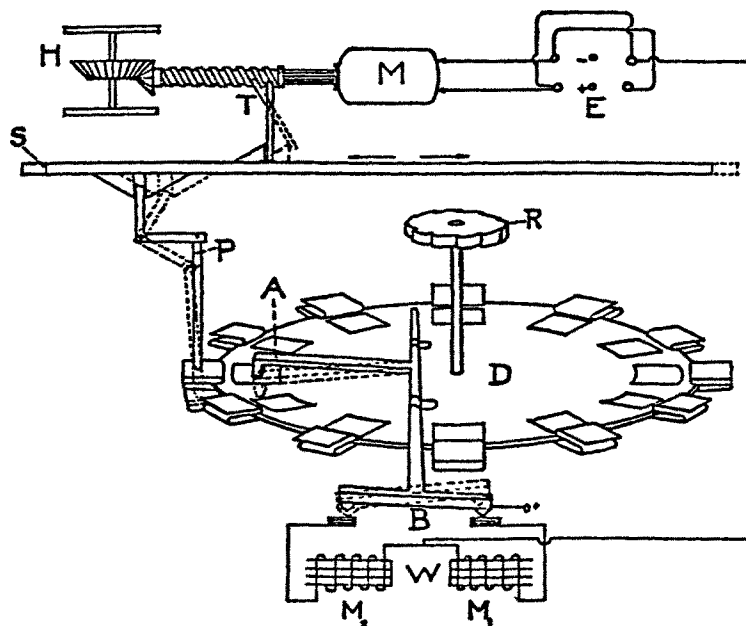
Fig. 6. (From Ross, 1938).

Minsky was to call it, i.e. by continually modifying the definition of intelligence in order to exclude all artificially reproduced phenomena.[17] However, apart from the antivitalistic dispute, Ross defined with extreme clarity the implications of the synthetic method which were to became *loci classici* in the culture of the artificial: the *testing* of psychological hypotheses by constructing artefacts, and the machine–organism *comparison* for respective performances. Presenting the project of some machines analogous to Rashevsky's (i.e. with the ability to discriminate stimulus patterns), Ross wrote:

"It is hoped that it may be possible to test the various psychological hypotheses as to the nature of thought by constructing machines in accord with the principles that these hypotheses involve and comparing the behavior of the machines with that of intelligent creatures. Clearly, this synthetic method is not intended to give any indication as to the nature of the mechanical structures or physical functions of the brain itself, but only to determine as closely as may be the type of function that may take place between "stimulus" and "response" as observed in the psychological laboratory or in ordinary uncontrolled learning and thinking. Only analogies which will work when elaborately executed are sought, not imitations of nerve, brain and muscle structure". (Ross, 1935, p. 387)

Shortly afterwards, Hull himself again referred to the synthetic experiments on which these "hopes" rested. In this case the anti-vitalistic conception of his psychic machine seemed to fit into a perfectly explicit naturalistic reductionism. In Hull's opinion, given the existing state of knowledge, it was impossible to realize such a reductionist project with regard to the psychology-physics relationship, i.e. it was impossible to deduce the postulates which are basic to adaptive and purposive behavior as theorems of microphysics, according to the "natural goal of science". Thus the idea of a "complete scientific monism" should be considered "only as a working hypothesis". In the meantime, Hull concludes, "a

kind of experimental shortcut" was to be preferred to the problematic point of view of the monistic-physicalistic theses – models had to be designed that would satisfy the postulates of Hullian theory. If such models behaved exactly as predicted by these axioms, *the theory would be proved true*, and it could be concluded that the "adaptive behavior may be 'reached' by purely physical means" (Hull, 1937, p. 29).

Hull's "shortcut" seemed to be clearly stated by Ross. In describing his automaton he focused his attention on two key-points of the culture of the artificial. According to Ross, in the expectation of the "hoped-for ultimate success of physiologists" in explaining cerebral structures, psychologists could consider the possibility of constructing mechanisms which are *functionally equivalent* to intelligent behavior. Such mechanisms could thus provide *proofs of sufficiency* for their theories:

"One way to be relatively sure of understanding a mechanism is to make that mechanism. To find the sufficient condition for learning we should try to make a machine that will learn. This has been done several times in the past few years, and has perhaps been somewhat overdramatized. A very persistent mistaken attitude to work of this sort is the idea that the builder of a machine which will learn must think he has built a mechanism physically like that underlying human or animal learning. Nothing could be further from the truth. What is demonstrated by the physical existence of a performing machine is that a machine is capable of that kind of performance. It is not demonstrated, however, that only this sort of machine can produce the given effects; for no truth is more commonplace in mechanics than that, in general, several alternative mechanisms, differing widely in superficial characteristics and forms of energy utilized, can produce the same end result". (Ross, 1938, p. 185)

Ross seemed, therefore, to make completely explicit the methodological implications of the culture of the artificial of his time. The temptation is to suggest that, in a very restricted sense, Ross' automaton fulfilled Hull's dictum on the central role played by the internal model in anticipatory behavior. In fact the memory-disk performs the (arbitrary) symbolic notation function which the automaton uses to record its position in the world. The blind path is represented by an unequivocal tab position: from then on the automaton will avoid it on the basis of this internal representation. Thus a certain (successful) action is anticipated by means of an automatic mnemonic recall process.

If we accept this interpretation, we might call Ross' automaton a "Craikian automaton". Johnson-Laird (1983, pp. 403–404) used this expression to describe, among others, simple mechanisms that appeared to satisfy the requisites formulated by Kenneth Craik in his famous hypothesis on the knowledge of the world obtained by organisms *and mechanisms* through a "model or parallel of reality", as he put it (Craik, 1943). Actually, however, all this should be taken with a considerable dose of caution (indeed it is very significant of the confusion that can be created in the field of the culture of the artificial). For example, without closer definition the "subjective parallel" or model of the world hypothesized by Hull could be reduced to the far weaker notion of a Watson-type (internal) proprioceptive stimulation pattern: a pattern capable of recalling without "cues" from the outside a certain sequence of action (a "chain reflex", in Watson's expression) previously learnt through exteroceptive cues.[18] Moreover, Ross' automaton (and actually also the simple so-called Craikian automata exemplified by Johnson-Laird) may be said to learn through a kind of feedback control, but only in a weak sense, i.e. in the very sense of the Law of the Effect

criticized by Mowrer – that of "learning by doing". As Mowrer sharply stressed, the cybernetic insight of automatic feedback control is much deeper than this.[19]

Craik's hypothesis is very well-known, but his appraisal of Hull's robot approach is less known.[20] Despite Craik's (deserved) success among the various computationalist cognitivists – from Johnson-Laird to Minsky, from Miller *et al.* to Newell and Simon – this theoretical link with Hull may help us towards a better understanding of Craik's ideas and, at the same time, cast some light on those who were to take up certain of Hull's vague as well as "prophetic" formulations concerning the ultra-automaticity of the psychic machine. In fact, Craik seems to stand at the convergence of trends in mechanistic psychology (the "non-rigid" machine) with the then very recent developments in automatic control theories (machines with feedback). As Craik put it, the intention of pursuing the robot approach "even more unswervingly" (Craik, 1966, p.80) than his inspirer and, in contrast, his criticism of certain trivial applications of the robot approach, underlie his distinction between the "analytic method" (essentially neurophysiological) and the "synthetic method" (essentially that of artificial models) in mind studies. However, Hull's arguments against "individuals unfamiliar with the possibilities of automatic mechanisms and (. . .) the professional designers of the ordinary rigid-type machines" (Hull, 1930b, p. 517) seem to find a natural echo in Craik's arguments against those who reject the synthetic method on the basis of the idea of a "rigid" machine, i.e. "lacking in feedback", as Craik put it to stress his disagreement with McDougall (Craik, 1966, pp. 19–20). And note that Craik, evoking such feedback machines, intended to give his own physicalistic solution to McDougall's old paradox of "causation from in front" – a paradox, as we saw, that Hull tried to solve through the postulation of the anticipatory goal physical mechanism.

In short, the fateful word had been uttered: the year 1943 saw anticipatory, "purposive" or "intelligent", behavior functionally identified with the behavior of a negative feedback machine. Craik's argument was on the way to becoming a *locus classicus*. "What would McDougall have said to Ashby's homeostat?", wondered Sluckin (1954, p. 36) returning to the problem in the fifties. But this is recent history.

Acknowledgements

Notes

1.   Note that I am not giving here a definition of cognitive AI as a scientific field (indeed a more difficult problem), but a statement of the background of philosophical, methodological and pragmatical assumptions fairly widespread among cognitive AI scientists – just a statement of culture of the artificial.

2. In the present paper I shall use the term *connectionism* to refer exclusively to that of the pre-Hebb kind. Hebb was inspired by Pavlov, Thorndike and Hull's connectionism, but his criticism of them as well as his new approach to connectionism and associationism cannot be disregarded. I stressed this point in an analysis of some pre-Hebb mechanical simulations of neurological processes (see Cordeschi, to be published).

3. In fact Hull used the expression "robot approach" only afterwards, in his *Principles of Behavior* (see Hull, 1943, pp. 27–28: this passage was quoted in their well known book by Miller, Galanter and Pribram, 1960). Nevertheless, such an expression can be legitimately used to denote his "synthetic" (mechanistic and antivitalistic) approach to the study of mind during the early thirties.

4. The description of the mechanism for the automatic finding of syllogistic conclusions (which is mentioned also by Martin Gardner, 1982, p. 124n) was not published. Hull hinted at it occasionally (for example, Hull, 1935, p. 219). That Hull's next interest for a more complex "intelligent" machine dates back to the mid-twenties results from Hull's posthumously published *Idea books* (see Hull, 1962). On the contrary, from the published pages of such a notebook, as one of its editors recognizes, it is not possible to know why, around the mid-thirties, Hull gave up his plans for constructing such machines. Only occasional explicit references to them are found in his next works (for example, Hull, 1937). And yet at first Hull seemed to take his leadership too seriously in this field. He wrote for example in 1930: "Fortunately my years of study and meditation on the subject will probably enable me to keep well ahead of the pack and be their leader and spokeman. My academic position as well as my age should aid in this." And he seemed aware of the possibility of criticism to his project. So he wrote in 1929: "I shall doubtless encounter sneers, criticism, and possibly even opposition from the conservative and unimaginative individuals. But if so, this can hardly be more than an incident. I must surround myself with a few sympathetic and intelligent individuals so as to protect myself from the negative suggestion associated with harsh criticism . . ." (Hull, 1962, pp. 838 and 829). On the other hand, his projects became quite popular among the psychologists as well as the media of his time. Anyway, it is a fact that Hull's robot approach did not go beyond simple models of conditioning, notwithstanding his deep insight in the methodology of the artificial – an insight which was to be better realized within a more developed thecnology.

5. The description of another mechanical device which simulated simple conditioning was published in 1930 (Walton, 1930), but more "for purposes of demonstration in introductory psychology", as its author said, than for testing a psychological (or neurological) hypothesis.

6. For example, Hull wrote in 1927 in his *Idea books:* "In designing a learning machine I felt the necessity of providing a device to vary the reactions so that trial and error could take place effectively" (Hull, 1962, p. 823).

7. See, for example, McDougall (1923).

8. Hull borrowed the term *redintegration* from the psychologist H. L. Hollingworth (Hull, 1929, p. 498). In Hull's opinion redintegration gave a more comprehensive view of old associationistic principles (for example, the simple chain reflex law might be derived from it: see Hull, 1930b, pp. 522–523). In a slightly different sense the term had been used by James, to denote the fact that one element of a past experience may, through association, awake the entire content of that experience (James, 1890, vol. 1, pp. 569 ff.). In the latter sense the old term became evocative among neoconnectionists (see Anderson and Rosenfeld, 1988, p. 3), who consider James as a forerunner of their own computational networks models of association. Of course, Hull too many be included among the authors who influenced the evolution of recent mechanistic models of association (just to make an example, think to S. Grossberg).

9. Hint: press S_u and, simultaneously, the conditioned stimulus complex S_1, S_2. Interrupt when "conditioning" is strong enough to evoke the maximal luminosity of L. If you press either stimulus button of the complex, both call forth the luminosity of L: though such a luminosity is obviously weaker than that evoked by the entire stimulus complex, it is the *same* in both cases. This does not occur in Baernstein and Hull's circuit (Fig. 1b), given the presence of the resistance, which on the other hand is necessary if you want to reproduce irradiation in it. The tendency to irradiation in Krueger and Hull's circuit (Fig. 2) turns out from the fact that, when S_1 is pressed in conjunction with S_u, the current from battery E tends to flow not only through S_1 to the storage cell E_1, but (for example) also to E_2 through the resistence R_1. Essentially, the latter, compared to the former is a fully analogical machine.

10. But note that Hull felt it necessary to correct "a frequent misunderstanding" on this point, which he ascribed to "the wide dissemination of the view of J. B. Watson": he never assumed that "the more complex forms of behavior are synthetized from reflexes which play the role of building blocks. This may or may not be true. His working hypothesis is, rather, that the *principles of action* discovered in conditioned reaction experiments are also operative in the higher behavioral processes" (Hull, 1935, p. 228 n.).

11. I refer here particularly to Hull (1929; 1930a; 1930b; 1931; 1932; 1934; 1935).

12. "Organism is here conceived as a completely automatic entity", Hull shall write afterwards, adding that the "complex automaticity" of human adaptive and learned behavior "is quite as automatic and self-regulating as simple automaticity". In any case, "highly complex automaticity constitutes no more evidence concerning the existence of an entelechy, or reason for indulging in anthropomorphism, than does simple automaticity" (Hull, 1952, p.347).
13. Through the introduction of such an "utterly new and different order of automaticity", Hull wrote in 1929, "mechanical engineering of automatic machines will be revolutionized to a degree similar to the introduction of steam engines and electricity" (Hull, 1962, p. 829).
14. In Hull's opinion, the mechanism of short-circuiting behavior sequences *per se* is not adequate to explain more complex trial-and-error behaviors, as maze learning. More radical criticisms were raised by Hull against the application of certain over-simplified principles of the early Watsonian type, such as the chain reaction principle, in the explanation of maze learning (see for example Hull, 1932).
15. See Pylyshyn's book for a discussion of other trends in cognitive science (e.g. Stich) with regard to the nature of representations.
16. Perhaps the most important treatment of the notion of ascription of purposiveness to artefacts preceding the well known Daniel Dennett's notion of intentional system is due, during the cybernetic era, to Donald McKay. He fully developed also a deep notion of mental representation, much appreciated (together Craik's symbolic model, which I shall expand on later) among early cognitivists (see Miller *et al.*, 1960, pp. 50–51). I stressed both points in Cordeschi (1985, 1988).
17. See Minsky (1963, p. 396). This objection is perhaps more popular in Douglas Hofstadter's version of the so called Tesler's theorem: "Artificial intelligence is whatever hasn't been done yet".
18. In spite of Ross' reference to Hull's psychic machine, this seems to be his own interpretation of learning and memory processes as based on kinesthetic stimuli (Ross, 1933, p. 208; 1938, p. 189). On the other hand, some hints on Hull's anti-Watsonian stance were given above.
19. In the case of learning by doing the automatic control concerns "not a response as it occurs, but rather a response or behavior of the organism in the same (or a similar) situation *on some future occasion*" (Mowrer, 1960, p. 267; but see the whole Chap. 7).
20. See Craik's *The mechanism of human action* (1943), which was edited posthumously by S. L. Sherwood, with an Introduction by W. S. McCulloch, L. Verbeek and S. L. Sherwood (Craik, 1966). (Craik died in 1945 at the age of 31.) Further details on Craik's mechanistic conception of mind may be found in Cordeschi (1986).

References

Anderson, J. A. and Rosenfeld, E. (eds.). (1988). *Neurocomputing*. The MIT Press.
Baernstein, H. D. and Hull, C. L. (1931). A Mechanical Model of the Conditioned Reflex. *Journal of General Psychology*, 5. 99–106.
Bennett, G. K. and Ward, L. B. (1933). A Model of the Synthesis of Conditioned Reflexes. *American Journal of Psychology*, 45, 339–342.
Cordeschi, R. (1985). Mechanical Models in Psychology in the 1950s. In S. Bem, H. Rappard and W. van Horn (eds) *Studies in the History of Psychology and the Social Sciences*. Psychologisch Instituut of Leiden.
Cordeschi, R. (1986). Kenneth Craik and the "Mechanistic Tendency of Modern Psychology". *Rivista di Storia della Scienza*. 3. 237–256.
Cordeschi, R. (1988). *Intentional Psychology and Computational Models*. *Conceptus-Studien*. 5. 69–77.
Cordeschi, R. (to be published). *Two Proto-connectionistic Machines*.
Craik, K. J. W. (1943). *The Nature of Explanation*. Cambridge University Press.
Craik, K. J. W. (1966). *The Nature of Psychology*. Cambridge University Press.
Ellson, D. G. (1935). A Mechanical Synthesis of Trial-and-Error Learning. *Journal of General Psychology*, 13. 212–218.
Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and Cognitive Architecture: a Critical Analysis. *Cognition*, 28. 3–71.
Gardner, M. (1982). *Logic Machines and Diagrams*. Harvester.
Hull, C. L. (1929). A Functional Interpretation of the Conditioned Reflex. *Psychological Review*. 36. 498–511.
Hull, C. L. (1930a). Simple Trial-and-Error Learning: a Study in Psicological Theory. *Psychological Review*. 37. 241–256.
Hull, C. L. (1930b). Knowledge and Purpose as Habit Mechanisms. *Psychological Review*. 37. 511–525.

Hull, C. L. (1931). Goal Attraction and Directing Ideas Conceived as Habit Phenomena. *Psychological Review*, 38. 487–506.

Hull, C. L. (1932). The Goal Gradient Hypothesis and Maze Learning. *Psychological Review*, 39. 25–43.

Hull, C. L. (1934). The Concept of Habit-Family Hierarchy and Maze Learning. *Psychological Review*, 41. 33–52 (Part I); 134–152 (Part II).

Hull, C. L. (1935). The Mechanism of the Assembly of Behavior Segments in Novel Combinations Suitable for Problem Solution. *Psychological Review*. 42. 219–245.

Hull, C. L. (1937). Mind, Mechanism, and Adaptive Behavior. *Psychological Review*. 44. 1–3.

Hull, C. L. (1942). Conditioning: Outline of a Systematic Theory of Learning. In *The Psychology of Learning*. Public School Publishing Co. of Bloomington.

Hull, C. L. (1943). *The Principles of Behavior*. Appleton.

Hull, C. L. (1952). *A Behavior System*. Yale University Press.

Hull, C. L. (1962). Psychology of the Scientist: IV. Passages from the "Idea Books' of Clark L. Hull". *Perceptual and Motor Skills*. 15. 807–882.

Hull, C. L. and Baernstein, H. D. (1929). A Mechanical Parallel to the Conditioned Reflex. *Science*. 70 (1801). 14–15.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge University Press.

James, W. (1890). *The Principles of Psychology*. Holt and Co. (Dover Edition, 2 vols.).

Krueger, R. G. and Hull, C. L. (1931). An Electro-Chemical Parallel to the Conditioned Reflex. *Journal of General Psychology*. 5. 262–269.

McDougall, W. (1923). *Outline of Psychology*. Methuen.

Miller, G. A., Galanter, E. and Pribram, K. H. (1960). *Plans and the Structure of Behavior*. Holt, Rinehart and Winston.

Minsky, M. (1963). Steps Toward Artificial Intelligence. In E. A. Feigenbaum and J. Feldman (eds.) *Computers and Thought*. McGraw-Hill.

Mowrer, O. H. (1960). *Learning Theory and the Symbolic Processes*. Wiley and Sons.

Pavlov, I. P. (1927). *Conditioned reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex* (English translation). Oxford University Press.

Pylyshyn, Z. W. (1984). *Computation and Cognition. Toward a Foundation for Cognitive Science*. The MIT Press.

Pylyshyn, Z. W. (1989). Why Foundations, and Where Next? *Artificial Intelligence*. 39. 248–251.

Rashevsky, N. (1931a). Learning as a Property of Physical Systems. *Journal of General Psychology*. 5. 207–229.

Rashevsky, N. (1931b). Possible Brain Mechanisms and their Physical Models. *Journal of General Psychology*. 5. 368–406.

Ross, T. (1933). Machines that Think. *Scientific American*. 148. 206–208.

Ross, T. (1935). Machines that Think. A Further Statement. *Psychological Review*, 42. 387–393.

Ross, T. (1938). The Synthesis of Intelligence. Its Implications. *Psychological Review*. 45. 185–189.

Sluckin, W. (1954). *Minds and Machines*. Penguin Books.

Stephens, J. M. (1929). A Mechanical Explanation of the Law of Effect. *American Journal of Psychology*. 41. 422–431.

Walter, W. Grey (1953). *The Living Brain*. Duckworth.

Walton, A. (1930). Conditioned Illustrated by an Automatic Mechanical Device. *American Journal of Psychology*. 42. 110–111.

Correspondence and offprint requests to: Roberto Cordeschi, Istituto di Filosofia, Università di Salerno, I-84084 Fisciano (SA), Italy.