A CYBERNETIC VIEW OF ARTIFICIAL INTELLIGENCE

José Mira and Ana E. Delgado

Received March 15, 2006

Abstract. In this work, taking advantage of the fiftieth anniversary of Artificial Intelligence (AI), we consider the disparity that exist between the excessive initial objectives of synthesizing general intelligence in machines and the modest results obtained after half a century of work. We mention some of the possible causes of this disparity (constitutive differences between man and machines, lack of a theory of computable knowledge and oblivion of the solid work made by the pioneers of Cybernetics). Then we go over the history until we arrive to the current AI paradigms (symbolic, situated and connectionist) in search of its cybernetics roots and conclude with some suggestions on the strategic decisions that could be adopted to progress in the understanding of intelligence.

1 Introduction In "Recolections from many sources of Cybernetics" [25] W.S. McCulloch wrote: "That is where cybernetics is today (1974). It was born in 1943, christened in 1948, and came of age five years ago in the early 1960's. In its short majority is has certainly done best for those fields where it was conceived. It has been a challenge to logic and to mathematics, an inspiration to neurophysiology and to the theory of automata, including artificial intelligence, and bionics or robotology. To the social sciences it is still mere suspiration."

Unfortunately, when we are celebrating the fiftieth anniversary of AI, we have to recognize that the most genuine initial motivations of cybernetics has been lost and that there is a great disparity between the initial objectives of synthesizing general intelligence in machines and the modest results obtained after half a century of work.

Among the possible causes of this disparity obviously the first is the enormous nature of the task (synthesizing general intelligence in machines), since we do not know the physiology and cognitive processes and hence are faced with the problem of reproducing something that we do not know. Natural intelligence is a very wide concept that encompasses a large number of skills other than the mere solution of scientific-technical problems in narrow domains.

The second cause of the disparity is associated with the large constituent differences between biological systems and machines. If all knowledge depends on the constituent materials and mechanisms of the system that known, it is clear that human knowing is different from machine knowing.

The third cause of disparity has to do with the excessive speed in engineering (developing applications) without the prior, necessary scientific support (a physiological theory of knowledge). This has led to some superficiality in Knowledge Engineering (KE) proposals that have disguised the lack of theory with an inappropriate use of the language. Many of the AI and KE proposals only exist in the observer's language. The unjustified use of cognitive terms (intention, purpose, emotion, comprehension, learning, ...), taking for granted that they have the same meaning in humans as in computation, is the third cause of disparity between what the observer names and what the computer calculates.

2000 *Mathematics Subject Classification.* 68T01, 92B20.
Key words and phrases. Artificial intelligence, cybernetics, neural networks.

Finally, a fourth cause of disparity could be the oblivion of the solid work made by cybernetics in the "bottom-up" approach to intelligence, in terms of neural mechanisms, instead of using the dominant representational approach. In this work we reflect on these causes of disparity from a cybernetics perspective.

We began in section two by distinguishing between the analysis of natural intelligence (AI as a science) and the objectives and methods of AI as engineering (KE). Then we summarize in section three the different historical stages of AI enhancing its cybernetic "flavour" (mechanisms underlying intelligent behavior). We thus arrive to the current state of AI characterized by some methodological results along with the recognition of the necessity of integrating the different paradigms (symbolic, situated and connectionist). In sections four, five, six and seven we analyze the cybernetics roots of each one of these paradigms in the forerunning works of Craik, Wiener, and the McCulloch's school, including von Neumann contributions. We then conclude with some strategic recommendations concerning the usefulness of adopting a cybernetics approach to AI.

2 Concept of AI as Science and Engineering Two are the long term purposes of AI: (1) To develop conceptual models, formal tools, programming strategies and physical machines in order to help understand and explain the set of cognitive tasks, which give way to human behavior normally labeled intelligent, and (2) Try to synthesize a non trivial part of these tasks using the available materials and methods for knowledge embodiment [32]. These two purposes correspond to the perspectives of AI as natural science (analysis) and AI as engineering (synthesis), usually embracing KE and Robotics.

There are in turn different methodological and formal approaches, paradigms in the sense of Kuhn [19], in each of the two branches, which constitute different ways of evaluating hypothesis, theories, experimental results and explanation mechanisms. Here we will consider three of these paradigms: (1) Symbolic or representational, (2) connectionist and (3) embodied or situated [44].

As a science of the *natural* the AI phenomenology is the set of facts associated to cognitive processes. For this reason, its formal subject matter partially coincides with that of neurology and of the cognitive sciences, and its method tends to approach that of physics integrating theory and experiment. To formalize this knowledge, AI uses all the tools at its disposal (logic, mathematics, algorithmic, and heuristics), along with new ones which have arisen as a result of the field's specific needs (rules, frames, "agents", and learning techniques) or the proposal of new models of natural computation, such as membrane or DNA computing [9], genetic algorithms and evolutive programming.

We also acknowledge that it is quite likely that we do not yet have available to us all the formal tools needed for the computational representation of the most genuine aspects of the behavior of the nervous system, such as understanding and production of natural language, perception, creativity, imagination and all the processes around the emotional sphere [31].

Finally, AI laboratory is simulation. Understanding the nervous system mechanisms underlying intelligent behavior can get profit from computational simulation of nervous activity at different levels of organization. The situated, symbolic, and connectionist models are programmed and the results of the programming are *evaluated*. As a consequence of that evaluation, either the models are re-formulated, or the inference mechanisms are re-designed and new conclusions drawn for new experiments and/or explanation hypotheses.

That which distinguishes AI as a science with respect to physics, is that now information and knowledge have become formal scientific subjects, alongside matter and energy. As the formal subject matter of AI, it will undergo taxonomic observation, analysis, modeling, explanation and transformation. What we seek in the long terms is a theory of knowledge, which is computable, with a capacity for prediction, which is analogous to a physical law.

In other words, a theory that is impersonal, experimentally verifiable, and transferable. It is obvious that the task is not a simple one. AI as a discipline of synthesis aims to become a new engineering (Knowledge Engineering -KE-), in the strict sense of the word, with the methodology and efficiency of other engineerings, which deal with matter and energy. Now we begin with a set of functional specifications and seek the synthesis of a system (formal model plus program plus machine), which satisfies them.

When engineering has to do with matter or energy, its objective (the need to be satisfied) is to design a physical system. In the symbolic perspective of KE, we work with information and knowledge, and both are pure form, totally independent of the physical system, which supports them. Consequently, the purpose of the design is always related with the identification, modeling, formal representation and use of that knowledge in inference. KE's counterpart to conventional engineering's initial needs is a perception, decision, planning or control task. The result we are looking for is a computer *program* on a specific machine and developed upon a model of knowledge supposedly used by the human operator carrying out that task. The connectionist and situated perspectives of KE have shifted the interest towards the embodiment of these tasks in a specific system (a robot) that has to interact with a real physical environment.

The tasks considered by the symbolic approach to KE are high-level tasks, corresponding to that which we call cognitive processes and can be classified into three broad groups, organized according to progressive difficulty. This difficulty is evaluated as a function of the variety of entities and the relationships, which make them up, as well as the degree of semantics needed for its complete and unequivocal description: (1) Formal domains, (2) technical domains, and (3) basic genuine functions of human behavior. In all these cases KE try to emulate logical aspects of human reasoning, having always in mind that, at the end, the user of these emulation programs will be a human being.

The tasks in formal domains take on generic structure of "*problem solver*" by means of searches in a space of states of knowledge, and they can be games (for example, chess) or *logico-mathematical* problems (for example, theorem deduction, geometry, symbolic integration, etc...). They are generally tasks in which there is no imprecision in the knowledge, there are few elements, and their behavior can be described in a nearly complete and unequivocal way. They are "formal micro-worlds", corresponding to very large simplifications of the real world. The results obtained here are difficult to extrapolate.

The tasks, which use *scientific-technical* knowledge in narrow domains have to do, for example, with *medical diagnosis*, *classification*, *decision making*, *planning*, and *design*. The characteristic aspect to these scientific-technical tasks is the limited character of the knowledge they handle ("narrow" domains) and the possibility of formalizing that knowledge with available techniques (representation by means of rules or frames and inference by rule manipulation or by activating certain fields of the frames). This formalization is possible because of the little variety and the low level of semantics accepted as enough to describe the concepts and relations regarding that task. These technical tasks in narrow domains, has grown spectacularly in the last thirty years and given way to *"Knowledge Based Systems"* (KBS) or "*Expert Systems*" which, in general, are not related to the biological mechanisms underlying natural intelligence. Alternatively, the synthesis of these systems, is based on natural language descriptions of what the human expert is doing when develops a task figure 1[42].

3 Cybernetics and AI: An Intermingled History The dream of mechanizing the thought process (what we now call "making human knowledge computational" or synthesizing its cognitive processes into AI systems) is very distant and comes, as is the case with almost everything, from the classical Grecian culture. Although it is usually acknowl-

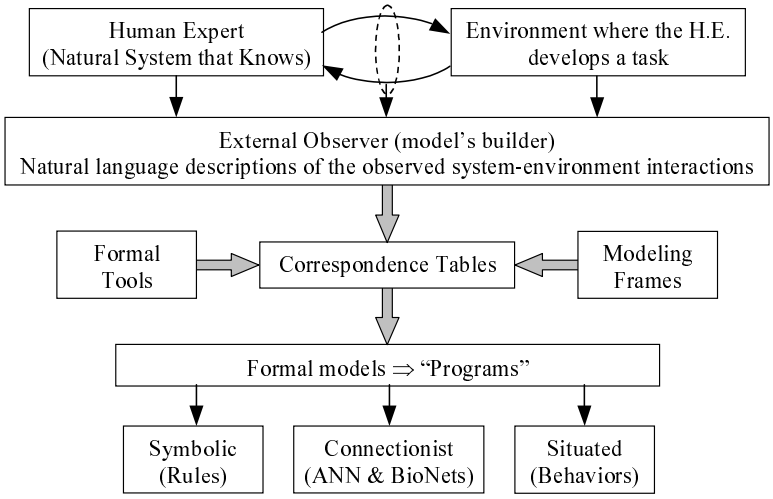

Figure 1: The natural language approach to AI and KE

edged that the birth of AI should be associated with the conference organized in 1956 at Dartmouth College [24], its real birth has its roots in Plato, Descartes, Boole, Leibnitz and Hobbes. Its neurocybernetics stage begins in 1943 with the pioneering works of N. Rashevsky [37], W.S. McCulloch and W. Pitts [26], J. von Neumann [4], N. Wiener [40, 49], J. McCarthy, A. Newell, M.L Minsky, S.C. Kleene, M.D. Davis, A.M. Uttley and C. Shannon, following a stage of formal domains (micro-worlds) and heuristics. The content of the book "*Automata Studies*" edited by Shannon and McCarthy in 1956 at the University of Princeton is essential to the understanding of the development of AI during this period [43]. The work of Alan Turing [45, 46] deserves special mention, both for its contributions to the computation model, and for its attempt to formalize AI by moving from the initial questions about whether or not machines could "think", to other more scientific and practical ones about the evaluation of the functionalities of a program (what we now know as the Turing test).

In the mixed history of AI and Cybernetics we can distinguish the following stages: (1) *Neurocybernetics*, (2) Symbolic computation, (3) Heuristics and micro-worlds, (4) Emphasis on knowledge representation, (5) *Rebirth of connectionism*, (6) *Situated approach (embodied knowledge)* and integration of paradigms.

3.1 Neurocybernetics AI began being neural computation when in 1943 Warren S. McCulloch and Walter Pitts introduced the first logical model of neuron [27] which we would today call minimal sequential circuit, formed by a logical function followed by a delay and in which programming is substituted by learning. Thus, a McCulloch-Pitts net of formal neurons is equivalent to Turing machine with the same memory capacity. If the net is programmable, then it is equivalent to a universal machine. The two models of computation (connectionist and symbolic) hence have equivalent power since its origins.

The basic contributions to AI during this period appear under the name of neurocybernetics and are based on the idea that both live beings and machines can be understood using the same organizational principles and formal tools (logic and mathematics). It is also believed that both should be broached at the *processor* level in terms of circuits and mechanisms, so that, to understand how the brain "computes", it is necessary to study

the anatomy and the physiology of the neurons and neural nets, model them formally, and observe how behavior emerges from local processes of feedback, convergence-divergence, lateral inhibition, reflex arches, associative memory and learning [3].

3.2 Turing, microworlds and emphasis on knowledge The 1950's see the passage from numerical to symbolic computation and programs for playing chess, or demonstrating theorems are developed, at the same time as Turing is proposing his well-known test. The recurring question since Descartes about whether "machines would one day be able to think" was reformulated in 1950 by Alan Turing in more precise terms by using the imitation game [46].

Following the neurocybernetics and computational precedents mentioned earlier, the year 1956 is usually considered the year of the birth of AI. It was coined by John McCarthy when he called the Dartmouth Conference based on the conjecture that "*every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it*" [24]. Four of the most influential researchers in this field were present at the conference: H. Simon, A. Newell, M. Minsky and McCarthy himself.

The first study, resulting from the search for universal deduction schemes, was the *"Logic Theorist"* program, which showed the real possibilities of AI, at least for the field of formal domains. Another study, which was representative of this period's thinking is the "General Problem Solver" program [36] which uses "means-ends" heuristics to resolve problems in a formal environment.

All the works of this first period (1956-196x) focused on problems characteristic of formal domains ("micro-worlds") which were idealizations of the real world: theorem proofs in propositional calculus, heuristic strategies in games, planning of actions and trajectories in the world of blocks, and games problems. In every case, AI is seen as a search in a space of problem states.

In the mid-sixties, gradual changes take place in the direction of AI, recognizing the limitations of the "general procedures for solving problems", bringing them closer to the problems of the real world and attributing increasing importance to the specific knowledge of the domain and, consequently, the problems associated with knowledge representation and subsequent use of these models of knowledge in inference, where there is a separation of knowledge and its use. The work of Minsky [29] and Winograd [50] are representative of the limitations of a methodology, which approaches natural language by limiting the domain in order to obtain a complete model of the structures and processes of that domain. Dreyfus [14] criticizes that period recalling that -in practically every case- the authors had an excessively optimistic view on the possibilities inherent in their programs, as has been proved to be the case forty years later. The so-called fallacy of the "first step" is very significant. It was said about practically every program that it was a first step towards the understanding of natural language, for example. However, the second step never came.

In the *intermediate* stage (197x-198x) emphasis was on knowledge-based systems (KBS) with a fairly generalized opinion about the limitations of logic as the sole representation language [5]. Rules, associative nets, frames and agents has been added as representational tools, but the problem of knowledge representation still persists, both in *theoretical* AI (what constitutes knowledge? what dependence relationship exists between knowledge and the structure that know?) and in *synthesis* AI (what representation is most efficient, expressive and complete? which one gives way to a more efficient subsequent use when that knowledge is used to reason? how can the problem of semantics be approached? how do we formalize approximate, non-monotonic, temporal, common sense, or qualitative reasoning?).

3.3 The rebirth of connectionism As we mentioned, AI was connectionist at its origin. The first studies were related to the synthesis of networks of artificial neurons capable of recognizing characters, storing information, or "reasoning". In turn, learning was understood in terms of processors. To learn was to modify the value of the parameters of a system to improve its behavior. Between this initial stage and around 1986, AI has been dominated by the symbolic or "representational" paradigm and followed the previously discussed lines. Finally, the last two decades witnesses a very strong rebirth of the Cybernetics roots of AI, with the rebirth of connectionism (artificial neural nets) and the shift of attention towards situated approaches. The symbolic paradigm of AI has been somewhat successful in emulating human decision-making but has been less successful in dealing with those situations in which we have more data than knowledge or we need to situate our KBS on a physical system (a robot). Along with this there is the generalized opinion that it will be difficult to progress in AI without considering learning in any system having to manage massive amounts of information and knowledge, which come from an unstructured and changing environment.

The growth of symbolic AI and the recognition of the limitations of formal perceptron-type neural nets put forth by Minsky and Papert in 1969 [28] slowed down connectionism at the end of the 60's. Nevertheless, in these AI stages (1956-1986) solid work in neural modeling is continued as summarized in figure2. We can mention the non linear analogic models of Caianiello [7], the probabilistic formulations of von Neumann [48], Selfridge [1], [Moreno and McCulloch [34], several models of associative memory produced by Anderson and Rosenfeld [2], and Kohonen [18], the neural pattern recognition systems around Fukushima [15], models of learning [16], and vision [22, 21]; models of cooperative processes to interpret the neural function at high level [17, 13] and the biophysical and collective models which demand an extension of neural computation towards the formulations of statistical physics as proposed by Ricciardi [38, 39], and some more comprehensive formal modeling frames [33].

The qualitative step in AI's connectionist perspective appears at the beginning of the 80's with Rumelhart et al's proposals [41][40] and Barton [6]. The essential difference is in the substitution of the threshold function by a derivable function, which allows retropropagation of the error function measured at the output layer to the inner layers, for which we don't have measured values, by the gradient method. This technical argument along with a certain sense of stagnation in the field of symbolic KBS's have given way to the rebirth of connectionism.

Unfortunately, in this rebirth of connectionism the same models of the past were used (multilayer perceptrons, radial basis functions, threshold logic, first order linear differential equations -"integrate and fire"-) forgetting the deep initial motivations of Cybernetics along with a lack of theoretical developments.

3.4 The current state: Some methodological developments, embodiment of knowledge and integration of paradigms Three aspects can be considered as representative of current state in AI. The first one is the recognition of the limited value of the computational metaphor when the calculus is described at only one level and without reference to the external observer. In 1981 Newell introduced the "knowledge level" [35] and Marr [21] the equivalent "theory of calculus", as a new level of description on the top of the physical and symbol levels previously recognized. As Clancey [11] suggests the Newell and Marr idea of levels is dominated by architectural metaphors. They consider the relation between the three levels of description of a calculus as correspondence tables (as mappings). In the nervous system these organizational levels are nested; all the mechanisms underlying the phenomenology of the three levels are operating concurrently, in a coordinate manner.

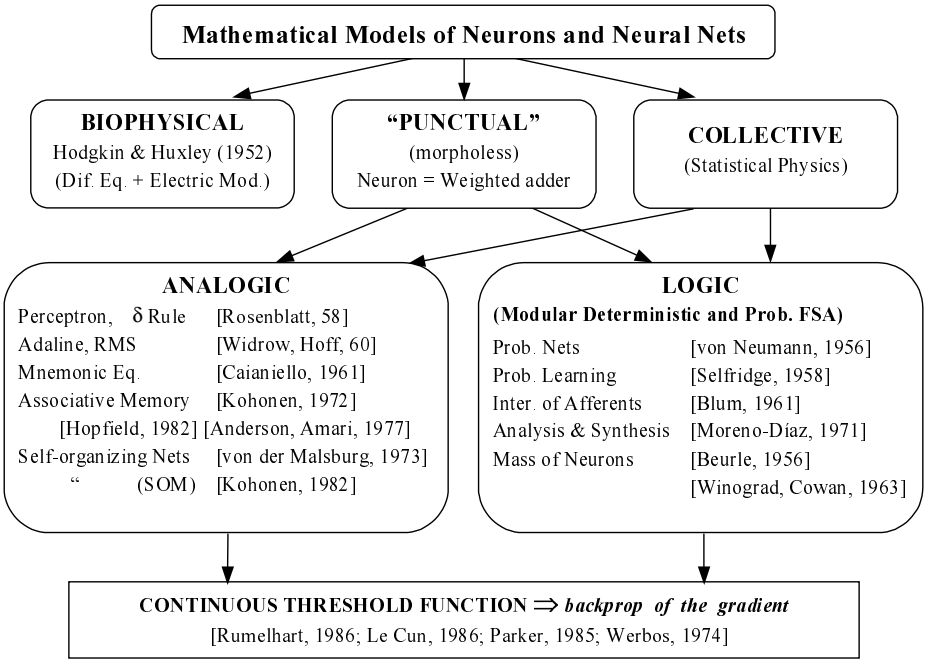

Figure 2: A summary of formal models of neurons and neural nets from 1952 to 1986, developed concurrently with the stage in which the symbolic paradigm dominated the AI scene

In order to complete our understanding of the meaning of a calculus at these three levels a new methodological distinction is introduced inside each level, between two description domains: the levels *own domain*, and the *external observer domain*. The introduction of the figure of the external observer and the differentiation between the internal semantics of a mechanism and that of its description comes from physics and has been re introduced and elaborated in biology by Maturana [23] and Varela [47] and in computation by Mira and Delgado[30, 32].

The second distinctive characteristic of AI current state is the growing importance of the situated paradigm that emphasizes two points: (1) All perceptions and all actions of a system are in the context of an external environment (including other systems) to which the system under consideration is structural and evolutively coupled, and (2) this dynamic and coupled interaction emerges as a consequence of a set of underlying mechanisms in both, the system and the environment.

The first point gives way to consider an ethological description of the interactions between the system and its environment as the first step of our conceptual models. The second point is the usual in cybernetics and highlights the importance of feedback mechanisms and the dynamic nature of intelligence. Instead of thinking in terms of the storing of descriptions in a programmable computer we would think in terms of a special purpose adaptive machine, whose function is modified as a result of the interactive activity of its mechanisms with the mechanisms of the environment.

During the last fifty year there has been an enormous lack of balance between KBS for human use and physical systems that have to deal with the real world through perception and motor actions. Taking the biological inspiration seriously implies building intelligence on top of a specific sensory-motor system (a "body").

The last aspect in the current state of AI is concerned with the need of integration. People working independently in each one of the three paradigms have recognized their limitations and decided to cooperate by seeking integrated formulations of the computational models from a set of libraries of "reusable components for knowledge modeling", including "tasks", "problem solving methods", "ontologies", "inferences", "roles", and formal tools. We believe that the connectionist, symbolic and situated perspectives to AI should be taken as mutually supporting approaches to the same problems. The extent of the global problem, which AI deals with, will oblige us to be modest for several generations. In figure 3 we summarize these methodological points.

4 Kenneth Craik and the Symbolic Paradigm The symbolic paradigm, as the other two, has its roots in the pioneering work of Cybernetics. In this paradigm human knowledge is represented in terms of declarative and modular descriptions of high level entities ("concepts") and relations between these natural language entities ("inferential rules") in such a way that reasoning is understood as a process of manipulations of these symbolic descriptions.

In his book *"The Nature of Explanation"* [12] Kenneth Craik interprets nervous system activity in terms of a set of processes geared at building an internal representation of the environment (a model) and using it to predict. To learn here is to accumulate knowledge by updating the model of the environment. Craik contributed to modern AI with two key contributions: *abductive reasoning and representational spaces*.

Inference in AI is associated to the individual or combined use of three types of reasoning: deductive, inductive or abductive. In *logical deduction*, we begin with a set of formulas considered to have general validity (axioms) and a set of rules or proof procedures are applied to them, which allows us to obtain new valid formulas. We go from the general to the particular and can guarantee the accuracy of the obtained result. The problem with

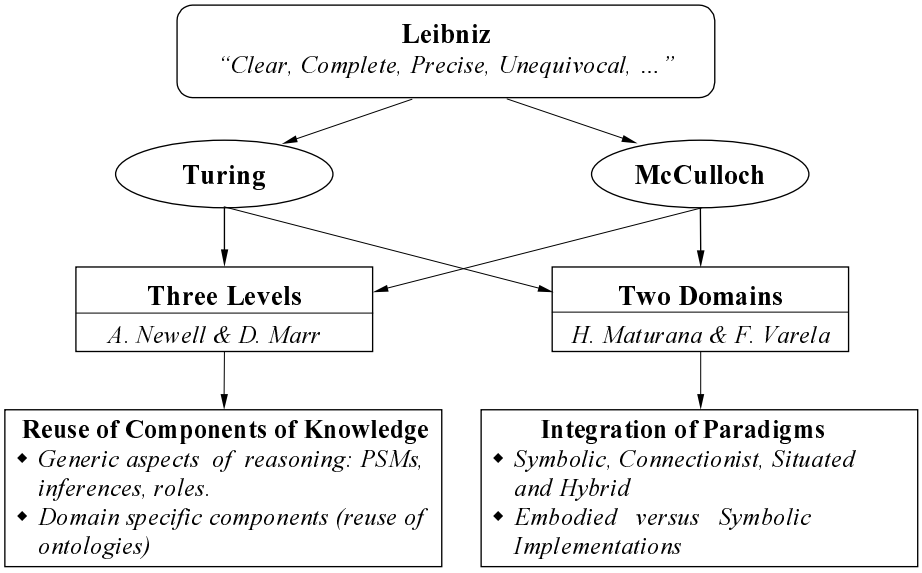

Figure 3: Methodological steps in the conceptual and historical perspective of AI

this type of inference is that we need to precisely establish the set of beliefs (axioms), which completely define the knowledge of the domain.

In *inductive inference*, we go from the particular to the general, generalizing the extractable information from particular cases. To do this we use "clues" (heuristics) with the knowledge of the domain, but we can never guarantee the completeness and accuracy of inference. There is always uncertainty as to the validity of the assumptions.

Finally, in *abductive reasoning*, we begin with a known conclusion and look for a fact to explain it. It is used in situations (such as diagnosis) in which the cause-effect relation is known along with one fact (effect) from which a hypothesis is made about its most probable cause. Craik, along with Pierce, are clear forerunners of the current work being carried out in the field of AI. In 1943, Craik made the first attempt at operational definitions of concepts such as causality, meaning, implication, and consistency, along the same lines as Turing later followed when he spoke about intelligence. The important question is not about what causality or intelligence is, but how they can be modeled (reconstructed) by creating programs, which duplicate them. That is to say, what structure and processes are needed for an electromechanical system to be able to imitate biological systems correctly? In the words of Craik, *"our question, to emphasize it once again, is not ask what kind of thing a number is, but to think what kind of mechanism could represent so many physically possible or impossible, and yet self-consistent, processes as number does.*"

The second suggestion in Craik's work is a mechanism, which reasons by analogy within the model of the environment where the formal implication is equivalent to causality in the physical world. Craik distinguished three processes:

1. A "translation" of external process into internal representation in terms of words, numbers, or other symbols (i.e. images).

2. Derivation of other symbols from them by some sort of inferential process.

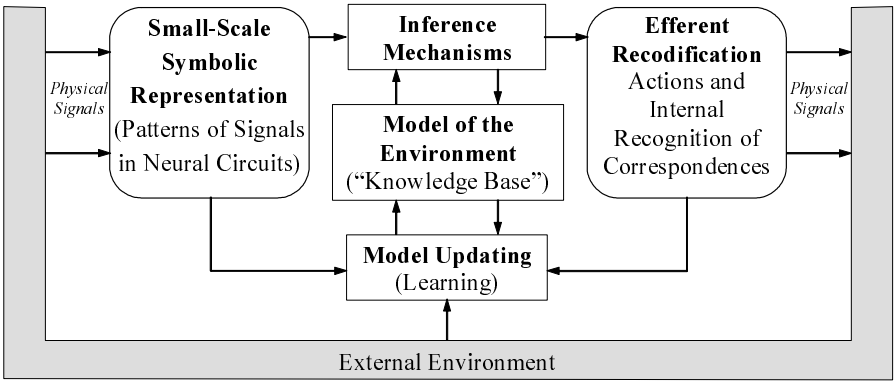

Figure 4: An up-dated version of the architecture implicit in the Craiks proposal for human reasoning

3. A "retranslation" of these symbols into actions, or at least a recognition of the correspondence between these symbols and external events, as in realizing that a prediction is fulfilled.

This inferential process in the model of the environment produces the symbolic results, which are equivalent to what the physical causality modeled, would have produced. The idea of a model of the environment has been present from the founding period to modern robotics and the new notions of knowledge modeling. The basic characteristic of what we now call a KBS is its power to model facts based on symbols. The idea implicit in Craik's work is that the coding of these symbols is not arbitrary and retains its identity from the sensory input, to the motor output. The internal relations between symbols maintain the consistency and interdependence of the external entities, which they represent. For Craik, thought models reality ("small scale models") and the essential feature of nervous system activity is not "the mind", "the self", "sense data", nor propositions but symbolism. Symbols in the model have similar relation-structure to that of the entities of the reality they model. In figure 4 we show an up-dated version of the architecture proposed in Craik's work, as a mechanism which reasons by analogy within the model of the environment.

5 Norbert Wiener and the Concepts of Information, Feedback and Purpose

The idea of using structural models which are analogous to those used by continuous systems theory to model non-analytical knowledge was introduced in AI by Clancey [10], along with Chandrasekaran[8], among others. These structural models, called "generic tasks", are inferential circuits of "functional blocks" (verbs "compare", "select", "evaluate", "measure"...), which contribute to the solution of an extensive family of problems.

Three important concepts underlying this idea of "generic tasks" were introduced by Wiener in 1943 [40] and 1948 [49]: (1) The notion of information as pure form, separable from the physical signal which carries it, (2) the concept of generic control and oscillation structure (feedback loop), and (3) the interpretation of *purposive* behavior in terms of a goal state and the negative feedback mechanism which enables the system to reach it. In figure 5 we show the up-dated version, in terms of roles and inferences, of the control loop proposed by Wiener. The *goal state* is the *input role* (to maintain the constant value of a magnitude or to make that magnitude to follow the temporal evolution of a "control signal"). In order to achieve this goal, three functions (inferences) are considered necessary

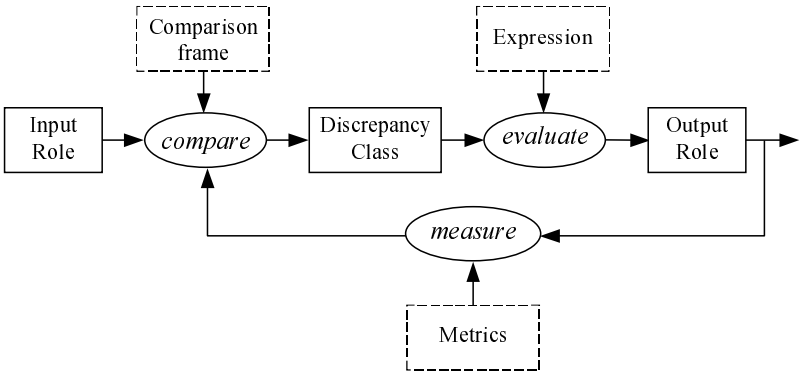

Figure 5: Inferential scheme of the control and oscillation tasks abstracted from the Wieners proposal to explain the neural mechanisms underlying purposive behavior

(comparison, evaluation and action performance measurement), as well as its connection structure: A feedback loop comparing the value of the output magnitude (output role) with the input role value and acting to minimize the different between the goal state and the current state, which is constantly being measured to inform the comparation inference. Note that this analysis and the feedback inferential scheme are totally independent of the physical dimension of input and output signals (we only speak of roles), and of the physical and formal nature of the inferences. In each specific case signals, physical implementations and operators will vary, but the inferential scheme and the organization of the task always remain.

6 The W.S. McCullochs School and the Situated Paradigm The situated approach is also known as "behavior-based", or "reactive". A behavior is defined as a direct mapping between a class of "input patterns" and a predefined "pattern of motor actions", so that the correspondence with a finite state automaton (FSA) is immediate: "Behaviors" are associated with states or sequences of states, including oscillatory loops, input patterns are associated with abstract symbols of the FSA input alphabet and patterns of actions with the elements of the FSA output alphabet. Observe however that in this external (functional) view of the situated approach the structural correspondence with the lower organizational levels is also lost, because we talk of *"behaviors"* "*patterns of perceptions"* and "*patterns of actions*", but we do not say anything about the correspondence between these natural language terms and the neurophysiological mechanisms supporting them. On the contrary, the pioneering work of McCulloch's school on epistemological physiology, clear forerunner of the situated approach, always took the knowledge embodiment problem from a dynamical point of view, based on the mechanisms that embodies this knowledge.

Figure 6 shows a diagram of the set of perceptual, associative, and motor schemes used by the situated paradigm to describe the interactive behavior of a living being and its environment Lettwin et al [20], would say the animal abstracts what is useful for him, from his surrounding, in order to get an efficient reaction in real time.

In 1952, W.S. McCulloch published the paper "*Finality and Form in Nervous Activity"* [26] where he mention a set of non-solved problems in theory of knowledge (self-reference, circular causality, closure, intentions, embodiment of ideas and values...) and clearly propose that these problems "must be stated and resolved as questions concerning the anatomy and the physiology of the nervous system". In terms of circuits (mechanisms) and modes of

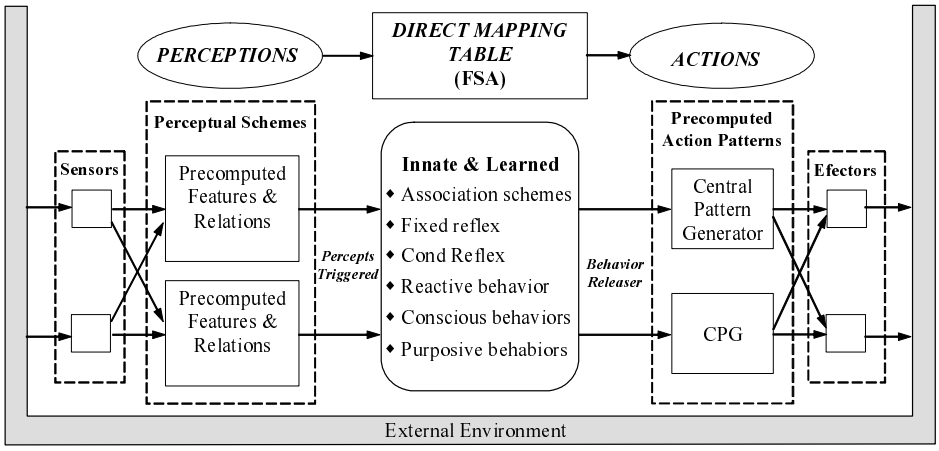

Figure 6: Summary of tasks describing the interaction of an "intelligent agent" with the environment, according to the situated paradigm

activity of these circuits. Intelligence (natural or artificial) is determined by the structure of the nested mechanisms from which it emerges. Figure 7 show a diagram of the three types of feedback loops proposed by W.S. McCulloch to explain circular causality, in the same line as proposed Wiener. In words of McCulloch, circular causality is embodied in the "*trains of nervous impulses that return to their place of origin, where they diminish or reverse the process that give rise to them.*"

W.S. McCulloch also addressed the embodiment of circularities of preference (figure 8) for which we need: (1) A closed loop for each "aim" or goal and (2) a scheme of dominance by means of specific inhibition patterns.

The paper of Lettvin, Maturana, McCulloch and Pitts "*What the frog's eye tells the frog's brain*" published in 1959 [20], can be considered of founding importance in situated robotics. We do not know better theoretical description, nor experimental evidence, of the existence of *"pre-computed perceptual schemes"* than the one proposed in this study of the frog's visual system. The animal abstracts what is useful fir him from his surrounding in terms of *four* parallel-distributed channels, each one tuned to different sort of patterns (local sharp edges and contrast, "bug" perceivers, local movements of-object's edges, movements or rapid general darkening). What the authors are stating is that the frog has "*a language of complex abstractions from the visual environment*" and that these abstractions are pre-computed in such a way that the occurrence of the corresponding situations only has to "trigger" or "release" the corresponding labels in order to enable a fast response. This codification produces dimensionality reduction and enables selective attention mechanisms.

The situated AI proposal on the release of pre-computed output schemes also has its roots in the Kilmer and McCulloch work on modal co-operative decision making [17] and the hypotheses that "the core of the reticular formation is the structure that commits the animal to one or another mode". The S-Retic model is characterized by: (1) Redundancy sampling of sensorial signals, (2) high connectivity (local and distal), (3) recursive lateral interaction between modules, (4) a probabilistic formulation of the consensus rule (an specific mode is selected if more than 50% of the modules support this mode with a probability value that surmounts 0.5). Some problems of the S-Retic, as the need of an external observer to "monitorize" the situation, are tackled in a more recent version of this model where the

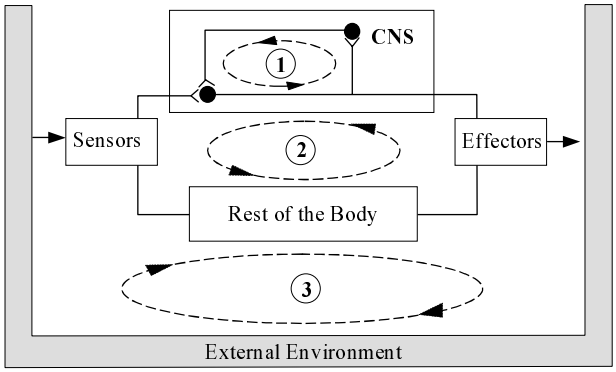

Figure 7: The three types of feedback loops proposed by W.S. McCulloch: Loops type 1 for tasks of control, timing, memory, central pattern generators, and other sequential functions. Loops type 2 for reflexes (involuntary, stereotyped and graded) and for fixed action patterns. Loops type 3 for purposes and values

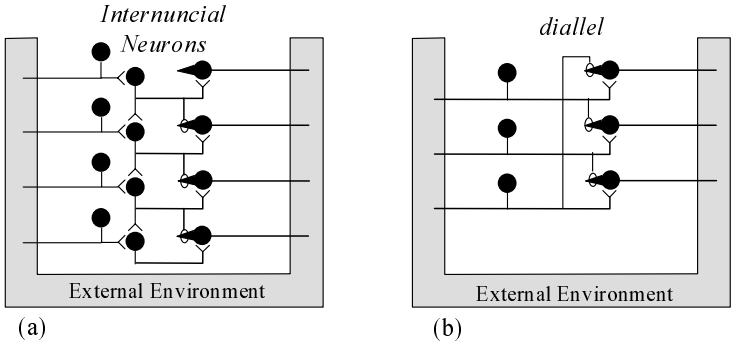

Figure 8: Circularities of preference: (a) hierarchy and (b) heterarchy

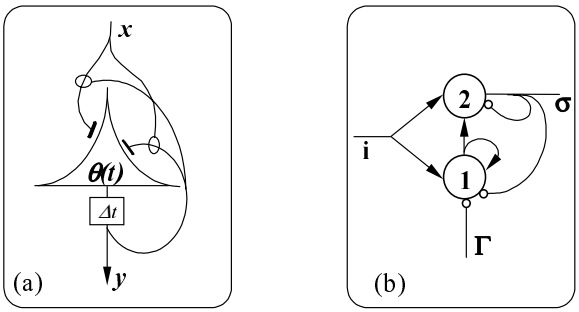

Figure 9: Two elementary examples of neural circuits. (a) McCullochs temporal evenness detector and (b) von Neumans module used as design component of his stored program computer

control is distributed among all the modules [13].

7 The Connectionist Paradigm A clear alternative to consider intelligence as stored descriptions of high level concepts and inferential rules in a conventional programmable computer is try to specific the neural circuits, mechanisms and organizations underlying intelligent behavior. This was the initial proposal of connectionism, as understood by cybernetics and under this interpretation the situated and connectionist paradigms are equivalent. In this way you never lose the causal structure of the entities and relations (connectivity patterns) constitutive of each calculus. At the same time you can incrementally contribute to build up a neural theory of knowledge. Von Neumann contrasted this connectionist approach to intelligence from axiomatically defined simple components. to Turing symbolic approach of giving an axiomatic presentation of the whole system without taking care of defining the individual components, nor the connectivity schemes between these components, from which the function emerges [4].

W.S. McCulloch used the simple circuit of figure 9.a to illustrate the idea of embodied calculus. This formal neuron only triggers when the number of times it has been activated is odd. It is a detector of the evenness or oddness of a train of pulses. In figure 9.b we can see one of the elementary mechanisms used by von Neumann to design the stored program computer. The idea is that each calculus has a dedicated mechanism from which the observed function emerges.

Current connectionism has lost this orientation and is considered as a way of modeling, formalization and programming problem solving methods in those situations in which we have more data than knowledge. The architecture of a connectionist system is modular and distributed over a set of layers in which each element of calculus is a "small-grain" parametric processor (figure 10) and where a part of the programming is substituted by learning, using supervised or unsupervised mechanisms. Stated in these terms, the artificial neural nets (ANN) are general-purpose adaptive numeric associators, as illustrated figure figure 11.

Independently of the value of ANN's as a complement to the symbolic methods of AI in solving problems having to do with changing, and only partially known environments, which need trainable, parallel and real time adaptive solutions, the connections with biological neurons and neural nets has been lost. A relevant set of anatomic and physiological properties of real neurons are not considered. Let us mention the differences between the

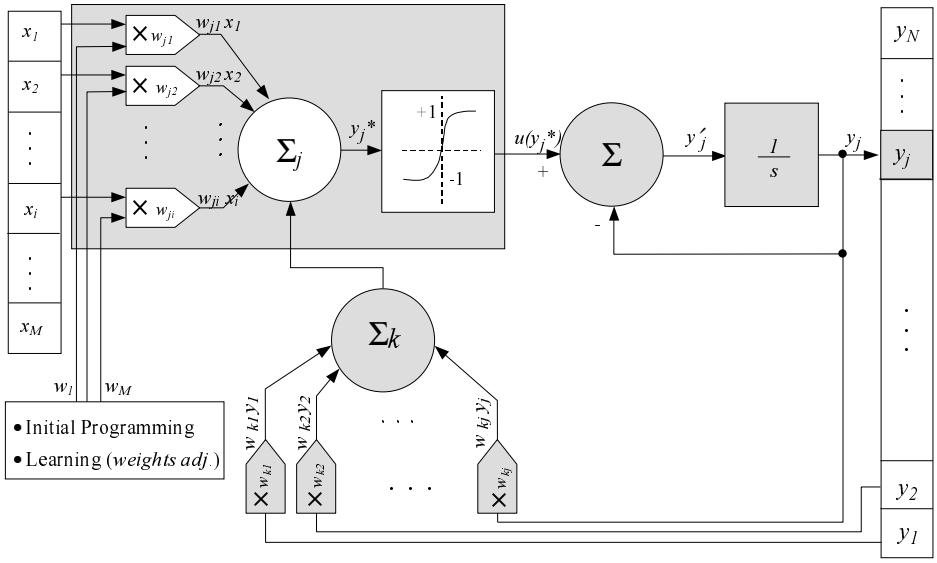

Figure 10: Integrate and fire dynamic model of artificial neuron, as currently used in the connectionist paradigm of AI. The static part of the model is a spatial and weighted sum followed by a threshold function to decide whether or not the axon is triggered. The dynamic part is a temporal integrator along with a local feedback loop. Learning is understood as a process of modification of the weighting factors

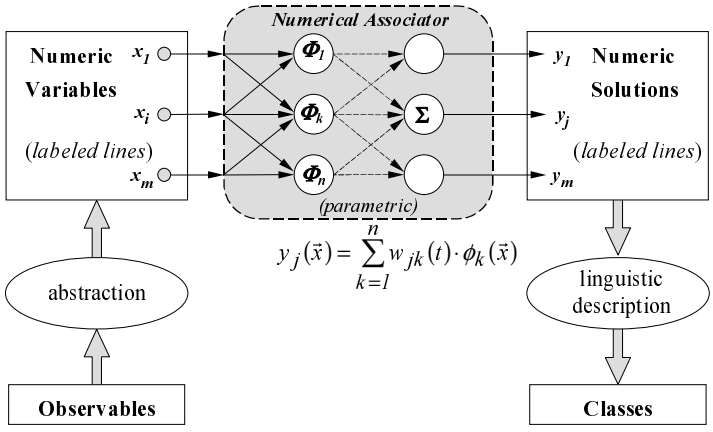

Figure 11: The ANN is a parametric numerical classifier that links a certain number of input labeled lines (patterns to be classified) with a few output labeled lines (classes). The values of the parameters are dynamically adjusted, and the user arbitrarily establishes the meaning of inputs and outputs during the data analysis phase

model of the synapses (a sigmoid) and the well-known characteristics of these unions between biological processors: (1) There are quick and slow (1 millisecond) synapses which allows the coexistence of two levels of intercommunication, response and learning. (2) They can be excitatory or inhibitory and with all analytical operators (add, subtract, accumulate, decrease, multiply, divide,...). (3) Chemical synapses are selective of different types of messengers (neurotransmitters) which allow neural computation to be interpreted as "the passage of messages" and "specific conditionals" with the wealth and diversity of proteins. (4) They act in both a cooperative and a competitive manner. The simplest expression of this cooperation is the spatio-temporal integration of the different contributions. The analytical sum is a very poor version of the biological concept of integration. (5) There are neurons with more than 100.000 contacts of this type, which means an average of more than $10^3 - 10^{10}$(10 billion) connections. (6) These contacts are plastic and constitute one of the structural supports of self-programming (learning) and memory.

It is our deep feeling that current analog and digital models of neurons and neural nets are not enough to cope with the computational complexity and wealth observed in biological neurons and that more realistic models are needed.

In the previous sections of this paper we have examined the other two AI paradigms. The aim of the final part of this section is to make some methodological considerations regarding similarities and differences between the three basic approaches to AI. We do it summarized in three points:

1. There is no essential difference between symbolic, situated, and connectionist AI at the level of physical processors. All computation is connectionist in the physical level. The difference between number and message appears in the domain of the external observers, when we assign meaning to the entities of the input and output spaces.

2. The jump from the knowledge level to the physical level is more severe in the symbolic and situated cases than in connectionism. For this reason, more effort must be devoted to the analysis of the tasks in the two first cases and more effort is needed in the analysis of data in the last case.

3. The main difference between current approaches to AI and Cybernetics is the substitution of *external programming* by the design of specific *neural mechanisms* to take care of the same tasks.

8 Conclusions In this work, taking advantage of the fiftieth anniversary of Artificial Intelligence (AI), we have considered: (1) the great disparity that exists between the initial objectives of synthesizing general intelligence in machines and the modest results obtained after half a century of work, and (2) the oblivion of the cybernetics roots of current AI paradigms.

To support these considerations we have begun by distinguishing between the analysis of natural intelligence (AI as a science) and the objectives and methods of KE (AI as engineering). The key point in this distinction is that KE does not require us to fully understand the mechanisms of human beings from whom a behavior emerges that we label as intelligent. AI must not aim to copy natural intelligence, but use it as a source of inspiration.

Then we have summarized the different historical stages and have recalled Newell, Marr, Maturana and Varelas proposals, which led to the methodological framework of levels and domains of description of the knowledge intervening in the full specification of a calculus. Only a small part of this knowledge (the formal underlying model) is finally computable, the rest is outside the machine, in the external observers domain.

With this methodological support we have analyzed the cybernetics roots of the three dominant AI paradigms.

We now conclude with two suggestions: (1) to recover the interdisciplinary atmosphere of Cybernetics and promote the interplay between Neuroscience and Computation to look after the neural mechanisms underlying intelligence. (2) To eliminate the most frequent mistake in these 50 years of development of AI which considers that the words that we use in the construction of our conceptual models are directly computable, that the description of a behavior coincides with the mechanism from where this behavior emerges. This historical mistake has led us to assume that our AI programs accommodate a lot more knowledge than they really accommodate. A large part of the supposedly computable knowledge is outside the computer, and here lies the real work, in achieving that new intermediate organizational layers (new models and formalization mechanisms) enable an increasing amount of knowledge to reside in a computer.

The basic question in AI is not what "computers can or cannot do" [22], but the amount and type of knowledge that we humans will be able to model, formalize and program so that finally it is computable in a machine. This is an open question whose answer depends on developing materials and architectures and making progress in modeling, formalization and programming techniques. In any case, by posing the question like this, we place it in the context of conventional science and engineering, without excessive nomenclatures or unjustifiable meanings, and consequently, we help limit the area in our quest for solutions.

Acknowledgments The authors would like to acknowledge the financial support of the Spanish CICYT under project TIN2004-07661-C02-01.

References

[1] *Pandemonium: a Paradigm for Learning. Mechanization of Thought Processes*, London, 1958. Proc. of a Symposium Held at the National Physical Laboratory, HMSO.

[2] A. Anderson, J, A. Pellionisz, and E. Rosenfeld, editors. *Neurocomputing 2: Directions for Research*. The MIT Press, Cambridge, MA, 1990.

[3] J. A. Anderson and Rosenfeld (eds.), editors. *Neurocomputing: Foundations of Research*. The MIT Press, Cambridge, 1989.

[4] W. Aspray. *John von Neumann and Origins of Modern Computing*. The MIT Press, Cambridge, Mass, 1990.

[5] A. Barr and E. A. Feigenbaum. *The Handbook of Artificial Intelligence*, volume I and II. William Kaufmann, 1981.

[6] A. G. Barton, R. S. Sutton, and C. W. Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans. on Systems, Man and Cybernetics*, 13 no. 5:834–846, 1983.

[7] E. Caianiello. Outline of a theory of thought-processes and thinking machines. *Journal of Theoretical Biology*, 1:204–235, 1961.

[8] B. Chandrasekaran. Generic tasks in knowledge-based reasoning: High-level building blocks for expert systems design. *IEEE Expert*, 1:23–29, 1986.

[9] G. Ciobanu, G. Paun, and M.J. (eds.) Pérez-Jiménez, editors. *Applications of Membrane Computing. Natural Computing*. Springer, Berlin, 2006.

[10] W. J. Clancey. Heuristic classification. *Artificial Intelligence*, 27:289–350, 1985.

[11] W. J. Clancey. *Conceptual Coordination*. Lawrence Erlbaum Associates, New Jersey, pub. mahwah edition, 1999.

[12] K. Craik. *The Nature of Explanation*. Cambridge University Press, Cambridge, 1943.

[13] A. E. Delgado, J. Mira, and R. Moreno-Díaz. A neurocibernetic model of modal co-operative decision in the kilmer-mcculloch space. *Kybernetes*, 18, n?3:48–57, 1989.

[14] H. L. Dreyfus. *What Computers Still Can't do*. The MIT Press, Camb. Mass., 1994.

[15] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:139–202, 1980.

[16] S. Grossberg. A neural theory of punishment and avoidance. part i: Qualitative theory. *Mathematical Biosciences*, 15:39–67, 1972.

[17] W. Kilmer and W. S McCulloch. *The Reticular Formation Command and Control System*. Berlín, 1969.

[18] T. Kohonen. Correlation matrix memories. *IEEE Translations on Computers*, C-21:353–359, 1972.

[19] T.S. Kuhn. *La Estructura de las Revoluciones Científicas*. Fondo de Cultura Económica, México, 1971.

[20] J. Y. Lettvin, H. Maturana, W. S. McCulloch, and W. H. Pitts. What the frog's eye tells the frog's brain. page 1940 1951. Proceedings of the IRE, 47. No. 11, 1959.

[21] D. Marr. *Vision*. Freeman, New York, 1982.

[22] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.

[23] H. R. Maturana. The organization of the living: A theory of the living organization. *Int. J. Man-Machine Studies*, 7:313–332, 1975.

[24] J. McCarthy, M. L. Minsky, N. Richester, and C. E. Shannon. A proposal for the dartmouth summer research project on artificial intelligence. Technical report, Hannover, New Hampshire, 1955.

[25] R. (ed.) McCulloch, editor. *Collected Works of Warren McCulloch*. Intersystems Publications, California, USA, 1989.

[26] W.S. McCulloch. *Embodiments of Mind*. The MIT Press, Cambridge, Mass., 1965.

[27] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[28] L Minsky, M and S. Papert. *Perceptrons: An Introduction to Computational Geometry*. The MIT Press, Cambridge, MA, 1969.

[29] M. Minsky. *Semantic Information Processing*. MIT Press, Cambridge, MA, 1968.

[30] J. Mira and A. E. Delgado. Some comments on the antropocentric viewpoint in the neurocybernetic methodology. In *Proc of the Seventh International Congress of Cybernetics and Systems*, pages 891–95. 1987.

[31] J. Mira and A. E. Delgado. Neural modeling in cerebral dynamics. *BioSystems*, 71:133–144, 2003.

[32] J. Mira, A.E. Delgado, J.G. Boticario, and F.J. Díez. *Aspectos básicos de la inteligencia artificial*. Sanz y Torres, SL, Madrid, 1995.

[33] J. Mira, R. Moreno-Díaz, and A. E. Delgado. *A Theoretical Proposal to Embody Cooperative Decision in the Nervous System*. Seaside, California, 1983.

[34] R. Moreno-Diaz and W. S. McCulloch. *Circularities in Nets and the Concept of Functional Matrices*, pages 145–150. Little-Brown, MA, 1968.

[35] A. Newell. The knowledge level. *AI Magazine*, 120, 1981.

[36] A. Newell, J.C. Shaw, and H. A. Simon. *A Variety of Intelligent Learning in a General Problem Solver*, volume Self-Organizing Systems, pages 153–189. Pergamon Press, 1960.

[37] N. Rashevsky. *Mathematical Biophysics. Physico-Mathematical Foundations of Biology*, volume I and II. Dover Pub. Inc., New York, 1938.

[38] L. Ricciardi. Diffusion processes and related topics in biology. In *Lecture Notes in Biomathematics*. Springer-Verlag, Berlin, 1977.

[39] L. Ricciardi. *Diffusion Models of Neuron Activity*, pages 299–304. The MIT Press, 1995.

[40] A. Rosenblueth, N. Wiener, and J. Bigelow. Behavior, purpose and teleology. *Philosophy of Science*, 10, 1943.

[41] D.E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning Internal Representations by Error Propagation*. Cambridge, MA, 1986.

[42] G. Schreiber, H. Akkermans, and R. de Anjo Anjewierden. *Engineering and Managing Knowledge: The CommonKADS Methodology*. The MIT Press, Cambridge, Mass, 1999.

[43] C. E. Shannon and J. (Eds.) McCarthy, editors. *Automata Studies*. Princeton University Press, Princeton., 1956.

[44] S.C. (Ed.) Shapiro, editor. *Encyclopedia of Artificial Intelligence*, volume I and II (2nd edition). John Wiley & Sons, 1990.

[45] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. pages 230–265. Proceedings of the London Mathematical Society (series 2) 42, (1936.

[46] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.

[47] F.J. Varela. *Principles of Biological Autonomy*. The North Holland Series in General Systems Research, New York, 1979.

[48] J. von Neumann. *Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components*. Princenton, New Jersey, 1956.

[49] N. Wiener. *Cybernetics*. The Technology Press. J. Wiley & Sons, New York, 1948.

[50] T. Winograd. *Understanding Natural Language*. New York, 1972.

José Mira, Ana E. Delgado
Departamento de Inteligencia Artificial, E.T.S.I. Informática
UNED. 28040 - Madrid, España
E-mail: {jmira,adelgado}@dia.uned.es