13
531

CONSIDERATIONS IN DESIGNING A CYBERNETIC SIMPLE 'LEARNING' MODEL; AND AN OVERVIEW OF THE PROBLEM OF MODELLING LEARNING.

Emin Der-Kureghian

A Thesis submitted for the
degree of Doctor of Philosophy
in the Department of Cybernetics at
Brunel University, Uxbridge, Middlesex.

1 9 8 8

TO MY PARENTS

ABSTRACT

Brunel University, Uxbridge, England, Department of Electrical and Electronic Engineering, Division of Cybernetics.

Emin Der-Kureghian

"CONSIDERATIONS IN DESIGNING A CYBERNETIC SIMPLE LEARNING MODEL; AND AN OVERVIEW OF THE PROBLEM OF MODELLING LEARNING."

1988

Learning is viewed as a central feature of living systems and must be manifested in any artifact that claims to exhibit general intelligence.

The central aims of the thesis are twofold: (1) - To review and critically assess the empirical and theoretical aspects of learning as have been addressed in a multitude of disciplines, with the aim of extracting fundamental features and elements. (2) - To develop a more systematic approach to the cybernetic modelling of learning than has been achieved hitherto.

In pursuit of aim (1) above the following discussions are included:

- Historical and Philosophical backgrounds;
- Natural learning, both physiological and psychological aspects;
- Hierarchies of learning identified in the evolutionary, functional and developmental senses;
- An extensive section on the general problem of modelling of learning and the formal tools, is included as a link between aims (1) and (2).

Following this a systematic and historically oriented study of cybernetic and other related approaches to the problem of modelling of learning is presented.

This then leads to the development of a state-of-the-art general purpose experimental cybernetic learning model. The programming and use of this model is also fully described, including an elaborate scheme for the manifestation of simple learning.

ACKNOWLEDGEMENTS

I would like to express my immense gratitude and indebtedness to my supervisor, Dr. Mike Elstob, who has been a constant source of inspiration, encouragement, and help during the long and arduous course of my research. Additionally, I would like to thank other members of the Cybernetics and Computer Science Departments for providing invaluable intellectual stimulus, and assisting me during the construction and implementation of my hardware project; and, also, to Dr. Darwin Minassian for supplying computing facilities for the compilation and printing of my thesis. Finally, I would like to extend my appreciation and gratefulness to my family who have supported and unreservedly encouraged all my endeavours in every possible way.

C O N T E N T S

CHAPTER 3:

TOOLS & TECHNIQUES FOR MODELLING LEARNING

CHAPTER 4:

APPROACHES TO MODELLING OF LEARNING - PART-I

<u>CHAPTER 5:</u>

APPROACHES TO MODELLING OF LEARNING - PART-II

CHAPTER 6:
DESIGNING A CYBERNETIC 'LEARNING' MODEL

PREFACE

Having embarked on the task of the study of the phenomenon of learning, and in particular its manifestation in machines, with a background in mathematics, electronics and control systems it was inevitable that an analytical mechanistic bias would be the overriding feature of the initial approach. As a medium for the investigation of artificial 'learning' in systems the construction of a state-of-the-art computer remote controlled experimental mobile-robot was undertaken, which was hoped to embody the implementation of a simple learning program that would enable the machine to gradually 'learn' from very fundamental criteria, after a process of interaction with its environment. The method of learning would be diametrically opposite the other major approach to 'machine learning', that of incorporating a lot of a-priori knowledge and information-base into the machine.

It was envisaged that a basic scheme could be devised which transformed the initial random behaviour of the model into a more ordered and purposeful, yet simple, patterns of behaviour. The elaborations of this scheme would, in turn, slowly lead to more interesting behavioural instances; and, indeed, even complex higher order learning capabilities could be displayed by later developments of the basic program. The actual embodiment of the 'learning system' in the physical construct was only an incidental characteristic of the design, chosen for purposes of clarity and ease of experimentation. Hence, the mobile robot itself was designed in a very versatile manner, its input-output features chosen in a non-machine-specific way that would allow its use as a general purpose tool for a variety of experimental modelling tasks.

The ensuing approach was, in fact, consistent with the intuitive 'natural' manifestations of the process of learning. Parallels could be drawn with the evolutionary and hierarchical ordering of learning seen in nature, and the continuity which is observed amongst various levels of learning behaviour (or their underlying mechanisms).

The completion of the hardware model was followed by various software developments which would emulate various simple behavioural patterns. A more elaborate 'learning' program was also planned that in a manner similar to some other research in this area would attempt to realize the 'learning' capability in the machine.

In a survey of the subject of "Machine Learning" it was seen that a great many fundamental aspects are either taken for granted, or various misconceptions and inconsistencies are evident between alternate viewpoints. Many hardware 'learning' models rely too much on the causalities of their physical fabrics, or use 'clever tricks' and engineering solutions to portray aspects of learning or human intelligence. Others incorporate a great deal of designer's own knowledge at the outset, and the inherent learning potential in-built within the model is sometimes misconstrued as "capability to learn from basic principles". Methods used for representation or implementation of 'learning' in artificial systems were also varied. Ranging from physical configurations, verbal descriptions, graphical illustrations, mathematical equations, logical propositions, to elaborate cognitive conceptualizations. The role of an external 'teacher' or 'supervisor' was also a rather prominent one in most 'learning systems', detracting from the autonomy of their 'learning' capabilities.

Hence, it was becoming increasingly evident that a much more fundamental analysis of the problem and a deeper understanding of issues involved was necessary, if we were not to get trapped in a self-propelling dogma of a blinkered and narrow view of the subject. The endeavour of investigating root notions such as teleological or entropic aspects of this class of models, and identification of their 'necessary and sufficient' features was seen as a worthwhile task on its own.

Immersion in a Cybernetics Department was the principal influencing factor that lead to this broad and generalist outlook. Since its introduction, in the 1940's, the science of Cybernetics has made an important contribution to the breadth of scientific thinking. Although, at times, suffering from the generality of its concepts, nevertheless, cybernetics has been able to put forward some specific methodologies for solving problems in more narrowly defined areas.

Cybernetics attempts to unify studies of living and inanimate purposive systems, and in its true spirit it would have been appropriate to have a methodology which could be equally applicable to both domains. A general cybernetic formalism on par with those in other well established sciences would allow the analysis of natural phenomena in the context of artificial systems. Indeed, if such a hypothetical formalism was forthcoming, then learning could have easily been manifested in systems. Yet, the complexity of the task of devising an abstract cohesive language is such that attempts so far have been grossly inadequate, and have only managed to highlight the underlying difficulties involved in such endeavours.

However, so far, the principal strength of cybernetics has not been in its techniques, but in its promotion of a fresh and expansive outlook or approach to problems. The broadness of perspective means that prejudices which sometimes hinder developments in other fields have been, generally, absent in cybernetic enquiries.

Hence, in this cybernetic tradition, it was deemed necessary that before we attempt to tackle our specific objective three main areas had to be scrutinized in more detail. Firstly, the natural learning phenomenon in its many forms should be looked at, and a variety of related issues addressed. Secondly, the characteristic problems involved in modelling a natural observation should be understood, and tools and techniques used for such modelling tasks outlined. Thirdly, a general investigation of previous attempts at modelling of learning should be made, to highlight relative merits, findings, or typical difficulties confronted in each paradigm.

As the above undertakings were being pursued the multiplicity of facets of the phenomenon of learning and its studies were being much more appreciated; and, at the same time, the vulnerabilities of a narrow approach to its simulation or synthesis becoming more prominent. Each field of study would lead to other subtopics or sublevels, and the interlinkings of issues and subjects was gradually compiling a very complex multidimensional picture. All aspects of description, investigation, analysis, simulation and synthesis of learning were characterized by a hierarchy which spanned from the system or behavioural levels to the molecular level.

The task of studying all elements of "learning" is an enormous one, and for this reason each segment of the composite picture is scrutinised within a predominant specialised discipline. This singularity of approach (without generally paying too much attention to alternate subjects) is seen in almost all literature on learning, even those which take on a multi-disciplinarian stance, normally, view the subject through lenses tinted by a particular bias. A variety of subjects such as Behavioural Psychology, Cognitive Psychology, Neuro-physiology, Education, Sociology, Cybernetics, Pattern-Recognition, Control Systems, Artificial Intelligence, etc. deal with learning; each having its own conception of learning and the relative importance of issues involved.

Hence, it was decided that as a joint objective of the thesis an overview of the various aspects of study of the learning process and its mechanisms be undertaken. Principally, the results of this non-standard approach will be of benefit to a prospective designer of a 'learning' model. Whereby, a very broad awareness of many topics, issues, and similar research can be attained, without cluttering the totality of picture with too detailed specifics of each paradigm. The aim is, however, not so much to try to unify the whole subject but to point to the particularities and discrepancies of diverse view, and the way they deal with various issues in learning.

The level at which each topic will be covered is, generally, governed by the its relevance to the second more specific objective of this thesis, namely, the considerations in designing a cybernetic simple learning model. At times, a summarized tabulation, classification, or definition of specific features of an approach will be made without much qualification; yet, at other times, a more detailed analysis or discussion of subtopics of a discipline will be attempted. The intention of this particular form of examination is to furnish an informative coverage of principal keynotes, trends, and points of contention involved in the variety of learning related disciplines which might have direct or indirect bearing to our specific problem. Additionally, by looking at the underlying historical and developmental aspects of subjects, attention could be drawn to their kinships and their commonalities of purpose.

All in all, it is hoped that the insight gained, and the deeper (and broader) understanding of the whole subject, from the conglomeration of the general and the specific pursuits of this thesis will provide a firm platform for later research into the much more difficult (and worthwhile) task of devising a true 'learning' model, which in itself is a life-time long scientific quest.

CHAPTER 1

THE NATURE OF THE PROBLEM

1.0 INTRODUCTION

In this introductory chapter we will attempt to prepare the background for our later more detailed analysis of various aspects of the learning process and its modelling. The schematic illustration of FIG.1.1 outlines the principal approaches and topics of interest to the phenomenon of learning, and shows the variety of facets to this multi-dimensional concept.

However, the representation of FIG.1.1 is not meant to be very precise or exhaustive, and research in many other disciplines, directly or indirectly, bear relevance to the process of learning - similarly, the inter-linkings of disciplines are not illustrated in FIG.1.1.

The approach of the initial part of this thesis will be a very broad 'global' one. Yet, at the latter parts, this generality will culminate into a more focussed leaning towards the simpler modalities of learning, in particular, its simple manifestations in models. Indeed, it was the design and construction of such a model which prompted this sweeping approach to the thesis.

We will begin by defining and analysing the nature of the problem. After a general discussion of the notion of learning the developmental and other aspects of learning (i.e., historical, philosophical, operational and technical) will be outlined. Then, more specific issues and problems of manifestations of learning in various models, and approaches to such endeavours, will be looked at in more detail. In particular, the case for building hardware 'learning' models will be argued. Finally, the general and specific objectives of the thesis, and the impetus behind the direction of its particular developmental path, will be discussed, and an overview of the chapters described.

1.1 WHAT IS LEARNING ?

Learning and adaptation are central criteria of life, influencing almost all aspects of human behaviour and many of animals' behaviour. The concept of "learning" in its intuitive human-learning sense is an age old notion, referring to the way that humans and animals increase their knowledge and improve their skills. But, the modern comprehension of this multi-faceted concept is a complex and a different one, referring to many other connected aspects, and implications of usage of the verb 'to learn' are far greater than its established common understanding would suggest.

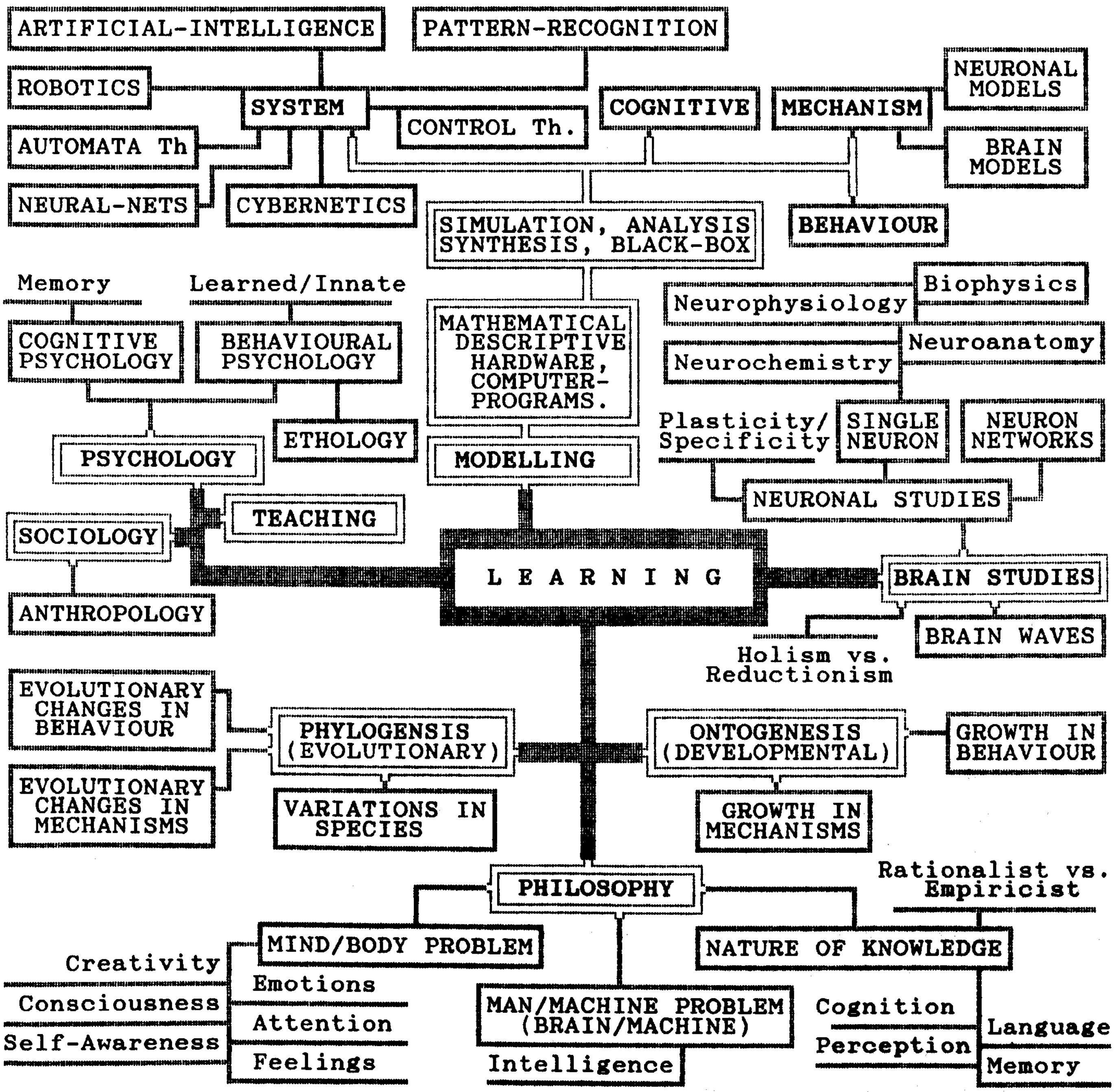


FIGURE 1.1. A schematic outline of "LEARNING" related disciplines, their domains of interest, and some major phenomenological issues or topics.

Some researchers have suggested that "learning" is too general a label to be used for such diverse range of activities and processes. In this thesis, also, it was found that a distinction was necessary, hence, we will follow a notation of enclosing the term LEARNING in single quote marks (' ') when it refers to its 'artificial' aspects, or its manifestation in a specific non-biological system (e.g., machine 'learning', a 'learning' robot). Conversely, without the quote marks when it is expressed in the global sense, or within its natural context (e.g., the learning process/phenomenon, animal learning).

In spite of the difficulty of devising an adequate definition of learning, which can cover all its manifestations, it is surprising that there is not much controversy or disagreement over the definition of "learning" itself. The main problems arise in the interpretations of such definitions. Later, various definitions of the learning process will be expanded within each discipline. But, an all encompassing simple expression of learning could be:

"TO PROFIT BY PAST EXPERIENCE".

An explicit premise of our extensive approach to learning is that the process of learning should not be regarded as an exclusive attribute of man and animals, and a machine or an abstract mathematical system satisfying various criteria, given the right interactive circumstances, should also be considered as a 'learning' entity.

The study of the learning process, as in many other brain related phenomena, seems to have three general facets: 'epistemological' (knowledge related) aspects, 'operational' (acquisitional or executive) aspects, and issues and problems involving both epistemological and operational aspects.

Epistemology is defined as the theory of knowledge, and knowledge itself can be regarded as a product of the process of learning (in its broad sense). Therefore, generally knowledge and learning are regarded as intrinsically related. Hence, epistemological considerations should always accompany the investigations of learning. Of course, in the lower forms of learning, such as reflex learning, epistemological aspects are of a trivial nature.

The term 'adaptation' is sometimes defined as: modification of behaviour due to negative feedback from the environment. But, in this thesis adaptation will be looked at as a simple manifestation of the process of learning.

1.2 A BRIEF HISTORICAL BACKGROUND

The illusive nature of human 'thought' and its attributes (e.g., learning) has intrigued man for thousands of years, and problems of trying to explain the mental and the behavioural characteristics of humans (and animals) have been in the forefront of science of philosophy, as well as occupying workers from many other disciplines such as mathematics.

But, generally it is regarded that a systematic approach to these problems only began in earnest from the 17-th century, when the physical studies of the surrounding world were starting to explain some of the dynamics of the inanimate phenomena. Galileo was one of the principal thinkers who attempted to explain the physical world in terms of mathematical relationships, he believed that mathematics was the 'language' of physical universe.

Some philosophers and mathematicians such as Hobbes had started looking at human thinking processes in terms of physical and mathematical abstractions, and with the introduction of Newtonian physics the mechanization of thoughts truly began its course of development.

It was Descartes, the famous French philosopher/mathematician, who in his attempts at unifying sciences explicitly addressed the mind/body problem, and developed a theory of mind. He regarded thoughts as symbolic representations which could be fundamentally equated with mathematics, but, he was also a 'dualist' who saw the mental and physical universes as separate and non connected. Descarte's conclusion was that mind and body were distinct phenomena, and hence divided human activities into two groups of 'mechanical' and 'natural'. Animals were declared as wonderful machines, yet human beings were considered as machines with minds in the form of a 'soul' imbedded deep within the brain.

Although, the notion of humans as machines was later argued by many scientists and was common place, such as in famous 18-th century philosopher Hume's concepts of "mental mechanics", a widespread belief in a discontinuity between animal and human intelligence persisted well into the 19-th century, until the introduction of Darwinian evolutionary doctrine.

During the 17-th and the 18-th century many attempts were also made to copy various aspects of human and animal actions, by constructing simple mechanical automata or devising models of mental processes. Some abstract models were based on the contemporary theories of mental phenomena, but the majority of these models were artifacts built by skilled craftsmen and watchmakers to demonstrate intricate human like (or animal like) movements.

During the 19-th century many other elaborate walking or even talking machines and automata based on gears, pulleys, levers and bellows were also constructed; some in the latter part of the century were using electromagnetic components in their models. Similarly, game-playing machines were devised, the most advanced of which could play end-games of chess quite cleverly.

In parallel to the introduction of such wondrous constructs developments were also being made in the field of design of calculating machines. A principal figure of the era was Babbage, who is sometimes considered as the father of computer-sciences. He conceived the idea of "analytical engine", and although did not fully realize his ambition of constructing such a machine, his conceptualization, together with the development of an algebraic foundation for binary symbols and postulates by Boole, prepared the groundwork for future computing machines.

The 'mind/body' problem, stated simply as: "How the mind and body affect each other", also gave rise to another duality of approach, the so called 'empiricism' versus 'rationalism'. The followers of the doctrine of empiricism, including philosophers such as Hobbes and Hume, believe that experiences and their associations are the only source of knowledge. They also subscribe to the 'mechanistic' view point whose thesis is as follows: "Mind is like a machine built from simple non-vitalistic elements". The rationalists, on the other hand, believe that reason and relation of concepts are the basis of knowledge, belief and action.

At the latter stages of the 19-th century and the early 20-th century, the search for mathematical counterparts of reasoning and thinking was still a principal pursuit of philosophers and mathematicians. Whitehead and Russell's studies of knowledge and logic, and speculations into the mathematical basis of abstraction, culminating in "Principa Mathematica" was a landmark achievement in the attempts at unification of fundamental issues of reasoning, logic and mathematics. During the same period, the theoretical and methodological foundations of modern trends of the science of psychology was also being established. The theories of learning were developed as a subsection of psychology in two distinct major camps:-

(1) - The 'associationist'/'stimulus-response'/ 'behaviourist': whose followers subscribe to the empiricist point of view, and included many of the experimental psychologists, namely pioneering adherents such as Thorndike and Pavlov.

(2) - The 'cognitivist': who are from the rationalist school interested in organizational aspects of human activity and reject the mechanistic doctrine, the principal protagonists were Gestalt psychologists such as Köhler and Tolman.

The mass of experimental results obtained about learning processes, and the theories proposed in various related fields, were incorporated in models which could either 'simulate' the results of an experiment or 'synthesis' a proposed postulate about learning, most such models were mathematical abstractions. It was the introduction of the science of Cybernetics by Wiener, and its explicit declaration of man-machine equation, which signified the start of the modern venture into the design of 'learning machines'. Cybernetics was the catalyst which promoted the use of well established mathematical abstractions to the problems of modelling of aspects of living organisms in machines.

A yearning for mechanization, in the tradition passed on from Descarte's time, compelled cyberneticians such as Ashby and many other system theorists to formalise behaviour and thought on basis of mathematical concepts; and later try to manifest such concepts in hardware models. A notable contribution was also made by the introduction of the 'information theory' by Shanon, which allowed the quantification of some previously vague concepts. As well as the modelling of 'learning' behaviour the work of some cybernetically oriented neuro-physiologists and mathematicians, namely McCullock and Pitts, was the prelude to a whole generation of neural-network type cellular 'learning' models.

The next significant development in this field was the advent of the digital computer, which without any doubt has proved to be the most important single innovation for the modelers of the learning process. The infrastructure for today's computers was laid over a century ago by the formalisms of Boolean algebra and the designs of calculating machinery, such as Babbage's "analytical engine".

However, the modern theory of computation owes a great deal to Turing's abstract analysis of computational machinery, in particular his conception of ideas of "Turing Machines" and "Universal Turing Machines". These simple abstractions were made of two basic components of a 'read/write head' and an 'infinite-tape', which were constructed for answering theoretical questions rather than for use in practical problems. Yet, they were influential in later developments of digital computers by other pioneering workers, such as von-Neumann who greatly contributed to the design of practical general purpose computers.

Turing's abstract machines were not suitable for representing human behaviours or mental processes, yet, they were able to mathematically prove an important assertion, which stated that if behaviour, language, or thought could be specified in a formal sense then all human activities could be represented by a Turing machine.

The most recent approach to 'machine learning' was established in the early 1960's with the widespread availability of digital computers. This so called 'information processing' approach is the dominant paradigm of today. The subject of "Artificial-Intelligence" or A.I. (a name coined by one of its pioneers, McCarthy) is the principal domain for research on the modelling of learning today; and it predominantly relies on the use of digital computers. Of course, during the past three decades, other computer structures were also introduced, to deal with a variety of, mainly cognitive, modelling tasks (e.g., parallel, analogue, LISP, production-systems).

1.3 PHILOSOPHICAL, GROWTH AND EVOLUTIONARY ISSUES OF LEARNING

(i) - Philosophy

As we have seen the psychological theories of learning developed from centuries of philosophical enquiry into the problem of mind/body distinction. Other philosophical introspections into issues such as 'mind', 'self-reference', 'consciousness', or 'awareness' have resulted in hypotheses about knowledge acquisition, or motivational aspects of behaviour; and have influenced the contemporary psychological thinking.

It is, therefore, hardly surprising that when conceptual descriptive terminologies of human mental states and functions are applied to alternate domains of machines, computers or abstract systems then so much controversy arises. Concepts such as 'consciousness', 'perception', 'cognition', 'learning', 'creativity', 'memory', 'thinking', 'feeling', 'awareness', 'emotion', 'attention', 'intelligence', 'free-will' are to name but a few. These concepts have been intuitively used for centuries without much attention to their underlying functional or biological basis. Hence, the explorations of 'machine' or 'mathematical' counterparts of these concepts are at best speculative; and loosely analogous to trying to name components of an automobile engine or define their functions, in terms of 'muscular' mechanisms.

Nevertheless, it is fair to say that philosophical scrutinies into the mechanization of these ideas, besides providing interesting debating topics, have shed light on the nature of many previously opaque concepts; and have allowed hypothesising about their integration. In some cases, however, due to the incompatibility of levels of descriptions involved, results of the translations of human concepts to machine domain have been of a very non–conclusive nature.

(ii) – Growth

The growth and development of an adult animal from a single egg–cell is governed by the blue–print provided in the genetic information within the cell. The development and changes of the brain and the nervous–system can be seen throughout the life of an organism, the crucial stages being the prenatal and early childhood periods. The learning capabilities and behavioural changes which accompany such neuronal changes are, hence, of interest to our wide perspective of this phenomenon.

(iii) – Evolution

Similarly, from an evolutionary perspective it can be seen that the learning process has gradually developed from the simple response mechanisms of uni–cellular primitive organisms to the most complex conceptual capabilities of man, which even allows him to introspect about the very nature of such processes and their mechanisms.

Hence, perhaps, the attempts of those modelers who only concentrate on the higher strata of learning hierarchy should be seen as a disregard for the clear example provided by the 'natural' modeler of this process (i.e., the natural selection). Demonstrating (in some 3.5 billion years) that the development of a conceptualising organism requires a progression of steps, from very simple to highly complex and rich in knowledge, both in its ontogenetic and phylogenetic ordering.

1.4 PSYCHOLOGY, NEURO–PHYSIOLOGY AND LEARNING

The studies of learning in its natural domain have a great deal of relevance to the 'artificial' modelers of this phenomenon. The well documented empirical observations of learning in psychology have provided many insights into the nature of learned behaviour, and also the organizational aspects of its structure. Many definitions, attributes and functions of different classes of learning behaviours have also been characterised. In addition, the studies of the brain and neuronal mechanisms have revealed illuminating underlying features of learning processes.

Today, it is unanimously accepted in learning related scientific fields that mental phenomena are the consequence of underlying physiological structures of the brain. This mechanization of cerebral functions is, also, being more and more validated by evidence from developments in sciences of neuro-physiology and psychology.

The interaction of 'analytic' and 'syntactic' aspects of studies of brain phenomena have provided some of the most challenging scientific questions of our time. Depending on the era (and fashionability of various paradigms) the brain has been described as: "a network of canals", "a water clock", "a mechanical clock", "a telephone exchange", and most recently "a digital computer".

The complete understanding of inner workings of the brain is a very difficult (if not impossible) task by the very nature of the problem. Hence, this problem has been tackled either by the studying, the theorizing and the manipulation of inputs and outputs to the brain (and its elements); or by the construction of organizational 'models' of neuronal mechanisms, and the studying of model's input-output relationships.

A great volume of experimental results has been amassed in psychology on the various manifestations of the process of learning in humans and animals. Augmenting these evidence is the rich source of findings of 'cognitive scientists' on the structural knowledge based aspects of the learning process, and also the investigations of 'neuro-physiologists' and 'physiological psychologists' on the learning related mechanisms of the central nervous-systems.

Some very precise scientific techniques have been developed for the rigorous recording and analysis of these experimental results, both in qualitative and quantitative terms. Yet, there is no general consensus about the explanations of these findings. Theories postulated within one field, generally have no direct relevance to other levels of investigation; and even within a specific line of approach many qualifications have to be made about the experimental observations before theorising. Therefore, for a given set of evidence, it is not only sufficient to propose a theory, but the reasons for putting forward such a theory should be justified, and the scope of its applications specified.

The above is a fundamental problem in science, typical of complex phenomena, such as learning, with many levels of descriptions. However, there are three principal ways of describing the learning phenomenon:-

(1) - 'Behavioural' or 'Generalising' descriptions which involve classifications of equivalent events and discovery of behavioural causalities.

(2) - 'Neuro-physiological' descriptions of mechanisms involved in learning.

(3) - 'Cognitive', 'Organizational' or 'System' descriptions which entail devising systems whose properties match with the observations of behaviour or mechanism.

It must be pointed out that mathematical abstractions could be applied to any of the above three levels of description; and such formalisms should only be regarded as a language for expression, and not a separate descriptive level.

Today, an obvious gap exists between neuro-physiologists or psychologists and designers of 'artificial learning systems', which is hardly surprising if we consider their differing backgrounds, techniques, skills, or terminologies used.

It is interesting to look at the significance and the contribution of various artificial 'learning' models to the natural domains of the study of learning phenomenon, for example, psychology, neuro-physiology, teaching, or ontogenetic and phylogenetic sciences. The views of the workers from such disciplines are indeed diverse, they range from skeptical repudiation, dismissals as only interesting diversions; to serious enthusiasm, and utilization of these 'artificial learning' models for purposes of simulation or better understanding of their own criteria.

Although, many 'non-biological' assumptions are made in the design of most 'learning' models, nevertheless, positive contributions have clearly been made to the studies of learning in its natural domains. For example, 'artificial' neuronal models and networks are commonly used in neuro-physiology for better understanding of neuronal plasticity; also, in the field of A.I. various discoveries have been made about the perceptual and linguistic aspects of symbolic knowledge.

1.5 NATURAL vs. ARTIFICIAL ASPECTS OF 'LEARNING' MODELS

The fundamental question of whether artificial systems (software and hardware) can in fact 'learn' has preoccupied many workers involved in the modelling of learning processes, and is a prime consideration in most A.I. and cognitive modeler's research. The basic issue hinges on the definition of human (and animal) 'intelligence'. In observing any natural learning behaviour three major questions are posed:-

(1) - What are the mechanisms involved?
(2) - What are the components of learning?
(3) - How does learning change with time?

Each of above questions have been partially answered at some level of observation in neuro-physiological and psychological sciences. Yet, the principal problem is that the manifestation of the learning process in man (and in animals) is not clearly understood. When a particular learning behaviour is scrutinised then a linguistic description or a mathematical quantitative evaluation of performance can be made; similarly, some aspects of neuronal changes can be observed. But, the hypotheses compiled are still very incomplete and many ambiguities have yet to be resolved.

Putting aside the vitalistic objections, arguments against artificial systems being able to 'learn' are, generally, either based on their superficiality due to mimicry/duplication, or use entropic considerations (i.e., order is only possible from order - a machine will only do what it is told or designed to do). Yet, it is a truism to say that many of the 'learning' models devised so far have been able to display novel behaviours which in some cases have even surprised their designers. Others, also, in an 'open' interactive manner have 'learned' to perform specific tasks with a proficiency which at times surpasses that of a human. In any case, the daunting question always remains present whether the manifestation of 'learning' in such an artificial model is functionally equivalent to the human learning process.

Turing (1947) foresaw that artificial systems could be made to manifest learning in human sense, and proposed that with technological development of larger computing machines experiments along these lines. The following Turing's criteria for 'learning' in machines can be considered on an equal footing to his now renowned "Turing test for 'intelligence'" (which will also be discussed later):-

> "Let us suppose that we have set up a machine with initial instruction tables, so constructed that these tables might on occasion, if good reason arose, modify these tables. One can imagine that after the machine had been operating for some time, the instructions would have been altered out of recognition, but nevertheless still be such that one would have to admit that the machine was still doing very worthwhile calculations. Possibly it might still be getting results of the type desired when the machine was set up, but in a much more efficient manner. In such a case one could have to admit that the program of the machine had not been foreseen when its original instructions were put in. It would be like a pupil who had learnt much from his master, but had added much more by his own work. When this happens I feel that one is obliged to regard the machine as showing intelligence."

Turing (1950), also, in a very knowledgeable, and historically significant, analysis of the problem of machine 'intelligence' again discusses the possibility of designing machines that can 'learn':-

> "Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulate the child'? If this were then subjected to an appropriate course of education one would obtain the adult brain."

Turing looks at the adult's brain as starting from an initial 'state of mind' at birth and going through a process of 'education' and 'experience'. The importance of the issues of 'reward' and 'punishment' (not in the emotional sense) are also emphasised in such a 'learning child-machine'.

An important analogy is also made by Turing in comparing the process of learning with the process of evolution. 'Natural selection' is seen as a very slow learning process which is governed by changes brought about by random 'mutations' - as opposed to changes brought about by 'motivational' or 'reinforcing' aspects of the process of learning. This 'random' element is regarded as a useful feature which could be utilized in some machine related 'learning' strategies; in particular, in non-trivial problems that have large combinatorial characteristics, where other systematic solutions are inadequate.

There is a well investigated and documented empirical observation in psychology, the so called 'superstitious behaviour', which shows a tendency in humans and animals to attribute order and organization to events which are totally random. Another aspect of this feature of human psyche, evident from the dawn of civilization, has been the desire to assign purpose and intelligence to inanimate phenomena surrounding us. Therefore, it is hardly surprising that machines or systems that seemingly act 'intelligent' are so readily ascribed with human intellectual capabilities.

In this respect, we can make the interesting observation that almost invariably one of the first considerations in the design of a laboratory robot, or an A.I. computer program, is the personalising of the model with a human sounding name (e.g., FREDDY, SHAKEY, HACKER, etc.); and these machines or programs are often referred to as "he" and not "it".

The proficient A.I. programs of today, such as chess playing programs, would have unquestionably been considered as 'intelligent' a generation ago. But, the tendency is to dismiss a behaviour as 'intelligent' once it is broken down to its logical components, examined, and a mathematical formalism abstracted for its synthesis.

One of the important ways which the more recent models of the learning process vary from their biological counterparts is in their widespread reliance on a 'teacher' or 'operator' element. Most learning in nature is attained by direct interaction and experience with environment, with an 'internal' evaluation of experiences prompted by 'goals', 'motives' or 'drives'. The earlier cybernetic 'learning' models did try to copy this aspect of natural learning, yet, were only capable of performing very simple tasks.

The various 'teleological' explanations as to why humans and animals should posses the learning capability have ranged from "entropic balance", "evolutionary adaptiveness", "survival", "biological control", "preservation of vital parameters", "purposiveness", "divinity" to many others. Similarly, such considerations are thought to be pivotal in the design of complex 'learning' systems.

It is conceivable that a genuinely autonomous machine with 'learning' capacity may, after its interactions with the environment, develope a notion of 'purpose' (in the abstract sense). Yet, still at the more fundamental level of design other internal teleological centers may be incorporated, which enables the machine itself to direct its behaviour, rather than to exclusively rely on the external judgments of its designer/teacher/operator/supervisor - examples of this type of primary design principle are 'seeking equilibriums' or 'seeking change'.

There are also many instances where the psychological and biological aspects of 'life' are too frivolously applied to machines, robots, computers or even abstract models. One extreme example is when some writers contend that machine-tools or robots are 'reproducing' themselves when they are engaged in the construction of other machine-tools or robots. These conceptualizations should be regarded with some degree of skepticism, and also care should be taken in the choice of 'natural' concepts when applied to various aspects of an inanimate design, alternately their definition should be qualified for such artificial domains.

So we have seen that many problems arise when various terminologies based on human attributes are scrutinised within disciplines involved with the modelling of the learning process. Contradictions, overlappings, and ambiguities between different terms become hindering stumbling blocks, resulting in much effort being expended towards arguing points of distinction. Concepts of 'intelligence', 'perception', 'cognition', 'memory', 'understanding', 'thinking', or 'knowledge' are some of the historically more controversial such notions. So it is understandable that many workers have a striving for the relative safety of mathematical descriptive methods where, generally, terminologies refer to precise concepts within the axiomatic framework of the particular formalization.

Formalising in general means changing the universe of discourse (or the language) by use of some translating rules. It is hoped that greater precision could be achieved in the new body of discourse. Yet, in the case of the learning process a lack of sufficient knowledge about underlying 'syntactic' and 'semantic' rules has meant that the translations to the new formalised domains have, generally, been in a very limited and narrow sense.

A recent trend in the modelling of 'mental' processes has been to 'externalise' the particular capabilities of the brain from its natural (neuronal) mechanisms. 'Language' and 'vision' are some other attributes of the brain which have been studied and theorized in this vain (Chomsky and Marr respectively), by inherents of their logical processes, rather than structural details of entities embodying these processes.

1.6 PARALLEL/SEQUENTIAL, ANALOGUE/DISCRETE ASPECTS AND LEARNING MODELS

In system theory various attributes of general systems are studied and their properties compared. A principal recurring argument of this nature which has been applied against the use of computers for the modelling of the learning process is their 'sequential' rather than 'parallel' character. It is contended that since nature, through the course of evolution, has equipped animals with a brain which essentially works in a parallel manner, then it is only logical that the same principle should be incorporated in the design of tools intended for the modelling of the brain's processes.

Partly for this reason, and partly for other technological or commercial considerations, the development of parallel processors and parallel computers have again become focus of much attention, and subject to much activity in recent years.

Considerations have also been given to the issue of 'digital' vs. 'analogue'. All human and animal primary senses and the changes in environment are by nature continuous (or analogue), the discrete (or digital) interpretations of such changes are only attributed at the higher symbolic representational levels of the brain. Even, the so called on-off action of the neurons is only a 'description' of their real activity which is continuous.

However, it can be mathematically shown that any analogue process may be approximated by an equivalent digital process, to whatever degree of accuracy required – by a process of 'sampling'. Examples of such an approximation abound in our verbal descriptions of surrounding world (e.g., hot, cold, dark, etc.).

Most of the early 'learning' models were based on analogue criteria, of course, they were attempts at the simulation of lower non-symbolic levels of the learning hierarchy. Some workers have also contended that the continuities of the environment can only be adequately represented by analogue systems. They see that changes of the real world, and continuities of perceptions of such changes, carry significances which are lost when the same changes are recorded and stored in discrete unconnected units. Hence, the digital computer is considered as a poor choice for modelling fast analogue changes of the world which implicitly carry a vast amount of 'bonding' information (whose preservation is deemed crucial).

So far the fundamental necessity of analogue processes in the task of simulation or synthesis of 'learning' has not been conclusively demonstrated, and with the development of much faster parallel digital computing structures, the arguments of relative merits of digital vs. analogue should be switching from the 'computing device' domains into issues of principle.

1.7 MODELLING OF LEARNING

Synthesis of a model provides grounds for further testing of a theory or its modifications. Normally, if theories are simple then deductions can be made either by verbal descriptions, simple hardware, symbolic logic, or mathematics. But, for the more complex theories linguistic descriptions become very tedious; and hardware, logical or mathematical descriptions will be too difficult to devise, or will be highly erroneous. It is precisely for this class of theories which a recourse has been made to computers – particularly, in explanations of higher mental activities of humans.

If a model is capable of exhibiting behaviour which functionally is a very faithful duplication of the phenomenon it is trying to simulate, then it can be said that the model has a close equivalence to the original. Additionally, if there were also close physical resemblances between the model and the original (i.e., texture, size, colour, etc.), then it is conceivable that the model can replace the original without noticeable distinctions.

Various mathematical equations and other forms of parametric abstractions have been devised to support different learning theories or specific learning observations. The techniques used in such abstractions utilize a variety of mathematical criteria, such as 'probabilistic' or 'set-theoretic'.

Yet, the fact that the process of learning can be interpreted in so many levels has led various researchers to express the theories of learning in terms of simple models which are more easily understood than complex mathematical abstractions of the observations, or their mere linguistic explanation. 'Learning systems' are realized as models in three basic ways: mathematical abstractions, computer programs, and specially built hardware devices.

The hardware models have the added advantage of incorporating the causalities of the physical environment within their design. But, most of all, hardware models are advantageous because of their practical considerations, and the clear insight they provide into the relation of structure and behaviour.

A well conceived hardware model can aid explanation as much as a well formulated abstract theory. Getting a physical model such as a robot or a machine to function properly is equivalent to demonstrating that the theory is internally consistent; and running the robot/machine through several tasks is logically equivalent to hypothesising about a particular explanation. Therefore, irrespective of the means used for realization (hardware or abstract) theories can be functionally equivalent. In other words, there is an 'isomorphism' (point-by-point parallel correspondence) between the behavioural patterns of such realizations.

Later, it will be seen that most 'learning' models, except perhaps the designs used for simulation of lower modalities of learning (e.g., reflexive behaviour, neural plasticity), have generally two basic design features - reflecting the distinction of human 'action' from 'thought'.

Firstly, the 'behavioural' features of the model, and secondly the 'information-processing' aspects incorporated within the model. While, the 'performance' of a model is judged from its behavioural features, the acquisition of knowledge and all 'learning' is centred around its information-processing faculties.

A third higher descriptive feature, as in the case of humans, can also be incorporated within a 'learning' model, and that is the capability to 'abstract' or 'model' an internal isomorphic representation of the external world (and the model itself). Whereby, inductive inferences can be made about the eventualities of the models behaviour without physically engaging in such behaviour. Generally, it is regarded that for this facilitation a 'language' capacity is necessary, to enable the abstraction and extrapolation of syntaxes within input data. Here, the philosophical issues of 'self' and 'self-awareness' could find scope for exercise.

'Learning' models could be categorized according to a variety of criteria (e.g., historical ordering, type of approach, type of data acquired, or domain of application). In an intuitive hierarchic sense, probably the simplest type of 'learning' models are those which demonstrate basic reflex conditioning behaviours or other innate features of animals. Next in the line of hierarchy are the models which use the reinforcements of favourable responses as their basis, and generally involve statistical criteria. Such models can be regarded as some kind of 'generalizing' or 'averaging' devices. Their elaboration, in turn, could entail adding secondary reinforcement centres which can determine the 'good' features that were before determined by an external 'trainer' or 'teacher'.

The 'concept learning' models which utilize logical representations of external world rather than their numeric or statistical features are, yet, higher in rank. Finally, the most complex types of models are those incorporating a broad spectrum of intricate learning methods, mostly based on knowledge intensive systems directed towards the acquisition of a specific skill, and involving the higher strata of human learning capabilities.

When various models of 'learning' are evaluated or discussed we must also consider whether they are "special purpose" or "general purpose" models, referring to their relative range of applicability. Many other considerations will also be outlined in later chapters.

The developments in the field of modelling of learning have been characterised by over optimistic claims of its researchers. Yet, failed overeager predictions of cyberneticians and A.I. workers are not the kind that disprove the whole theory, but only shows weaknesses in the choice of its direction or pace of development.

In the 1950's and the 1960's, such modelers having analyzed the problem and proposed a blue-print thought that it would only be a matter of time, and improvement of appropriate technological tools, before their simple systems could be elaborated into fully fledged sophisticated 'learning machines'. This scientific over-expectation is typical of many disciplines, and does not imply that the workers involved have anything but the best intentions.

The advent of digital computers with their ever increasing computing powers (faster and with more memory), although predicted by the early modelers of the learning process, did not provide the revolutionary breakthrough expected. The aspiration of pioneers of cybernetics for a general purpose computational machine, with nearly unlimited potential, which can be used for the simulation and synthesis of the learning process is almost realised today in the shape of digital computers. But, in spite of these powerful tools, solid and successful global 'learning' theories which can be applied to different tasks and systems have not been forthcoming. The recurring theme has, therefore, been an under estimation of computer capabilities, and an over estimation of its potentialities. In a sense, we can say that theoretical advances have not been keeping pace with technological innovations.

1.8 DIFFERENT APPROACHES TO THE PROBLEM OF MODELLING OF LEARNING

Some early designers of 'learning' systems had anticipated the two developmental paths that would be taken by researchers in the field of 'machine learning'. Which were the two diametric strategies of building-in (or programming-in) a 'lot' or 'little' complexity into the initial state of the system. They contended that, on the one hand, complex logical inferences could be incorporated at the initial state of the machine, hence, the pattern of the 'intelligent' behaviour would be present from the start. Or, on the other hand, by incorporating just a minimal amount of information and a capability to 'learn' the machine would be able to develope the 'intelligent' behaviour gradually itself. The actual development of 'learning' models within the past four decades has traversed quite an interesting (and logical) path, which closely emulates the general pattern of these early predictions.

As it will become evident from our later discussions of various approaches to the modelling of the learning process, and from many examples cited, the notion of learning played a much more prominent role in the earlier endeavours of designers of 'intelligent' machinery or systems.

Following the proliferation of the theories of learning, the development of 'learning' models generally progressed on a parallel lines to such theories, and mainly incorporated various 'natural' criteria in their designs. These psychological models were principally used for the simulation and better understanding of the theories of learning.

Later, such challenges were taken up by cyberneticians, control-engineers, automata theorists, or general systems mathematicians, who used many more abstractions in their models. Similarly, other 'intelligent' models were developed in subjects such as Neural-Nets, Logical-Nets, Self-Organising Systems, and Pattern-Recognition which also featured many aspects of the process of learning.

The advent of digital computers was a significant landmark. One group of modelers continued on the path of simulating natural learning processes, using these newly acquired tools and techniques. However, a substantial proportion of the followers of the new 'information processing' paradigm decided that the constraint of adhering only to the 'natural' aspects of learning was too limiting, specially since progress was slow, and was dictated by the pace of progress in neuro-physiological and psychological sciences. Hence, they embarked on designing 'cognitive' computer-oriented 'learning systems' using various 'ad-hoc' techniques, and abandoning the emphasis and constant references to 'natural' aspects of learning. They contended that their methods could achieve the end result pattern of a 'learned' behaviour in a much faster and more efficient manner - particularly, in specific task domains.

The workers in this paradigm try to incorporate a great deal of prior 'information' and 'knowledge' into the initial structure of the system. Yet, generally, not much attention is paid to the 'acquisitional' aspects of the design; also any 'learning' mentioned normally refers to a specific problematic domain, and not to its 'global' sense.

Today, all the above approaches are still being pursued in parallel with varying scales of followings and popularity. Adherents of some rejecting the methodologies or even the viability of some alternate approaches, yet, others seemingly adopt a more accommodating and tolerant views of disparate approaches.

In any case, intellectual feuds between followers of opposing paradigms seemingly resulting from argumentation of relative merits of each approach is not a helpful pursuit. Specially, in view of many precedences in the history of science, where alternate approaches have coexisted in parallel until conclusive settlement (e.g., quantum theory vs. wave theory) - although, each being argued with fervour and believed to be true by its followers.

Nevertheless, to find the real weaknesses of a discipline or approach it is best to listen to informed counter arguments of its critics. The frailties of opposite views are invariably exposed more, and the relative merits of their own discipline are highlighted out of proportion. Therefore, to get a balanced view, the two (or maybe more) accounts of the same story should at least be heard.

It is a prevailing belief of workers in fields such as A.I., pattern-recognition, robotics, cybernetics, computer sciences, etc., that within the next 40-50 years there will be technical possibilities of constructing machines with intellectual capabilities of an adult human. Although, the scenarios leading to such a development vary tremendously, ranging from an optimistic utilizations of such machines to their doomsday type dominations. Here, an implicit assumption is also made that parallel discoveries in brain sciences and psychology will make new and significant contributions to this pursuit of 'intelligent' machinery. An important postscript to above conviction is that if human intelligence can be attained then it surely can be surpassed, and the next level of speculation will yet bring to surface the enormous possibilities which 'super-intelligent' developments will entail.

In the following we will attempt to elaborate on the two resulting principal approaches which have come to dominate the field of modelling of 'learning'.

1.8.1 THE CYBERNETIC "BOTTOM-UP" APPROACH

The usage of term "cybernetic" is not a universally exercised practice. Various other researchers refer to this distinct approach as "brain studies", "natural", "neural-nets", "Self-Organising Systems", or "pattern-recognition" approach. But, the appropriateness of using the 'cybernetic' label in this context will be justified in our later discussions of the subject of Cybernetics.

A point of relevance, here, is that the wide perspective which this thesis is advocating has been a characteristic of the cybernetic view of most disciplines. George (1986), in his discussion of the development of cybernetics, A.I. and related subjects, highlights the issues involved in the diversification of a field of knowledge. He contends that distinctions in approach (i.e., cybernetic vs. A.I.) to problems of modelling of human activities are non-fundamental, and each approach of these overlapping disciplines only signifies "the different ways of slicing a cake", rather than the divisions themselves.

The early cybernetic learning models relied heavily on ideas of control-theory, such as 'feed-back', and the intricacies of their designs were generally attributable to clever interconnections of few electronic, electrical or mechanical components and devices.

The pioneers of this paradigm did, in fact, approach the problem from a very logical point of view. Their intention was to manifest the elementary modalities of learning in abstractions or artifacts, utilizing observations from simple behavioural traits (mainly in animals) or simple neuronal mechanisms. Their approaches are today, generally, described as mundane, simplistic, weakly-specified, non-productive, or unfashionable. But, an important point is lost in such assessments of these early trends, it was indeed not the intention of their designers to show or to make models which could immediately demonstrate the higher intellectual capabilities of humans. They were more interested in the underlying fundamentals, and a slow progressive, yet versatile, path to the higher strata of intelligence hierarchy - hence, the lack of spectacular achievements.

However, it is true to say that cyberneticians, neural-net scientists, or other adherents of the "bottom-up" approach to 'learning' have not exhausted all possibilities of their paradigm. Maybe, today, some 30 years after the high point of their elementary 'learning' models, with the introduction of new technological tools and techniques, and increased theoretical and empirical knowledge, once again attention should be focused on this intuitive approach; and some of the fundamental aspects of the modelling of the process of learning scrutinised in much more detail.

1.8.2 THE INFORMATION PROCESSING "TOP-DOWN" APPROACH

The centuries old argument of brains vs. machines surfaced with a new vigore with the flourishing of the new disciplines of computer sciences, A.I. and Robotics. Of course, the principal goal of the research in these fields is not to cleverly mimic 'natural learning' (in a science-fiction sense), but to scientifically create machines which are genuinely able to 'learn', and use this capability, fundamental to all 'intelligent' behaviour, in improving their performance or knowledge.

The 'computer' or 'information' revolution is generally considered as the second industrial revolution. But, in fact, the course of the mechanization of thought processes can be regarded more as an evolutionary process, which is developing in parallel to the mechanization of muscular powers, currently manifested in industrial robotics and automation.

Today, developments of computer technology, currently in its 5-th generation with 6-th and 7-th in planning, are taken for granted. Their power, speed and capacity are increased an order of magnitude every few years.

The general tendency in information processing is towards computer based models. At the root of this trend lies the central belief that mathematical formulae, machines and robots are incapable of conveying 'mental' symbols and carrying out complex computations on such symbols; after all, the ultimate objective is to have systems which can display facets of 'thinking'. However, the computer is seen as a suitable vehicle for such abstractions. This trend, in turn, accounts for the way that the learning process (and other endowments of the brain) has come to be 'externalised' to the brain, and its explanations based on higher organizational features of the structure rather than the specifics of mechanisms or symbols themselves. In other words, the 'fibre' of the model is no longer considered relevant, whether it is biological, chemical, electronic, mechanical, or abstract; and the computer is simply regarded as a convenient tool at disposal.

On reflection upon this latter point, if some of the current disciplines involved with the modelling of 'learning' (e.g., A.I.) are looked at in isolation from its 'computer' component, then distinctions between such subjects and psychology, linguistics, sections of philosophy, or some other fields become less clear cut. It is the appreciation of this underlying kinship which has compelled some workers to describe all these disciplines under the umbrella of "cognitive sciences". The science of "Cybernetics", in the same vain, can be regarded as the broader interdisciplinary approach which encompasses biological, cognitive and mechanical aspects of 'learning'.

A characteristic of the very title of many information-processing disciplines of today such as "Pattern-Recognition", "Speech-Recognition", "Artificial-Intelligence", "Problem-Solving", or "Theorem-Proving" is that their names actually refer to the ultimate 'goals' of their workers, rather than signifying the 'course' of their scientific endeavours, as is the case with most other scientific topics. If, indeed, the objective of true artificial-intelligence was attained then there would not be a point for such a pursuit.

Some early Information Processing models were clearly engaged in either or both aspects of 'simulation' and 'synthesis'. But, this distinction has become much more vague in the increasingly more complex models of A.I., where the appreciation of underlying difficulties of the task has made the distinction, to some extent, immaterial.

Within the doctrine of Information Processing no real attempt is made to incorporate in a model the biological structures or the natural ways of achieving the same end. It is also believed (although not with very firm conviction) that a conglomeration of sub-models, each capable of simulating one aspect of human intelligence, will at the end lead to the attainment of true artificial 'intelligence'. 'Learning' models in this approach are generally devised for four major categories of 'learning' tasks:-

(1) - Rote Learning: memorizing a sequence of tables for later use.

(2) - Parameter Learning: learning by adjusting or discovery of values of certain parameters - e.g., in pattern-recognition.

(3) - Method Learning: algorithmic learning of procedures applicable to different situations.

(4) - Concept Learning: building a knowledge structure out of elementary concepts.

In recent years, the trend in this field has been away from the 'pure' aspects of learning, and towards the 'problem-oriented' aspects. Various economic considerations have compelled many workers to seek more practical and industrially-applicable results. Therefore, the early 'learning' models based on 'natural' phenomena are now, in the main, considered out of fashion (and to some extent futile) by the adherents of information-processing approach to modelling of learning - the emphasis has shifted to tackling specific clearly defined problems, or class of problems, at the 'representational' level.

1.9 WHY BUILD HARDWARE MODELS AND ROBOTS ?

As we discussed previously, hardware devices, as models, have some intrinsic qualities which can facilitate a better understanding or synthesis and simulation of a phenomenon. In particular, the reason why hardware simulations of 'learning' are attempted is because most behavioural, cognitive or neuro-physiological experimental results are difficult to interpret; and theories based on them are, normally, imprecise and their consequences hard to follow. Hence, the way that mathematics provides a rigorous foundation for theorizing in the physical sciences, the machine simulation can be regarded as fulfilling a similar function in behavioural sciences.

The conception of modern notion of "robotics" is generally attributed to Czech play writer Karl Capek (term "robot" is a derivation of Czech word for 'workers'). Science-fictional robots aside, there is not much doubt that real robots have been very useful in assisting man in situations where the environment is dangerous, inhospitable, or inaccessible; or the task is mundane, repetitive, routine, or dangerous; or when simply some improvement of a performance criteria is required.

Today's sophisticated robots (hardware models) are mostly used in conjunction with a digital computer, and, therefore, when referring to their 'intelligent' behaviour it is basically the 'intelligence' of their accompanying computer program which is insinuated and not the hardware construct itself.

In addition to the advances made in computer technology, which are so relevant to 'learning robots', numerous other areas of engineering have also been providing their specific contributions to such designs.

Generally, the computer is regarded as the 'brain' of the robot, the manipulators or mobile vehicles as the 'body' or 'effectors', and the various devices which can provide almost any imaginable quantified measure of the physical environment (e.g., light, colour, vision, heat, touch, smell, sound, pressure, speed, movement, sonar, etc.) as the 'sensors'. The inputs of these sensors are converted to electrical signals which, in turn, are translated to digitized codes of computer language.

The technology of robotics is forging ahead at a brisk speed, and constantly more precise, faster, more powerful or more versatile manipulators and robot vehicles are constantly being developed. The wondrous tasks that some of these robots are able to perform are quite astounding, a recent Japanese experimental robot can play a whole musical score on the piano without any training, by recognising the musical notes with a pattern-recognition system and using delicate finger type manipulators; other advanced walking or hopping robots can now show very high degrees of dynamic balancing capabilities. Yet, in spite of all these achievements very little actual 'learning' is included within their designs - most of the 'learning' achieved in the current generation of robots only refers to the memorizing of spatial trajectories or paths.

Another category of robots, such as the project undertaken for the purposes of this thesis, are the 'experimental robots'. The forerunners of this class of robots were the mechanical and electrical devices ('turtles') which were constructed by cyberneticians interested in simulating specific behavioural traits. Since the early 1960's, various laboratories in academic institutions have been engaged in building either 'roving' mobile robots or 'hand-eye' manipulative robots with many interesting and varying features.

Some mobile robots had manipulators or TV-cameras mounted on the carts, others used legs rather than wheels for movement - here, the studies based on human or animal limb movements have provided a helpful source of data. Many of the early experimental mobile robots had their logic decision making mechanisms embodied within the hardware of the design, however, the later versions had the processing centre incorporated in a digital computer, which was either on-board or housed separately and in communication via cables or infra-red/ultrasound/radio signals - this dissection of the "brain" from the "body" of robots was considered as an important breakthrough.

The computer controlled laboratory robots, as well as being interesting scientific curiosities, are built for a variety of reasons. In some instances it is an obvious practical problem solving or industrial application which dictates the necessity of this type of research. Other workers are simply using these vehicles as tools for experimentation or demonstration of particular psychological or educational criteria; here, robots are regarded as a kind of simple roving animal which interacts with a rich sensory environment.

Similarly, various cognitive postulates on perceptual problems may be investigated using robots that have visual sensory devices attached to them; and, indeed, they have been influential in a better understanding of the processes of formation and acquisition of knowledge structures. Some experimental robots have aided the development of various programming languages and tools (e.g., LOGO, PROLOG). Finally, an experimental complex robot can be a good medium for combining various isolated ideas within a single autonomous body, such as combining a visual system with a navigational system or a manipulative system.

The developments of laboratory robots have been progressing almost in parallel to the technological advances of computer sciences, having nodes and troughs which are attributable to particular trends of an era. For example, in the early 1980's such developments received an impetus from the widespread use of LOGO programming language in educational establishments. LOGO needed a simple 'turtle' for demonstration and many such devices were constructed both in laboratories and also marketed commercially.

On the whole, the principal reason why these models did not quite realise the early aspirations of their cybernetic designers is the lack of supporting 'general' software developments, which would have allowed much more interesting behavioural emissions from such models. Hence, all such laboratory exercises have concentrated on very specific behaviour patterns or tasks. To some degree, their designers have been frustrated that despite having a versatile physical artifact, with many potentialities, they cannot find appropriate methodologies to embody interesting behaviour within it.

1.10 THE OBJECTIVES OF THESIS: DESIGN OF A SIMPLE 'LEARNING' ROBOT

Before we attempt to outline the objectives of this thesis in detail, it would be a good idea to briefly explain the background, the progression of ideas and endeavours, and the circumstances which were influential in determining the direction and the development of the thesis in this particular extensive way.

The original topic of interest was, of course, the modelling of learning, principally, in the simpler manifestations of its hierarchy. The objective was to start from the lowest level and elaborate to the higher strata of learning. After a preliminary study of the subject, the design and the construction of a computer controlled mobile robot was decided upon as a project. The aim was to construct a very versatile experimental tool for implementing various learning hypothesis, or using it for synthesis and simulations of specific learning related criteria. The task of designing and building 'learning' machines or systems which in some sense can act 'intelligently' has an intrinsic fascination. The robot was to represent a very simple organism, being able to interact with its environment in a primitive way.

The processes of design and construction of the robot was an interesting and challenging endeavour, and will be discussed briefly later. Various considerations had to be made as to what features to include so that an experimenter may interact freely with the machine, without various 'house keeping', 'operational' or design constraints which characterised many other such experimental devices. Frequently, these constraints hindered the latter software developments of the machines, forcing their designers to change the focus from 'fundamental' issues to 'machine specific' issues.

The degree of the complexity of the robot was also a crucial consideration. Since if it had too many input or output devices, then the fundamental issues of interest could not be represented faithfully. Similarly, if the potentialities of the model were too trivial, then no 'learning' behaviour of any significance could be manifested. It was decided that a radio controlled device would be built, which would be in constant radio communication with a computer that can continuously monitor the states of its touch, shock, and sonar sensors; and, also, could transmit direction and speed instructions out to the robot. This device, which in certain functional capabilities could be possibly compared to an amoeba or a virus, would be able to autonomously rove in the environment of a laboratory, and should have enough sensory complexity to demonstrate simple obstacle avoidance type 'learning' tasks.

The actual building of the robot was an arduous task itself, and involved many considerations and hurdles which had to be overcome, such as design and construction of motor-drive, power supply, charger, sonar, data-communication, radio-transceiver and computer-interface circuitry, as well as the chassis, mechanical parts, contact-switches and various other components.

After the completion of the 'body' of the robot and its successful testing, the next stage was to start the development of an appropriate software which would enable the robot to demonstrate, within its universe of discourse, a 'learning' behaviour, more or less the way it would be manifested in a hypothetical organism of same complexity. Although, it was envisaged that even with such primitive senses the same machine, with the aid of the computer, could be made to conceptualize about its environment in a trivial sense. Yet, this higher level of description would be a choice of the designer, and not intuitively detectable from the design. It would clearly stretch the capabilities of the robot which was not equipped to deal with the conceptual level of description in the first place. This point emphasises the parity of behavioural complexity and underlying neural structural complexity observed in nature; and should be an important consideration when attributing 'intelligent' aspects of behaviour to machines, or when dismissing a machine as incapable of achieving such 'intelligent' behavioural traits.

Various simulation programs were written for operating the robot within the laboratory environment. These programs (written in FORTRAN) ranged from simple traversing of graphs or mathematical functions, to the simulation of a reflexive behaviour which entailed the machine roving around and moving away from obstacles upon collision. The behaviourally most advanced program involved the sonar scanning of the room by the robot and drawing a kind of map of the environment on the computer display. This allowed the robot to plan a path and explore the room without bumping into various obstructions.

At a lower level of behavioural complexity, also, a simple algorithmic procedure would enable the robot, using its sonar signal, detect a course of collision and take avoiding actions. The program itself was quite long, and included many 'machine supervision' secondary aspects, yet, the essence of the, so called, 'intelligent' aspects of the behaviour was contained in a few lines of programming (a simple mathematical deductive process). Now, the real challenge could be redefined as: how we could manifest the same logical deductive process by a gradual 'learning' in the machine starting from almost random behaviour, and using basic criteria.

The search for the implementation of such a 'learning' process involved looking at similar areas of research and designs of 'learning' robots; and even the blue-print of a simple scheme was devised which without doubt would, unfailingly, lead to the attainment of such a desired 'learned' pattern of behaviour. Yet, more and more, it was becoming evident that the objectives of such simulation exercises must be clarified in the first place, and more fundamental questions answered before embarking on such an endeavour.

It was observed that too often the designers of such programs, with the aim of achieving interesting and quick results, delve in the specifics of a particular formalism, methodology or behavioural pattern (normally from their previous backgrounds), without considering the full implications of such endeavours. The results of this kind of hasty approach, in general, only reinforced some preconceived ideas, and had very few surprises in store; normally, leading to an even narrower perspective of the subject, and leaving the elements of initial beliefs of the designer largely untouched.

Hence, a principal shift of emphasis took place, and it was decided that before the specific preferences (prejudices?) of our design was set a journey through the genesis of the phenomenon of learning be undertaken.

The approach taken was to get a kind of 'snap-shot' of the state of research in various disciplines at this moment of time; and also to find out the applications, future possibilities, and aspirations of workers in each field. The underlying philosophical, historical, biological, evolutionary, and social aspects would, in addition, be covered so as to understand some of the interrelations of various disciplines, their common features, or backgrounds to their development.

The challenge of a comparative study of all aspects and approaches of learning was considered as both rewarding and daunting. Yet, such a multi-approach compilation was regarded as a valuable contribution to our later more 'objective' analysis of the 'learning' models, and also to any designer of such models at the preliminary stage of their task. Specially, since most studies of the learning process and its modelling are carried out in the rigid boundaries of a single discipline, which reflect the common interests or the typical backgrounds of researchers in its domain.

An attempt will be made to highlight the shortcomings of each paradigm, as well as citing its achievements and strong points. At the final analysis, however, it can be said that the definitive statements of the weaknesses of each approach to design of 'learning' models are much more conclusive than the hopeful possibilities professed by their adherents.

Yet, deliberately, no particular angle of the arguments will be promoted on exclusive basis, or pursued as a theme. Since, not only it is the striving for objectivity which culminated in this wideness of perspective, but, it is the express advocation of this thesis that such a 'broad' view should be adopted in "the modelling of learning". Hence, in simple terms the objective of this thesis can be summed as: "MAKING QUALIFIED JUDGMENTS"

It is, however, not contended that for a real breakthrough a revolutionary abrupt step need to be made in the development of 'learning machines' (i.e., starting from basic alternate criteria, possibly non-related to the natural occurring processes). But, the aim is to stop and view the whole topic with the widest possible perspective and try to clarify few goals, and deviations from such goals; and also to look at the 'total' context in a much more detail without being prejudiced by a blinkered uni-disciplinary bias. In other words, we would like to see what conclusions, interpretations, commonality of concepts, or inter-relations exist between the branches which have sprouted from the mainstream of the psychological and philosophical studies of learning.

1.11 THE OUTLINE OF THE THESIS

In starting an undertaking such as this project there are certain characteristic pitfalls which may pose problems for the researcher. "Solving the World" and " Ambitious Paralysis" are, according to Bundy (1984), two of the traps which may hinder progress or even force the worker to abandon the endeavour in frustration. However, due to the very broad nature of this thesis it was inevitable that at times deviations from the main objectives would happen, and keeping the underlying aims in focus was one of the major obstacles of this endeavour.

There is a saying that: "a little learning is a dangerous thing", and as we are trying to 'learn about learning' in this wide ranging manner then the difficult decision has to be made as to what depth should we probe the various fields involving this ubiquitous phenomenon. Is it enough to skim over an approach, or should we familiarise ourselves with the particular methodologies and formalisms of the approach.

Due to the extensiveness of task in hand, a compromise had to be made to limit the investigations of subjects to a broad yet uniform sense. With only occasional indulgences made into detailed analyses which are considered essential to the core problem - hence the fundamental aspects (both knowledge and behaviour related) concerning 'learning' models are more emphasised. Typical or historically important works of each discipline will also be cited, their notable proponents mentioned, and the basic criteria and viewpoints of the paradigm discussed.

We will not try to engage in very deep philosophical discussions of the periphery concepts of learning. These debates are typified by the scrutinies of the concept of 'intelligence' and its various aspects in many A.I. related literature. Such a pursuit, although worthwhile and interesting in a philosophical domain, in a sense will probably narrow down the definition of learning in the sweeping manner we wish to study, and will deter us from covering the widest possible range of approaches to 'learning' (and 'adaptation'). The notion of 'learning', hence, will be discussed in its consensus understanding of the concept, which, incidently, varies from discipline to discipline. In this work we will attempt to characterize the process of learning from different points of view, and also to specify distinctions or unifying features.

A decision had also to be made as to what levels the mathematical abstractions should be scrutinised - in spite of personal preferences for such abstractions, due to a mathematical background. Based on the broad nature of the thesis, the general criterion used was to include only formalisms which were pertinent to the underlying fundamental issues, or where a direct relevance could be seen to the specific goal (i.e., the design of simple 'learning' machines); and the understanding of which would not require rigorous mathematical background.

Therefore, the outline of this thesis can be summed up in the following:-

(1) - Natural learning observations studied in a broad sense, including both physiological and psychological aspects.

(2) - Hierarchies of learning identified in evolutionary, functional, and developmental sense.

(3) - Various features and components of different types of learning processes identified.

(4) - Models and modelling discussed in general; and specific principles, tools, formalisms and techniques involved in the modelling of 'learning' examined.

(5) - Approaches to the problems of 'simulation' and 'synthesis' of 'learning' considered, their development outlined, some notable examples discussed, and relevant issues addressed or argued.

(6) - The design and construction of a hardware mobile robot discussed as a possible tool for experimentation, simulation or synthesis of various simple 'learning' schemes.

(7) - Features deemed to be essential in design of 'learning' systems discussed, both in general terms and also in the context of above experimental hardware model.

1.12 AN OVERVIEW OF THE CHAPTERS

Finally, here, we will briefly outline the composition of each chapter, and the course of development of the thesis. The intermediary chapters or sections, however, could be studied also in isolation, since they are separated by 'natural' methodological, paradigmatic, or phenomenological breaks in approach or level of enquiry.

In Chapter-1 we began by a general definition of the problem which is of interest to this thesis, namely the process of learning and its modelling. The various developmental aspects of related sciences were looked into, and specific key issues of contention were analyzed in more detail. Next, the modelling aspects of the phenomenon of learning and various approaches involved were discussed. Finally, the case for our particular global approach to the issue was argued and the objectives identified.

In Chapter-2 we will be examining the three principal approaches to the empirical studies of learning in man and animals. The learning theories developed in psychology, brain-sciences and cognitive studies will be discussed, and the backgrounds to their development pointed out. Various definitions of learning, its attributes, and related issues will be addressed; and their underlying neuronal mechanisms described. Also, different taxonomies and hierarchies of the learning process and its mechanisms will be outlined.

Chapter-3 will be concerned with the general problem of the modelling of learning, hence, discussions of some fundamental issues of modelling will be undertaken; principally, involving its facets of 'simulation' and 'synthesis'; as opposed to the empirical studies which mainly involve 'analysis'. 'Learning' models will be broadly categorized into 'natural' vs. 'artificial'. Finally, abstract tools, mathematical theories, and techniques used in the modelling of 'learning' will be reviewed, and some teleological considerations outlined.

In Chapter-4 and Chapter-5 an attempt will be made to describe the various synthetic (artificial) approaches to the modelling of the learning process. The order of discussion approximately reflecting the historical sequence of appearance or prominence of each approach. After outlining the essential features and problems within the various disciplines involved, typical 'learning' models of each paradigm will be described. In addition, evaluation and analysis of each approach, its achievements and its goals will also be undertaken in relation to the 'complete picture' we are trying to put together on the topic of learning. The principal approaches covered will be: 'adaptive control', 'neural-net', 'automata theory', 'self-organising system', 'cybernetic', 'pattern-recognition', 'artificial-intelligence', and 'robotics'.

In Chapter-6 the specific objective of the thesis, or the analysis of various aspects of design of cybernetic 'learning' models, will be pursued. The robot experimental vehicle, referred to earlier, will be briefly discussed; yet, the actual technical details of construction or the developed software will not be included. Additionally, some fundamental issues in 'learning' models will be discussed, the possibility of their realization in simple hardware models (e.g., our particular robot) considered, and the blue-print of a hypothetical system capable of 'learning' will be outlined. Next, in Chapter-7 we will summarize our discussions and make final conclusions.

Finally, in the Reference section the various literature used in the course of writing this thesis, in particular, the large number of papers accumulated over the years on the subject of learning and its modelling, will be listed. However, again, we will not include the many technical and computing books or references used in the course of our hardware development.

CHAPTER 2

EMPIRICS AND THEORIES OF LEARNING

2.1 INTRODUCTION

Observations based on the dynamics of the living and the inanimate world surrounding us show an intriguing variety of learned and adaptive processes.

A child learns to coordinate its sensory-motor activities, or to recognize the features and significances of its environment; a man trying to acquire a physical skill learns to develop and control his muscular movements; a student learns the meanings of new collections of symbols and grammatical rules, and learns to solve a variety of conceptual problems; an animal in an experimental psychology laboratory learns to tackle a complex maze-running problem; a circus animal learns to perform astonishing feats. All in all, animals in nature from mollusks to mammals show a diverse range of learned behaviour, and even simple organisms such as amoeba or bacteria can exhibit adaptiveness to new stimuli in their environment.

Similarly, a robot in an industrial plant, or a roving mechanical turtle in a laboratory, can be made to improve its performance by using data from past experiences; a chess or checkers playing computer program can 'learn' to play at proficient level, and compete against human opponents, apparently by 'learning' from its mistakes; a complex pattern recognition system 'learns' shapes of objects or human faces, and gradually makes fewer matching errors.

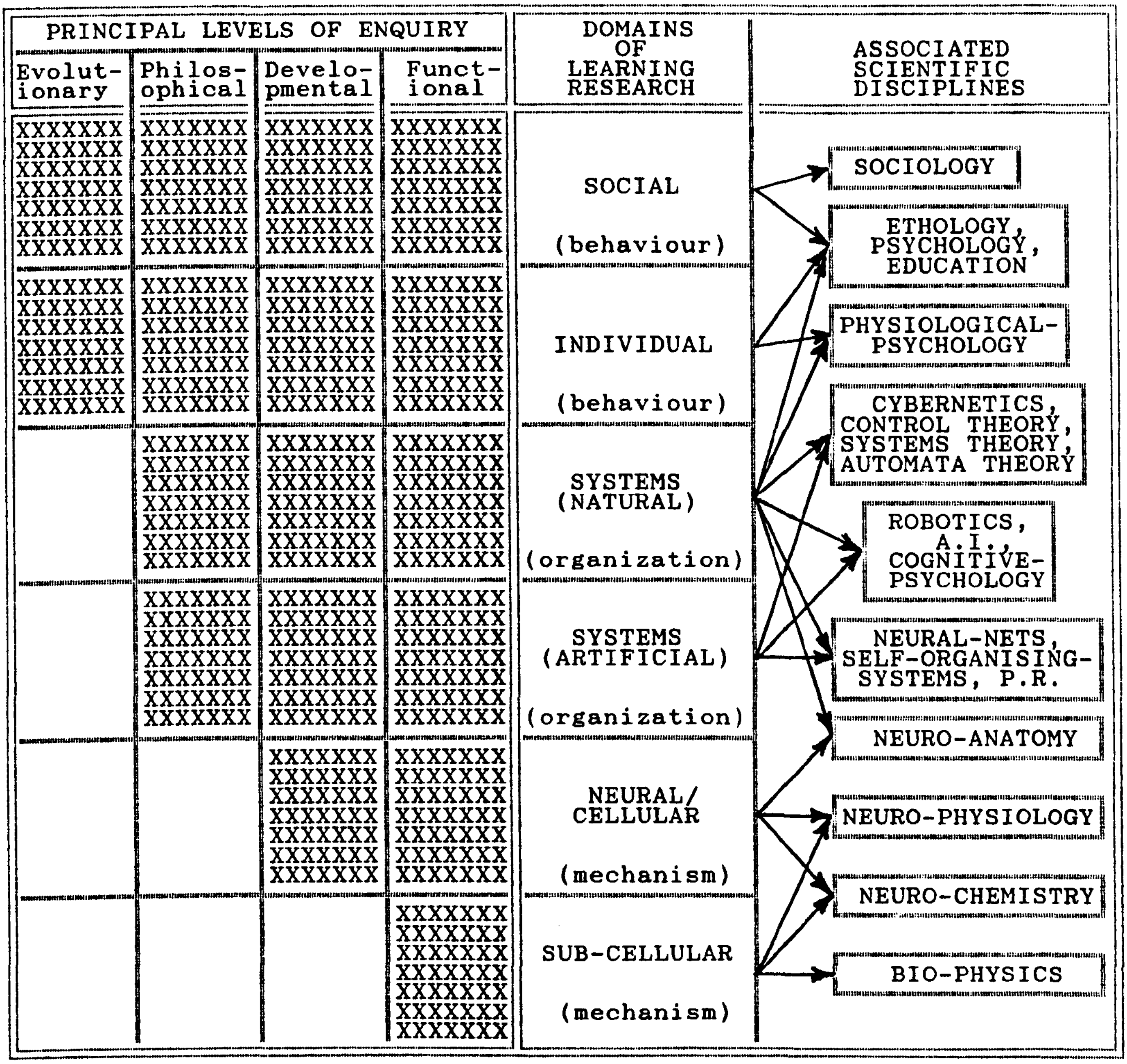
Problems involving learning or adaptation have puzzled the minds of researchers from many diverse fields of science, as well as occupying the 'pure' researchers of the learning phenomenon. Some of the principal points of discussion and controversy are the definitions of criteria shared by different types of learned and adaptive behaviour, and also the identifications of underlying features of such activities - for example, in the above diverse range of 'learned' behaviour.

2.2 VARIOUS LEVELS OF INVESTIGATION AND STUDY OF LEARNING

"Learning" in its general sense can be defined as: 'the process of adaptive interactions between a physical entity or an abstraction and its environment'.

Yet, the phenomenon of learning, as seen from the hierarchical representation of TABLE 2.1, is a multi-faceted and multi-dimensional concept, whose investigations entail many specific descriptive or mathematical aspects and levels.

TABLE 2.1. The main disciplines associated with various domains of Learning Research are outlined in a hierarchy; the first four columns show the principal levels and aspects of study.



In fact, most scientific enquiries are characterised by a hierarchy of explanations, and a variety of investigatory levels. In learning sciences the levels of descriptions range from speculative issues about the nature of 'mind' and social aspects of learning, down to cellular and atomic features of its manifestation.

Scientific research, in general, involves the scrutiny of a subject in a specific descriptive framework of reference. Hence, normally the results which are obtained are consistent and valid, if merely looked at from within that particular body of knowledge. The principal distinction between different hierarchical levels are in their operational considerations. In other words, the views that are held, the techniques that are used, the questions that are asked, or the type of answers which will satisfy such questions.

A look at the historical development of most scientific enquiries shows that the initial explanations and observations of a phenomena mainly begins from the higher strata (e.g., the verbal characterizing) of the hierarchy of its descriptive levels. These observations, in turn, generate the more precise lower levels. The general consensus is that the statements at the lower hierarchical levels are much more solid and fundamental, and can be applied over a wider range of problems; yet, their explanatory powers lack the depth and the reasoning features of the higher levels.

Of course, the boundaries of many such levels are only loosely defined, and hence ambiguities arise when attempts are made to cross over these descriptive boundaries. Similarly, when results occurring at one level are predicted as a consequence of events at another level then various inconsistencies become evident. The principal reasons for such difficulties are the lack of direct correlations between different descriptive levels; or, the need for additional dimensions of enquiry, such as 'ontogenetic' and 'phylogenetic', to be taken into account in the higher levels of description or investigation.

It can be seen from TABLE 2.1 that there are many specialized areas of current scientific investigation interested in the process of learning. However, most learning-theories have been developed from within the following three principle branches of science:-

(i) - COGNITIVE SCIENCES: the cognitive scientists are interested in the study of the accumulative and organizational aspects of learning related temporal experiences. Hence, normally, knowledge structures and other perceptual issues are investigated within cognitive sciences. Most of the research in 'Artificial Intelligence' (A.I.) could be included in this category.

(ii) - BEHAVIOURAL SCIENCES: the workers in this field investigate the external and empirical manifestations of the learning-process in human beings and animals. The formation and the development of learned and adaptive behaviours are the primary subjects of interest in this area.

(iii) - BRAIN and NERVE-SYSTEM SCIENCES: the researchers in this area study the internal mechanisms of the learning process, observed by the changes which take place within the nerve-system during learning. The brain scientists are in pursuit of the neural correlates of adaptive behaviours.

2.3 DEVELOPMENT OF LEARNING THEORIES - PHILOSOPHICAL/HISTORICAL PERSPECTIVES

There is little doubt in the minds of researchers in learning that this science has developed from the philosophical enquiries into the nature of man's mind. Over many centuries the problems of 'mind' and 'body', and their relation to each other, had occupied the attention of philosophers and later psychologists. Elaborate concepts (e.g., consciousness, purpose, desire, etc.) had been defined to explain the 'intelligent' activities of man. A typical theoretical framework as an explanation for such modes of human behaviour was the doctrine of 'Hedonism' in the 17th and 18th centuries, which asserted that individuals are motivated by a desire for 'pleasure' and by an aversion to 'pain'.

Historically, human subjects were the principle domain of investigation for the researchers in behaviour. Much attention was given to the reasoning and intellectual faculties of man, especially, the effect of experiences on the mind and the organizational aspects of such mental impressions. During the past hundred years, the presence of two opposing philosophical views of 'Rationalism' and 'Empiricism' has led to a major divergence of ideas in the field of learning sciences.

2.3.1 AN OVERVIEW OF LEARNING-THEORIES WITHIN COGNITIVE SCIENCES

In rationalism, reason rather than the sensations input to an organism is the basis of knowledge; the animal's experiences are dealt with in a non-reductionist and holistic fashion, and the raw data are interpreted only according to certain forms of innate perceptual assumptions and relations. The internal processes and concepts embodied in the brain are considered to be as important as the experiences themselves. The case for the Rationalist view is much stronger in describing the higher forms of learning such as problem-solving or language-acquisition, and also in explaining the structural organization of perceptual data. Cognitive scientists and philosophers share the rationalist view point.

A major school of thought, the 'Gestalt', was developed within the rationalist view-point, in opposition to the 'reductionist' analyses of experiences. The Gestaltists were originally concerned with perceptual organization and the unitary nature of sense information. But, later, using similar concepts, endeavoured to explain the problem-solving and learning abilities of animals. This was not done by considering only the animal's input–output interactions, but by taking into account notions such as: knowledge, thinking, planning, inference, intention, insight, goal-directedness, expectation etc. In cognitive explanation of leaning, the worker is interested in large-scale (molar) rather than small-scale (molecular) action descriptions.

Many elaborate learning theories have flourished from this rationalist/cognitive view point. Even, some mathematical models or pseudo-behavioural learning models (e.g., Tolman's) have been developed with cognitive inclinations. Similarly, many of today's Problem-Solving, Artificial-Intelligence and Pattern-Recognition models also subscribe to the cognitive and organizational view.

The discovery of various knowledge related structural or information retrieval (i.e., 'memory', 'remembering' and 'forgetting') aspects of the learning process is also attributed to Cognitive-Psychologists. In future chapters we will investigate the details of some cognitive learning models, and discuss them more fully.

2.3.2 AN OVERVIEW OF LEARNING-THEORIES WITHIN BEHAVIOURAL SCIENCES

The doctrine of Empiricism has the view that experience is the only source of knowledge, and our 'ideas' are derived either by a direct copying of sensory impressions in the brain, or by combining several simple (or complex) ideas into one simple (or complex) idea. The majority of behaviourists and physiological psychologists are proponents of this view of knowledge-formation, and most of the associative theories of learning have been developed within this school of thought.

Empiricism has four major components:-

(i) – SENSATIONALISM: all knowledge is derived directly or indirectly through experience.

(ii) – REDUCTIONISM: all complex ideas are composed of a basic pool of simple ideas and are, in turn, reducible to these simple ideas.

(iii) – ASSOCIATIONISM: ideas or mental events are connected through the proximity of their occurrence in time or space.

(iv) – MECHANISM: the mind is a machine like entity, built from simple components, with no mystical dimensions.

The study of human and animal learning behaviour has been one of the most intriguing subjects for the inquisitive scientists throughout the ages. Initially, hampered by the complexity of the task and the vagueness of methodologies, or other socio-religious obstacles, no significant advances were made into finding a scientific understanding of the 'learning process'. It was

Darwin's study of animals' common ancestry and the subsequent development of evolutionary theories which laid a solid foundation for the systematic research by behavioural-physiologists and behavioural-psychologists; namely, principal adherents like Thorndike, Pavlov, Watson, Lorenz, Hull and Skinner. The work of these behaviourists have contributed to the rich body of contemporary knowledge on animal-learning; and their rigorous empirical enquiries triggered an explosion of scientific interest which has led to the enormous variety of research into this subject.

It is relevant here to mention a dichotomy which exists regarding the methods used for the investigations of learning in animal-behaviour sciences:-

(i) - EXPERIMENTAL PSYCHOLOGY: the experiments in this field are performed in controlled laboratory environments using isolated stimulus-response investigations of a behavioural pattern, which may not normally occur in the animal's encounters with nature. There is a clear and deliberate omission of all mentallistic references (e.g., consciousness, mind, free will, imagination, emotion, etc.). The main task is the prediction and the control of behaviour by the sole use of observed data; introspection plays no part in this method.

(ii) - ETHOLOGY: or the study of animals in nature, is historically the established branch of behavioural-sciences, but, 40–50 years ago it was overshadowed by the rise of experimental psychology. However, more recently, the rigid settings and the unnatural surroundings of mazes, boxes and pulleys of experimental psychology laboratories have compelled some behavioural scientists to return the domain of study back to the animals' more accustomed surroundings in nature.

2.3.3 AN OVERVIEW OF LEARNING THEORIES WITHIN BRAIN AND NERVE-SYSTEM SCIENCES

During the 18th and the early 19th century, the accumulation of physiological knowledge was gradually localizing the brain as centre for 'mind' and 'soul'. This, later, followed by the general acknowledgment of the nervous-system as the sole organ responsible for control and coordination of all animal activities. By the beginning of the 20th century, the discovery of electricity and investigations of its effects on nerve and muscle tissues, together with an increasing knowledge of electrical, anatomical and physiological details of the brain (and its components the 'neurons') had led to a much clearer view of the make-up of the brain and the nervous system.

Work pioneered by the famous Russian Scientist Pavlov into isolating specific aspects of an animal's learning behaviour, and other early analytical studies of response mechanisms, resulted in the introduction of associative learning theories. These endeavours were the inspirations for the brain-researchers to embark on a methodical study of the brain's functions and structures.

The theories of 'Neural-Pathways' introduced by Hebb (1949), as a progression of Pavlovian ideas was an attempt to find a coherent explanation of learning in terms of sensory impressions on nerve structures. The convergence of behavioural and physiological sciences helped to establish many inter-disciplinary subjects. However, several strictly behaviourist or specialized neuro-physiological branches of learning sciences were also developed. Although, there have been many significant advances in all fields of nerve and brain sciences, physiological theories that can offer global principles or laws for different types of learning have not been forthcoming. The most solid achievements have been based on the empirical observations of specific isolated aspects of learning such as 'conditioning' or 'habituation'; and have, mainly, involved experiments on simple nerve-cell preparations or lower animals.

The spectacular progress made in the science of Genetics, from the speculative observations of Mendelian Genetics to the descriptive explanations of Genes, followed by the molecular discoveries of DNA, RNA, and finally the mechanisms of genetic coding, can be a good example as to how, ideally, the theories of learning should develope. Therefore, the study of the learning process at the cellular level should really be complementary to the studies at the behavioural and cognitive levels, and not isolated within its domain. Since, only a combined broad approach will enable the attainment of significant breakthroughs, on par with the science of Genetics.

2.4 BEHAVIOURAL APPROACH TO CONCEPTS AND ISSUES IN LEARNING

To understand some of the methods used by behaviourists in exploring the mysteries of this enigmatic subject, we should look at some typical definitions for "learning" in related literature.

LEARNING IS:-

"A change in the strength of an act through training procedure."
(Hilgard and Marquis, 1940)

"The reassortment of animal's responses in a complex situation."
(Skinner, 1961)

"A process which manifests itself by adaptive changes in an individual's behaviour as a result of experience." (Thorpe, 1963)

"The process through which life experiences leave a mark on an individual and permits an animal to develop new adaptations." (Dethier and Stellar, 1970)

"A relatively permanent change in behaviour which is not directly observable (different from performance) and comes about as a result of experience or practice." (Mednick, Pollio and Loftus, 1973)

"An adaptive modification of behaviour towards a stimulus that can be traced to a specific experience in an animal's life with that stimulus or a similar one." (Alcock, 1975)

"The change in behaviour or behaviour potential to a given situation brought about by repeated experiences in that situation, provided that the behaviour change cannot be explained on the basis of the subject's native response tendencies, maturation, or temporary states." (Bower and Hilgard, 1981)

We can see that in the above selection of definitions, regardless of their generality or specificity: (a) - there is an underlying mechanistic approach to the explanations of the learning-process; and, (b) - there is emphasis on the externally apparent aspects of learning, namely, the 'behaviour', and little reference to any of its internal features.

A typical behaviourist's experiment deals with the categorization, the isolation, and the scrutiny of a particular external behavioural aspect of an animal's learning-process. Under a desire for objectivity (perhaps misplaced), and a quest for theorizing the learning-process, the behavioural scientists have demanded that all knowledge should be operationalized in terms of behavioural responses. The enormous volume of research amassed over many years in the behavioural-sciences has resulted in the formation of a solid framework of concepts and criteria about the behavioural aspects of the learning-process. Now, without trying to go into much historical, experimental or developmental details, firstly, we will outline the basic assumptions which have been the guidelines for researchers in this field, and secondly, we will enumerate the major concepts that have emerged from this body of science.

2.4.1 CHARACTERISTICS, ATTRIBUTES, CONCEPTS AND LAWS OF LEARNING

All classes of learning and adaptive processes in animals are intuitively considered to possess a unique origin, this is in view of their common phylogenetic background. Hence, the differences seen between various learning processes are not merely an indication of their accompanying structural complexities, but, also, reveal the levels of an evolutionary ascent. This hierarchy being the result of the endeavours of various species in trying to cope with increasingly complex environments and learning tasks. Ideally, some unifying concepts and theories should exist to encompass and explain all classes of adaptive-processes. But, in reality, the empirical analyses of these processes have compelled the behavioural scientists into defining rigid boundaries and characteristics for different types of 'learned' behaviour.

2.4.2 LEARNING vs. INNATE BEHAVIOUR

One of the most notable points of discussion in behavioural sciences, as well as being historically significant, has been the dilemma over 'innate' against 'learned' behaviour. This strict classification of behaviour led to years of debating amongst researchers in animal-behaviour.

The question of instinct vs. learning was also at the root of the so called 'Nature-Nurture' controversy. On the one hand, the hereditarians claimed that behaviour is mainly the result of an internal genetic process. On the other hand, the environmentalists maintained that most behaviour is the product of experience.

The validity of this distinction in a fundamental way is not conclusive, and empirical results have shown a mutual dependance in the development of innate and learned behaviours. The Innate (unlearned) responses have been classified into three major types:-

(1) - REFLEXES: simple neuro-muscular reactions to stimuli.

(2) - INSTINCTS: complex patterns of inborn activities.

(3) - TAXES: orientation of the organism with respect to a stimulus.

Instincts are generally referred to the 'stereotyped' patterns of behaviour which: are usually complex in nature; are found universally amongst the members of a species; are evoked without the need for prior learning or experience; are seen in their complete form at their first occurrence; are relatively invariant; and are elicited by a specific releasing stimulus.

The birds' song-learning or nest-building abilities and numerous other ritualistic, predatory, maternal and reproductive activities of insects and animals are classified as 'instinctive'. However, it has been observed that the proper development of an instinctive behaviour is dependent on the external presence of the appropriate environmental cues, as well as the internal genetic make-up and hormonal secretions.

In the case of a young bird learning to sing, ethologists have made the observation that the young bird must be exposed to the singing of a mature bird at a critical phase of its development in order to acquire the complete singing proficiency of a fully grown bird. Similarly, the nest-building capabilities of birds, although quite complete in their first occurrence, are refined in successive seasons as a result of gradual adaptations to various environmental factors.

In addition, some experiments have shown that sensory or other deprivations, or the administration of artificial physio-chemical stimuli during the development of an animal's nervous-system interfere with its acquisition of normal learning-ability. Although, the precise nature of the correlations between biological and environmental events are unknown, a clear interdependence has been demonstrated.

A different version of the nature-nurture controversy is the attempt to apportion the relative contribution of genetic and environmental factors in the formation of a particular type of behaviour.

Today, instinctive-behaviour is no longer considered in a rigid either/or fashion. The phylogenetic composition of an animal determines the pattern of such stereotyped activities which are, in turn, shaped to the individual animal's ecological niche, or its other environmental conditions.

In a sense, the genetic native response patterns, inborn in the animals' nervous-systems, may be considered as the link between phylogenetic-adaptations and ontogenetic-learning.

One of the most important characteristics of innate behaviours are found in their 'Releasing-Mechanisms' (i.e., the sensory cues which initiate a set of genetically programmed actions). Innate behaviours such as: the reflexive constrictions of pupils, the seasonal mating habits of birds, or the orientation skills of insects, are all controlled by releasing-mechanisms. Releaser stimuli could be either perceptual/physio-chemical/environmental inputs, or internal releasers such as the hormones which regulate and dictate many aspects of an organism's life.

2.4.3 LEARNING vs. OTHER SENSORY CHANGES

Further distinctions are normally made between the behavioural changes resulted from the process of learning and other behavioural changes. Some of these changes show a striking resemblance to some features of a 'learned' change brought about by conditioning or extinction, but they are generally thought to belong to a different class of behaviour. However, it is not clear yet if there are completely distinct mechanisms at work at the neuro-physiological levels of some of these apparently similar changes. These non-learned sensory changes are as follows:-

(1) - MATURATION: the developmental changes associated with growth process, such as learning to walk in humans or learning to fly in birds.

(2) - FATIGUE: a recoverable change to performance due to motor exhaustion, such as reduced response time due to tiredness.

(3) - SENSITIZATION: a reversible adaptive alteration to sense organs or sensory habituation, such as can be seen in the heat sensing mechanism of the skin, where sensitization to excessive but non-damaging temperatures may occur.

(4) - DAMAGE: a permanent change to behaviour due to a physical impairment or decay of sensory motor system.

2.4.4 LEARNING vs. PERFORMANCE

Since learning itself is not directly observable (except possibly in a very elementary sense, in the associative conditioning of a simple animal's nerve-cell preparation), a crucial distinction is made between 'learning' and 'performance' (i.e., the only external measure of the process of learning). Simply observing performance, without taking some underlying parameters such as motivation, attention, arousal and drive into account, will not give a clear view of a learning-process. It can be said that performance is an integral 'factor' or a 'variable' of the learning-process.

2.4.5 ATTRIBUTES OF LEARNING

To narrow down the definition of an organism's activities which could be called as 'learned', and exclude all other non-learned behavioural changes, generally, the following three principles have been attributed to learning within the literatures of behavioural sciences:-

(1) - PERMANENCE: learning may result in an enduring change in an animal's future behaviour.

(2) - ADAPTIVENESS: learning comes about as a result of practice or experience which culminates in adaptive or advantageous change.

(3) - CONTROLLABLE: learning can be altered by experimental, instructional or environmental intervention, and is a reversible proccss.

2.4.6 CONDITIONING

A survey of psychological and animal-behaviour literature shows that the origin of modern learning theories is, generally, attributed to Thorndike's work on 'trial and error' learning at the end of the 19th century - in experiments involving caged animals that learned to escape from puzzle-boxes. These experiments led to the development of a conceptual framework, and signified the basis for the methodology used by later behavioural researchers in learning.

Another major landmark in the development of learning theories was the Russian Physiologist Pavlov's experiments on the conditioning of animals' reflex responses (e.g., the salivation response in a dog). This work, in turn, signified the start of a variety of 'Stimulus-Response' (S-R) associative theories. The 'Pavlovian/Type-I/Classical Conditioning' can be, generally, thought of as a simple type of learning. While, lacking the richness and the descriptive/interpretative qualities of a complex learning process, this form of conditioning has the advantage of being easily controllable and demonstrable experimentally; and the theories that have been developed from this type of work have enabled the explanation of many aspects of simple learning in organisms.

The physiological processes involved in the response conditioning of simple animals have also been investigated, mainly, by noting the neuro-chemical and neuro-physiological changes that take place during conditioning.

Later, 'Skinnerian/Type-II/Operant/Instrumental Conditioning' was introduced which, yet, further developed the idea of stimulus-response conditioning, and founded many new hypotheses and terminologies. Unlike classical conditioning which was mainly involved with the reflexive and automated modes of behaviour; here, other more complex criteria were also postulated by workers on this second type of conditioning to explain various facets of higher order behaviour.

The main distinguishing feature of Type-I/Classical Conditioning from Type-II/Instrumental Conditioning is that in type-I the stimulus which is to be conditioned occurs independent of animal's response, while, in type-II such stimulus ('reinforcer' or 'punisher') is dependant on the emission of some designated response. Generally, the essence of a conditioning experiment is to isolate a specific behavioural pattern, and observe the establishment of a bond between a previously neutral input (Unconditioned Stimulus) and an elicited output (Responses/Reflexes) of the animal. Such associations are governed by the proximity of input-output occurrence, and also by the desire of the animal to achieve or attain a favourable response.

An extension of simple Pavlovian conditioning notions has been the ideas of 'secondary' or 'higher-order' conditioning. This is the observation that a second neutral stimulus could be paired with a Conditioned Stimulus (CS) that had already been conditioned to produce a response, and later this secondary stimulus will itself evoke the response without the presence of CS.

There have also been interesting investigations of the so called 'Introceptive-Conditioning' or the conditioning of internal involuntary reflexes (e.g heart-beat) to some external stimulus. This type of conditioning has been demonstrated dramatically by experimental results - animals that had been trained to reduce their pulse-rate in return for the electrical stimulation of certain pleasurable areas of their brain, would, in fact, reduce their heart-beat to a fatal point.

2.4.7 CONCEPTS ON CONDITIONING AND STIMULUS-RESPONSE THEORIES

In addition to the learning theories which were introduced by the rigorous investigations of the phenomenon of conditioning, many other S-R learning theories have been put forward, namely by: Thorndike, Hull, Guthrie and Estes; and generally, from the reductionist-behaviourist point of view. The following are some general principles which have emerged from the various postulates on conditioning and other S-R theories:-

(i) - THE LAW OF CONTIGUITY

The principle of 'association' or 'connectivism' is one which prevails in all S-R theories, it is the assertion that a connection is created between experiences which occur closely together in time or space. The extension of this law to the classical conditioning ideas was to attribute the bonding of Conditioned-Stimulus (CS) and Conditioned-Reflex (CR) to the temporal proximity of their occurrence.

Experiments on the relationships of the elapsed time intervals between the Conditioned-Stimulus (CS) and the Un-Conditioned-Stimulus (UCS) have shown that a time lag of under one second between CS and UCS is optimal in establishing the corresponding bond. In other words, the rewarding stimulus to the animal should, optimally, follow the occurrence of the stimulus to be conditioned in little less than a second. The associative feature of experiences, as outlined by the law of contiguity, has been further highlighted by the 'inferential' nature of stimulus and response linkage in the higher-order conditioning.

(ii) - THE PRINCIPLE OF REPETITION

The conditioning experiments have shown that the repetition of a CS and UCS pairing strengthens the CR. This observation is universal in all types of learning.

(iii) - THE CONCEPT OF GENERALIZATION

In the process of conditioning, the response once established to a particular stimulus, will be evoked by a stimulus similar to CS. This is called 'Primary-Generalization' to distinguish it from the learned similarity of two stimuli, which is called 'Secondary-Generalization'.

(iv) - THE CONCEPT OF DISCRIMINATION

This is the complementary concept to generalization. 'Discrimination' is the process by which stimuli are differentiated, and the desired ones filtered out from a range of stimuli.

(v) - THE PRINCIPLE OF REINFORCEMENT

This concept, in certain forms also known as the 'Law of Effect', is the process essential for the strengthening of associative bonds between experiences. Reinforcement is, basically, about the role of 'reward' and 'punishment' in the consolidation or the weakening of associations. It is the general consensus that animals would learn to do things better if they were rewarded with food, water or other rewards.

The major ideas of reinforcement were developed from the work on the Skinnerian type Conditioning, it was discovered that both 'positive-reinforcement' (reward) and 'negative-reinforcement' (punishment) were influential factors in motivating an animal to learn. Partial-Reinforcement has also been investigated, and attributed with the interesting property of establishing a stronger S-R bond than a total reinforcement. 'Goal', 'motivation' and 'drive' are relevant notions here, since many psychologists believe that these are the forces which activate an animal's learning. Although, these concepts could be correlated to some physiological mechanisms, yet, they can only be inferred empirically from the performance of an animal. In the language of behaviourism; it can be said that an animal is 'directed' towards 'learning' by striving for a 'goal', and the energizing source for 'motivating' such behaviour is the 'drive'. Here, some workers, in attempting to explain the nature of the initiating forces of learning, have speculated that 'drive-reduction' is the primary reinforcer in the learning process, and that a stimulus is bonded to a response if and only if the animal's drive and motivation level are reduced after such a response.

(vi) - THE LAW OF INTERFERENCE

This law covers the case of 'Forgetting' and the 'Extinction' of conditioned/learned responses; it states that a learned or conditioned pattern of events or behaviours may be weakened or inhibited by similar learning endeavours. The permanent nature of learning mentioned earlier is also emphasized by this law. The typical gradual extinction of a conditioned/learned responses is believed to be due to 'inhibition' and not the result of an irretrievable loss. This can be demonstrated by a so called 'spontaneous-recovery' of a conditioned-behaviour which had been forced into a complete apparent extinction by the retraining of the animal – the conditioned-behaviour is, normally, restored to half its original consistency after a rest period.

2.5 THE COGNITIVE APPROACH TO CONCEPTS AND ISSUES IN LEARNING

During the development of behaviourist explanations of the learning-process, all knowledge had been considered to be reducible to stimulus-response bonds. This belief and its apparent inability to explain many complex issues arising from the investigations of learning had impelled many researchers to voice their criticism to the mechanistic view of the subject. Cognitive-Scientists refuted this observational and empirical emphasis in the analysis of the learning-process, and introduced a new outlook towards this subject, the key notions being Information, Perception and Organization of Knowledge.

In trying to tackle the problem of learning from a different angle, there have been many re-definitions of the behavioural concepts using the organizational approach, as well as the introduction of a number of new criteria. However, in certain cases it is difficult to appreciate the validity or the necessity for separate terminologies. Numerous concepts within cognitive and behavioural outlooks refer to identical properties of learning and many underlying principles are common to both views. The distinctions are seemingly due to differences in:-

(1) – Levels of observations,
(2) – Interpretations put on empirical results,
(3) – Approaches to various problems,
(4) – Preferences given to specific experimental methods.

A typical behaviourist's experiment is that of an experimental animal involved in a maze-running or lever-pressing task, while a cognitive psychologist typically engages in experiments on human subjects in recall or recognition of a string of symbols. In the following without pinpointing the underlying critical or diametric views, or the detailed experimental evidence we will briefly enumerate the various definitions and concepts that have emerged from the study of learning by cognitive-scientists.

2.5.1 DEFINITIONS OF LEARNING

The question of 'learning' in cognitive-sciences is approached in a fundamentally similar way to the behaviourist's view, it is considered that learning should possess the same principle attributes of:-

(1) - ENDURANCE (2) - ADAPTABILITY (3) - CONTROLLABILITY

The holistic nature of outlook to problems has led the workers in this field to define the process of learning in more abstract terms. Sommerhoff (1974) emphasizes this point: "The main correlation established during learning appears to be between objects and goals rather than between specific neural inputs and outputs."

2.5.2 PERCEPTUAL CONCEPTS IN COGNITIVE-SCIENCES

The cognitive-sciences have historically evolved from the Gestalt school of thought and have inherited many established principles of the theories of Perception. The main underlying principles of perception are:-

(i) - THE LAW OF FIGURE-GROUND: A distinction is made between a figure and its background, using the perceptual features of it's boundaries, this law can be equated to the 'concept of discrimination'.

(ii) - THE LAW OF SIMILARITY: Items that have some common features are grouped together, this law can be equated to the 'concept of generalization'.

(iii) - THE LAW OF PROXIMITY: The elements of a field of objects are grouped together according to their nearness to one another, this law can be equated to the 'law of contiguity'.

(iv) - THE LAW OF COMMON-DIRECTION: A set of percepts will tend to group together if they appear to be in a sequential order or points on a simple curve.

(v) - THE LAW OF SIMPLICITY: The perceptual information is divided and organizes into simple and regular figures.

The 'Trace-Theory' of perception emerged from these principles, it was an explanation for the neural processes which are active during perception. This theory states that: "The perceptual experiences are reflected in neurons as an enduring trace, in the same form as the original neural-impressions of such perceptions". Ideas or events are associated not by the formation of a bond between two or more separate entities but by the creation of a new 'whole' unitary percept. 'Recall' or 'Remembering' is the reactivation of a given memory trace. A trace continues to exist in the nervous-system and can be retrieved by the selection of the appropriate cues, and the subsequent amplification of the trace. 'Forgetting' occurs in either the retrieval-failure of the trace, or the disintegration or modification of it. Trace-theory and other theories of perception and cognition have been instrumental in the formation of a cognitive explanation of the learning process. Various notions to the behaviourist's concepts of contiguity, reinforcement, repetition, interference, discrimination and generalization have, hence, been developed; and, while some of these redefinitions are only a different way of construing a concept the cognitive approach has, indeed, shed light on some specific aspects of learning, namely that of memory-formation, knowledge-acquisition and the concept of 'expectancy'.

2.5.3 STIMULUS-STIMULUS (S-S) THEORIES OF LEARNING

An alternate class of explanations of the learning-mechanisms, namely, the Stimulus-Stimulus (S-S) theories have emerged from the cognitive school of thought. The pioneers of learning research within Gestalt idealogy such as Koffka and Köhler founded the lines of scientific work on the subjects of insightful-learning and problem-solving. This was later followed notably by Tolman's (1959) work on cognitive aspects of learning, especially Latent-Learning (solving detour-problems and cognitive-mapping by animals).

The resulting S-S explanations of the adaptive-processes which were involved in these areas of learning research were contrary to the Hullian S-R theories, the Pavlovian-Conditioning notion of stimulus-substitution is possibly more akin to these S-S ideas. The S-S theories state that: "The perception of relationships between stimuli and the anticipation of consequences are the primary functions in the modifications of behaviour during learning, also, association is made between different stimuli rather than a stimulus and response".

The two varying S-R and S-S explanations of a learning-process is illustrated in FIGURE 2.1, where a Type-I/Classical Conditioning task (e.g., conditioning of a reflex response) is analyzed using these two interpretations.

TYPE OF EXPLANATION	PHASE 1 Pre-Test State	PHASE 2 Training-State	PHASE 3 Testing-State
S-R	UCS → UCR	UCS → CR	CS → CR
S-S	(UnCon.) (Stimulus) (UnCon.) (Response)	+ Neutral Stimulus (Con.) (Resp.)	CS=UCS → CR

FIGURE 2.1. The S-R and S-S explanations of a typical conditioned reflex experiment, the principal difference of these two theories are in the final phase of learning. In S-S the Conditioned-Stimulus and the Unconditioned-Stimulus are thought to have functionally fused together, while in S-R a new associative bond is thought to have been created between CS and CR.

2.5.4 MEMORY

The proponents of conditioning experiments had contemplated the question of memory-formation. In their view a memory of experiences or events was created and organised, and was an intrinsic and integral part of the learning-process. The explanations of learning-mechanisms involved in the neural level was based on the Stimulus-Response (S-R) notions, the essence of these theories was the consolidation of associative bonds between stimuli and responses.

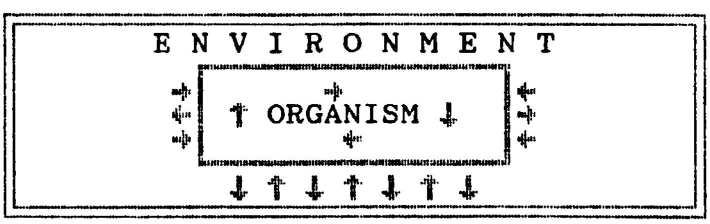
The S-S approach to memory-formation has culminated in a host of theories and findings in the cognitive sciences. To elaborate some of the issues that are involved, we will attempt to analyze a hypothetical learning process.

Any organism or animal as an entity could be thought of as having three general modes of activity:-

(i) - 'Inputs' (sensory or physical) from the environment to the body.

(ii) - 'Outputs' from the organism which brings about changes in the environment (or its movements).

(iii) - 'Internal Processes' occurring within the body of the organism.



Within this universe of actions learning is an isolated phenomenon, with some of its attributes enumerated earlier. The existence of a persistent sensory-impression of experiences at the neural level, and the relevance of such a storage of information to the learning-process, was intuitively taken for granted by early researchers. The principles of 'directiveness' and 'repeatability' of a learning-process are the main witnesses in support for such a 'Memory-Trace' or 'Engram'. The presence of a storage for information in the brain can also be inferred, by the subsequent testing of the performance or the knowledge of an animal that has undergone learning.

If we take for granted certain a-priori concepts regarding the environmental constraints and the internal modes of an organism (e.g., the feasibility of response, motivation, drive, attention), then a tri-level explanation of a typical learning process could be depicted by the sequences of TABLE 2.2. Whereby, the various explanatory stages of verbal-descriptive, state-of-knowledge, and memory-trace are outlined.

TABLE 2.2. The different phases involved in a general task-learning activity, the corresponding states in the knowledge level and the memory-trace modes are shown.

TIME	LEARNING STAGES	STATES OF INFORMATION	MEMORY-TRACE MODE
PHASE 1	PRE-LEARNING	PRIOR KNOWLEDGE	PRE-LEARNING TRACE
PHASE 2	INITIAL LEARNING OR TRAINING	ACQUISITION OF INFORMATION	TRACE-FORMATION
PHASE 3	END OF PRIMARY LEARNING	PROCESSING AND STORAGE OF INFORMATION	TRACE-RETENTION
PHASE 4	TESTING	RECOGNITION AND RETRIEVAL OF LEARNED INFORMATION	TRACE-RECALL
PHASE 5	RE-LEARNING	REHEARSAL AND REFINING OF LEARNED INFORMATION	TRACE-CONSOLIDATION
PHASE 6	FORGETTING	TOTAL OR PARTIAL LOSS OF STORED INFORMATION	TRACE-DECLINE
PHASE 7	USE OF LEARNED SKILL OR KNOWLEDGE	EXECUTIVE MODE OF INFORMATION	TRACE-UTILIZATION

The explanations of TABLE 2.2 is by no means a blue-print for all learning processes, various phases could be omitted, repeated or may occur in different order, depending on the experimental or specific environmental circumstances.

2.5.5 DIFFERENT TYPES OF MEMORY

In the micro-level of observation of a learning-process, all the elements of the brain (i.e., the 'neurons') may seem to be functioning, in a continuous and seemingly random mode of electrical activity. However, in the macro-learning sciences various explanations and definitions have been devised in an attempt to separate the brain into areas whose functions could be associated with specific cognitive processes (such as 'retention', 'retrieval' or 'forgetting'); also, different types of memory-storage with distinctive mechanisms have been elaborated.

Referring to TABLE 2.2, the phases 1 and 2 which refer to the static and transitional learning stages of memory-trace will not be discussed in more detail here, since the acquisitional aspects of learning were covered earlier in this chapter.

The investigation of memory-trace retention, the phase 3, has resulted in the categorization of memory into different groups according to the type of information stored. These various types are 'Semantic-Memory', 'Spatial-Memory', 'Sensory-Perceptual-Memory' (audio, visual, touch, etc.) or Procedural-Motoric-Memory (skills).

A further classification of memory is, also, made into 'Iconic-Memory', 'Short-Term-Memory' (S.T.M.) or 'Long-Term-Memory' (L.T.M.), according to the duration of information retention within each type.

(1) - Iconic-Memory: This is described as a very short lasting visual memory, iconic-memory refers to the temporary retention and recall capabilities of visual and nervous systems when exposed to a brief visual percept.

(2) - Short-Term-Memory: This type of memory is generally referred to the temporary storage of perceived data for future evaluation, utilization or disposal. The dominant feature of S.T.M. is the fast decay time of the information. In the case of animal experiments involving place-learning or latent-learning, it was noticed that an animal that was allowed to see the placing of a food object in a room established a S.T.M. of that event, and if after a fairly short period of time the animal was released in that room, it displayed a 'delayed-action' by remembering the location of the food and subsequently searching for it. Other experiments on human subjects have established the capacity of an average human's S.T.M. to be 7 ± 2 symbols, letters, numbers or words.

(3) - Long-Term-Memory: This type of memory contains a very large volume of information, experiences of an animal's life are accumulated in this type of memory as well as the semantic-memory in the case of humans. The information stored in the primary level of S.T.M. is thought to be transferred to this higher and more permanent level. The factors contributing to or inhibiting such transfers are motivation, arousal and attention of the animal and the novelty, significance or repetition of the experiences. The rehearsal of the learning material in phase 5 of TABLE.2.2, facilitates the transfer of information from S.T.M. to L.T.M.

2.5.6 DISTINCTIONS BETWEEN S.T.M. AND L.T.M.

The categorization of memory into L.T.M. and S.T.M. has been under constant scrutiny by scientists. Yet, many critics have pointed out that there is no conclusive evidence to justify such a distinction. This is mainly in view of many similar and common principles which are observed in retrieval, interference or decay aspects of both memories.

The physiological evidence of partially damaged brains or lesions of parts of an animal's brain has indicated that such damage could impair one kind or other of memory, but whether this conclusively proves the dual nature of the memory mechanisms and locations is not clear yet.

For simple associative learning tasks in lower animals the 'plasticity' of the brain as a homoeomorphic entity has been demonstrated. The technique normally used is to observe the effect of lesions to various areas of the brain on learning or recall abilities of an animal. It has been shown that there are no specific regions in the 'cortex' (grey matter) of the brain, which are uniquely responsible for the acquisition or the retention of this type of learning.

The 'holographic' nature of memory-storage has, also, been demonstrated by noting that the removal of a percentage of an experimental animal's cortex does not extinguish a specific learned task, but only results in the reduction of resolution of certain aspects of the animal's performance, speed or learning-capacity. In other words, such dissections lead to the reduction of the resolution of memory, rather than the complete loss of chunks of information.

It has also been noticed that the total loss of the 'hypocampus' of an experimental animal's brain, only inhibits the transition from S.T.M. to L.T.M., and the prior contents of L.T.M. or the current information in S.T.M. is not directly affected.

The results of these and other experiments have shown that, although, distinctive memory processes seem to be involved, physiologically the existence of separate brain locations for the two memory mechanisms has not been positively verified. The main objection to the methods involved in this type of work being the interference of such experiments in the actual mechanisms under investigation.

2.5.7 REMEMBERING

The current cognitive-psychology research has been mainly involved with the organizational aspects of memory formation and retrieval. Remembering or recall of an experience involves procedures of 'recognition' and 'retrieval' of memory-traces.

In the act of recognition the animal given a multiple choice of stimuli chooses one which has a particular set of cognitive attributes or cues. In higher learning processes, such as concept-learning or problem-solving, this task extends to a 'search' for cognitive rules or procedures.

The retrieval of information from memory has also highlighted a distinction between S.T.M. and L.T.M. Experimental results have shown that, the retrieval of data from S.T.M. is done by duplicating or replaying the original perceptual experience, for example, in the recall of a set of words which are stored in S.T.M., only the sound is remembered and not the meaning, while in recalling from L.T.M. the meaning or the context of an experience is remembered.

The process of information retrieval calls for a 'search' of the memory-store for the appropriate attribute, the procedures involved in this search could be one of a number of conceivable alternatives, such as serial/parallel or holist/partist. The experimental results in this direction point to the search in S.T.M. being done by a Successive-Scanning of memory in search of target information. In L.T.M., also, a hierarchal and contextual ordering of information, and search, has been demonstrated which indicates that data is structured according to its meaning or context, and experiences or inputs to a learning organism is classified in groups of similar percepts.

2.5.8 FORGETTING

Finally, in phase 6 of TABLE 2.2, the memory-trace formed during learning and relearning is subsequently diminished by the passage of time. There are five major explanations of this change as follows:-

(i) - Decay Theory: This is possibly the simplest explanation. It considers the decay of the memory simply as the erosion of blocks of storage by time, and although it explains the reduction of recall at the S.T.M. level quite adequately, it does not account for the persistence of some trivial information or skills over extremely long periods.

(ii) - Trace Change Theory: This theory attempts to explain forgetting as the result of a gradual rounding-off or generalization of various details of memory of an event, and not the loss of strength of a memory-trace. The shapes of percepts are thought to, either change into familiar objects, or evolve into simpler shapes. Hence, although some novel or unique events are preserved almost in their original form, the mundane or routine events are rounded-off into their nearest typical occurrence. Trace-change theory is not unanimously accepted by all learning researchers, it started from the perception-theories of Gestalt Psychology, and it presupposes a continuous rehearsal of memory-traces. The experiments used in proving this type of theory are normally based on the successive evaluation of the learned-material, this is done by comparing the reproductions of the original learning at different time intervals. The main criticism to this type of experiments is that the forgetting could be simply due to the distortion of data during the continuous rehearsals or reappraisals of information.

(iii) - Retrieval Failure: This phenomenon indicates that forgetting is not due to the loss of the learned-material, but is the result of a failure in the retrieving of information from memory. The experiments conducted in support of this theory show that a memory which seems to have been lost can in fact be recalled by the introduction of new retrieval cues.

(iv) - Intentional Forgetting: Motivational influences on forgetting have been shown to have a significance in extinguishing memories associated with painful or unpleasant experiences, this type of forgetting is most commonly evident in individuals suffering from repression, also in controlled experiments people could intentionally forget a learned memory by consciously deciding to dispose of the contents of a S.T.M..

(v) - **Interference Theory**: Historically, this theory originated and was established within the behaviourist school of thought, the interference aspects of learning has been exhaustively investigated and a great deal of supporting experiments exist. The observation that two similar experiences over a period of time tend to interfere with the retention of each others memory-trace has led to the development of 'Interference-Theory', which in fact is comprised of two sub-theories of 'Retroactive-Interference' and 'Proactive-Interference'.

(a) - Retroactive-Interference: The main claim of interference theory is that, memory once established does not decay or become lost, but is only displaced as a result of some new similar material. Hence, mere passage of time does not contribute to forgetting. Retroactive-Interference is a phenomenon which states that newly learned material can replace the memory of previously learned similar material.

This observation has been demonstrated by experiments in which half of a group of subjects in an task-learning experiment participate only in the first part the experiment, and are rested in the second part. While, the remaining subjects continue with the learning experiment. The relative high recall ability of the rested group in remembering the material which was jointly acquired, shows that the continued training has in fact interfered with the previously learned similar material.

(b) - Proactive-Interference: This is the complementary notion to the one above, it describes the effect that previous experiences or memories can have on the acquisition of new material. This type of interference is also demonstrable by empirical results. By using similar techniques to retroactive-interference experiments, it has been conclusively shown that, the capacity of learning a particular type of material may diminish, as the result of interference from previously learned similar material.

2.5.9 OTHER ISSUES IN COGNITIVE SCIENCES

As pointed out earlier, cognitive sciences have been investigating primarily the complex modes of learning-behaviours in higher animals and humans beings. Some existing concepts have been further developed and other new ones introduced. For example, the introduction of cognitive-search methods, cue-selection procedures involved in concept-learning, and also in insight and latent learning the idea of 'expectancy' of an event are some of the notions put forward in this paradigm. These issues seem to be relevant to our discussion of learning and will be discussed here.

2.5.10 PROCEDURES AND STRATEGIES IN CONCEPT-LEARNING

The learning of concepts is basically an attribute of man, although it has been shown that some higher animals possess conceptual capabilities in learning symbols or audio-visual commands. Yet, to conclusively demonstrate the notion of concept-formation, it would be necessary to know what an animal thinks, hence, language capability could be considered an implicit component of this type of learning process.

One of the key issues in concept-learning is, the methods involved in the search of attributes/factors/cues of a concept in a specific context. The procedures for the 'search' and 'recognition' of these cognitive-cues bears a close similarity to the activities involved in the brain in 'recognition' and 'retrieval' of memory-traces. Experimental results of the investigations in this area have shown that the search and sorting of such attributes could be done in the following two principle ways:-

(i) - Reception/Serial Procedures: where a subject forms a hypothesis according to the successive encounters of the occurrence of an experience. To test such hypotheses there are basically two distinct strategies. First, the 'holistic-strategy', which is to remember as many occurrences of a 'significant' experience and to try to verify the hypothesis on the basis of such data. Second, the 'partist-strategy', in which the hypothesis is tested in each incidence of the 'significant' experience. Here, the 'significant' or 'positive' instances of a concept-rule are those events which reinforce the subject's actions - it may also be an instruction from a teacher, or a reward.

(ii) - Selection/Parallel Procedures: are the complementary notions to serial procedures. In this case, the experiences are confronted all at once. The task is to extrapolate the significant attributes from a selection of inputs. Here, also, principally two main strategies are used. First, the 'scanning-strategy', in which the subject forms a conceptual-rule (a set of attributes) and tests for the significant/positive instances of such a concept. This is done by successively scanning the context in a random manner, and by evaluating the hypothesis at each stage. The second strategy is the 'focusing-strategy', which is to initially choose a general 'exemplar' with the highest positive conceptual attributes available. Subsequently, conceptual events are selected in such a way that each attribute can be tested. This individual verification and rejection of attributes, finally refines the general exemplar, and yields the desired conceptual-rule.

2.5.11 EXPECTANCY

The notion of 'expectation' or 'anticipation' is also another important concept which has emerged from cognitive-sciences. This issue will be elaborated here, the background to the development of such ideas outlined, also the prominent role of Tolman (1959) in this area emphasised, and his theories of learning using this criteria briefly described. The particular interest in this fundamental issue is in view of the possible applications of its concepts to the problem of designing simple 'learning' models, as outlined by the objectives of this thesis.

The survival of an organism requires the maintenance of certain physical inputs of the animal within tolerable levels, such as food, water, oxygen, etc. A notable feature of behaviourist's description of the learning-process is the underlying assumption that learning is initiated as a result of an animal's need to stay at its optimal physiological state. An explanation of a learning task achieved by a hungry animal, using the terminology of behaviourists could be as follows:-

The animal 'motivated' by a 'drive' for the 'goal' of attaining food is 'emitting' an 'exploratory' behaviour, this 'appetitive' task could bring about, by chance or by experimental design, certain events which will 'reinforce' the animal's behaviour. The 'contiguity' or the 'drive-reducing' properties of such events helps the animal to learn and remember the behavioural sequences leading to these events for future use. The higher the motivational level or the drive of the animal the better the quality of learning.

The 'goal' in such learning-tasks may be an object which is acted upon or ingested, or it could be the execution of a certain behavioural pattern, or, more generally, it can be considered as a change in the stimulation of the animal. New goals and sub-goals may be learned if they are instrumental in attaining the initial goal. The organization of such goals are considered to be in a hierarchal form in the order of survival-value priority.

Similarly, the motivation could be a physiological need (e.g., food, water, pain-aversion), an acquired or learned motive (e.g., in experiments on primates, chimpanzees can learn to work for symbolic tokens which could be later used to acquire food), or could be a motive with not a very obvious purposive nature (e.g., the manipulation of toys by infants as noticed by Piaget (1977), or a general need for perceptual change in stimuli).

The emphasis on the 'purposive' and 'goal-directed' nature of learning behaviour is, also, a major feature of the cognitive theories of learning such as Tolman's. He pointed out that: (1) - behaviour should be analyzed at the level of purposive actions rather than movements; and (2) - behaviour is "docile", in the sense that it should be considered adaptable to various changing circumstances, and the means available to the animal determine the choice of actions it takes. In his view, the actions of a motivated (e.g., hungry) animal are determined by the so called cognitive-mapping, which is the use of the knowledge of the spatial paths and other perceptual aspects of its environment. The organism is capable of putting together such information, or inventing a solution for reaching the goal event. The current behaviour is guided by the belief of the outcome of the actions rather than a reflexive response to a given goal.

The analysis of behaviour of a hungry animal could be done at two levels, on the one hand it may be considered as a collection of muscular responses and on the other hand it can be looked at as a sequence of purposive mental events. However, as mentioned earlier, like most complex phenomena an organism's behaviour may be described at several different levels, and the distinction here seems to be the degree of emphasis given to the physical or mental aspects involved.

Tolman was a proponent of the temporal and mentallistic approach to the explanations of behaviour. He had the view that the organism engaged in a deliberate reflection about problems. Such notions, although familiar in human common sense, had a metaphysical flavour which resulted in a critical reception of his ideas by other scientific researchers of behaviour. Hence, in pursuit of a more objective explanation of his ideas he engaged in the investigations of the problem-solving abilities of animals, and other cognitive processes involved in learning. Although, human subjects would have been intuitively a more suitable domain for this type of enquiries, he mainly used animals for his experiments.

In doing such research Tolman laid down the basis of an elaborate learning-theory. He postulated that the internal representation of events 'signified' 'what-leads-to-what'. This representation was a strictly un-observable mental process, and it involved the 'anticipation' of stimuli and the 'expectation' of an event occurring. He emphasised the distinction between the 'acquisition' of learning and 'performance'. At the physiological level he favoured the S-S interpretation of the learning-process.

Some of Tolman's experiments were based on observations of place-learning and latent-learning abilities in animals, these experiments involved the remembering of mazes or spatial paths by the organisms with no apparent immediate rewarding or reinforcing agents.

The 'Law of Effect' which was introduced by Thorndike at the end of the 19th century, and also the later 'contiguity' principles, stated that, behaviour is only learnt if it leads to the satisfaction of certain motivating conditions. Hence, their explanations of learning-processes could not wholly account for the observations made by Tolman. Therefore, he introduced the "principle of expectancy" which although not contradictory to the law of effect, yet, attempts to explain the learning-process at a deeper level. The 'reward' not only reinforced a specific response, but also went one step further by reinforcing the expectancy of a stimulus leading to such a response. According to Tolman's view, "learning takes place by confirmation of expectancies."

These theories explained the results of a class of experiments on primates in which monkeys having previously seen the placement of a fruit (banana) under a box upon the subsequent release into the environment expected to find the goal-object by lifting the box. But, if the banana was replaced with a different fruit (lettuce), then the animal's reward expectation of the first kind of fruit was not confirmed, and it rejected the second reward and carried on looking for the original goal-object, although it would have normally accepted the lettuce as a primary reward. Habits and their formation is an important aspect of human behaviour in which the ideas of expectancy and the confirmation of such expectancies play a prominent part.

For Tolman, the reward is not a necessary element of acquisition of learning, and the law of effect is not a universally valid principle for explaining adaptive learning changes. His principle learning theories based on the notion of 'what-leads-to-what' stated that:-

(a) - Two or more stimuli events are combined in S-S relations to form the basis of knowledge and.

(b) - A stimulus S1 its response R1 and another stimuli S2 that follows R1 form a three-term expectancy, written as S1-R1-S2. These expectancies are strengthened if the subsequent occurrence of S1-R1 sequence is followed by the expected stimulus S2. Similarly, it is weakened if S1-R1 occurred but were not followed by S2.

An interesting feature of expectancies was that, they could be integrated into larger strings of expectancies (S1-R1-S2-R2-...), such as in learning a maze, the animal connects the expectancies at the choice points, it is subsequently able to amalgamate these sequences into a single chain of actions, which can be activated at the start of the maze.

Another issue raised from this line of research was the inferential nature of thought processes, it was believed that an animal with a S1-R1-S2 expectancy, and engaged in its appetitive exploratory behaviour, learns associative relationships between pairs of events such as S2 and S*. It subsequently makes an inference for a new expectancy of S1-R1-S*. This is a process by which the discovery of perceptual associations by an organism (such as S2-S* pairings) at the cognitive level can help to solve a particular problem, and effect the behaviour at the response level. Using this method the animal acquires a large number of Sx-Sy and Sm-Rm-Sn connections. This knowledge base could be organised into a sort of neural or cognitive map, by which the animal is able to find the shortest or the most economical chain of actions for attaining a desired goal.

Tolman used the metaphor of brain being a map-control room rather than a telephone-exchange, where the incoming stimuli are not connected by single one-to-one switches to the outgoing responses but are turned into a tentative cognitive like map of the environment, which determines what responses if any the animal will finally release.

The main criticisms of this type of research have been: firstly, their subjective and mentallistic nature; and secondly, the lack of structural theories which can specify ways that these expectancies could lead to the appropriate actions. However, many important issues have emerged from this line of work, especially regarding the internal representation of the dynamics as well as the statics of the environment in the brain. The main question in this respect is: how can the external flow of events be symbolised by internal processes?

A proposal was made by Sommerhoff (1974) regarding the nature of changes involved in the transition of external dynamics to internal representations, on a more objective grounds. Using the concepts of 'directive-correlation' and the notion of 'anticipatory-reaction', the following speculations was made, as to what kind of processes could exist in the internal mechanisms of expectancy-formation.

(a) - It is possible that the internal representations of expectancies bears no direct correlations with behaviour, and the internal processes involved in the realization of such expectancies work at a different level.

(b) - The internal expectancies could be represented by a stochastic process, based on the probabilities of events and outcomes and the consolidation or weakening of such notions.

(c) - A deterministic model of expectancies is also conceivable, involving the creation of isomorphic associations between internal representations of events and their outcomes.

One of the main contributions of Tolman's theories to the contemporary science of learning has been to emphasise the two distinct levels at which the learning-process should be investigated. A learning animal is constantly making a mental model of its behaviour by constructing expectancies and neural-maps; and, also, by using its behavioural responses and stimuli from its environment it is able to confirm or reject these expectancies. Hence, this dual "processing + executive" nature of explanations has become the prevalent feature of many of today's information-processing models of the learning-process.

2.6 SOCIAL ASPECTS OF LEARNING AND TEACHING

Some relevant issues of social-behaviour, social-learning (in animals and humans), 'intelligence', and the general area of education and the impact of the learning theories on teaching will be discussed here.

2.6.1 SOCIAL-LEARNING

Learning-processes which have been discussed so far have been mainly the outcome of an individual organism's interactions with the environment. These adaptive behaviours were emitted to enable an animal to survive. However, such behaviours almost invariably occur in a social context. The social organization has its own evolutionary survival advantages in facilitating reproduction, rearing of the young and providing defence against predatory and other challenges of the environment. At the same time, it contributes to emotional stress, disease and conflict. Social behaviour is the interaction/influence of individuals on each other. Various elements of social-behaviour are:-

(i) - Reproductive-Behaviours: To some degree a social element exists in this type of behaviours, for most species.

(ii) - Developmental-Factors: The critical inputs to the organism, that play a major part in the setting of the pattern for social roles. Physiological examples are the hormones that determine morphology, and psychological and perceptual examples are the events that determine response tendencies, such as the factors involved in 'imprinting' or the setting of social-dominance hierarchies.

(iii) - Communication: Most animals have means of communicating between individuals of species, varying from simple physio-chemical signals to complex symbolic processes such as language.

(iv) - Territorial-Arrangements: The aspects of social behaviour influenced by animals' breeding, habitat or food-source locations.

(v) - Characteristic Species-Specific Behaviour: These are the actions which are determined on the basis of particular ecological factors surrounding a species; such as maternal care or family organization.

A study of ethological and behavioural literatures reveals the variety of behavioural adaptations that exist in the social level of animal's behaviour. Examples for each of the above categories are: the nest-building and courtship-behaviour of birds; the imprinting of young birds to the early experiences of their lives; the ritual dancing-display of stickleback fish at the time of mating as a means of communication; the territory establishing behaviour of sea-gulls at the nesting time; and the food-begging behaviour of young chicks. The intriguing feature of all these examples is the intricate balance which is kept between the learned/environmental and the innate/genetic/hormonal contributions.

In the case of humans the theories of the social-learning have been developed mainly within the science of social-psychology. This type of work generally has a combined cognitive and behaviourist outlook, and it analyses the learning, motivation and reinforcement of social behaviours in terms of cognitive events intermixed with external behavioural events.

The emphasis in the theories of social-learning is on learning by observation, rather than learning by doing or personally experiencing. In social-psychology and education the development of intelligence, and the 'Intelligence Quotient' (IQ) have been major topics of investigation; also, a matter of special interest has been the heredity properties of intelligence.

The relation of intelligence or IQ with learning ability is a controversial issue and has also been rigorously investigated. It has been shown in some experiments that in certain instances the IQ can be correlated to a specific learning ability of a subject. However, this is not a universal observation, and is dependent on various environmental backgrounds or the training of individuals.

Another fact which has emerged, is that the learning-proficiency can not be transferred between tasks with different basic elements. For example, learning can not be transferred between manual and formal tasks.

The scientific study of higher forms of learning such as problem-solving, concept-learning, and language-learning has established yet another area in learning sciences, that of the instructional aspects of social-learning, namely, 'teaching'.

2.6.2 TEACHING

One of the most important contributions of the findings in psychological learning-sciences has been to the development of instructional theories and teaching-programmes. Although, the work of many researchers on learning deals with its purely scientific aspects, some investigators have ventured into studying the ways this knowledge may be utilised in education. 'Teaching' or 'training' can be viewed as the control of a learning-process, or alternatively as the transfer of knowledge or skill.

The investigations of learning as highlighted by the variety of theories, approaches and views, and by the even less precise understanding of its mechanisms, has not yet yielded a unifying descriptive science. Hence, as anticipated, many consensus also exist in the field of teaching-theories.

Additionally, the validity or effectiveness of many contemporary teaching techniques are questionable, this is in view of their inability to improve the 'quality' of the learning, rather than simply providing environmental facilitation for learning or recall. The earlier theories of instruction were embodied within the 'curriculum theories' of the science of education, these theories basically promoted the presentation of the learning-materials in some orderly fashion. In the past 30-40 years programmed-learning has been introduced, in the development of which experimental-psychologists played a major role.

Recently the accessibility of inexpensive computers and other audio-visual aids, has facilitated the introduction of teaching-tools such as simulators, films, TV, video, tape-recorders and computer assisted instructions, to many schools and educational establishments.

The impact of this mechanization of teaching has been to increase the importance of this field of study, and also to emphasise the need for an exact theory of instruction, which will help the devising of automatic programmed-teaching strategies.

Cognitive and S-R theories as well as some notions from motivation and personality studies, have contributed to the following ideas which are currently used in various teaching theories.

(i) - MOTIVATIONAL ASPECTS

Learning is best fostered if a goal or a hierarchy of goals are set for the learner, this could be a behavioural objective, such as obtaining a specific proficiency-level after a number of lessons. This type of goal is usually more effective than aiming for a general understanding or appreciation of a subject. Also, the interest of the learner should be aroused, by setting interesting tasks and goals with optimal degree of difficulty.

(ii) - PRESENTATIONAL ASPECTS

A task or a skill which is to be taught should be broken down into its appropriate components and organised into a hierarchy. Thus, the learning material should be ordered into sub-units and presented from simple to the more complex units. Also, the learned material is better retained and more easily transferred if it is fully understood, than if the same material is learned by rote-memorising.

The perceptual structure of the teaching material is also important, conceptual information could be learned better, by the proper organization or presentation of learned material and the placing of vital visual cues. The environment of the study or the classroom has to reinforce the relevant task-learning behaviour, poor concentration, poor attention, poor habits or other distractions are all interfering with the optimal teaching-processes.

(iii) - STUDY TECHNIQUES

Recital; practice; manipulation of learned material by responding or relating one part to another; asking of relevant questions; searching the text for answers; will all enhance the learning of material. These techniques also help to form automated habits which improve the performance in future (e.g., the learning of some mathematical rules). In studying a text various prescribed study guides are devised, these generally involve following a specific sequence of steps, using which a student can improve the learning or remembering of studied subjects.

Also various mnemonic or memory devices are elaborated, these are special techniques of making the material more meaningful and hence, improve the recall and learning abilities.

2.7 A TAXONOMY OF ADAPTIVE AND LEARNED BEHAVIOURS

In the following, a classification of different adaptive and learned behaviours will be outlined. No special phylogenetic considerations are made in this stage. The list is simply a taxonomy of adaptive activities of animals in order of their behavioural complexity, as defined in behavioural and cognitive sciences. Also certain adaptive observations may be accounted for in more than one of the specified catagories.

(1) - INSTINCTIVE BEHAVIOURS: The simplest kind of adaptive behaviours in animals, instincts are inborn-response tendencies which provide automatic mechanisms for adapting to the recurring situations in life. Instincts are normally complete in their first occurrence and are triggered by a specific stimulus. There is normally, an element of learning or modification present in most instinctive behaviours.

(2) - IMPRINTING: A highly specific and limited form of learning, it could also be considered as an externally triggered instinctive behaviour. Imprinting can be clearly observed in birds, during the early periods of their lives. The young birds in a 'sensitive' period after hatching, will take any large moving object to be their mother, this impression will stay with the birds for the rest of their lives.

(3) - HABITUATION: Possibly the simplest type of true learning universally seen within different species. Habituation is the waning of normally occurring behaviour. An animal gradually decreases its normal response to a stimulus, which is repeatedly being exposed to. Although, habituation is similar in characteristics to fatigue or sensory-adaptation, it is different in the sense that, it persists over long periods of time. Habituation has the important survival function of, filtering-out the insignificant and mundane aspects of life.

(4) - CLASSICAL CONDITIONING: A very simple type of learning, involving associations between stimuli. This phenomena was originally observed by Pavlov, in experiments on the salivation response of dogs to food stimuli, it was noticed that, a previously neutral stimulus, such as the sound of a bell, could be bonded associatively with the food stimulus, and after an appropriate number of training procedures, the sound of the bell alone would evoke the salivation response.

(5) - INSTRUMENTAL CONDITIONING: Instrumental/Operant Conditioning, brought to prominence by Skinner, is essentially similar to Classical Conditioning, the main feature of this type of conditioning is the choice that an animal has over the stimuli it receives and the responses it makes. An animal that learns to depress a lever to get a food pellet makes an association between the food-obtainment and the lever-pressing, based on the reinforcement or the reward of its needs.

(6) - TRIAL AND ERROR LEARNING: The term "trial & error learning" was brought into wide usage by Thorndike, this type of learning is also referred to as 'habit-formation', it has the basic elements of instrumental conditioning, and normally, contains a subsidiary element of classical conditioning. Trial and error learning involves choosing successively a response from a collection of possible responses, and evaluating the consequences of such response for future utilization.

(7) - ONE TRIAL LEARNING: This type of learning is associated with Gutherian learning-theories, which state that, a stimulus-pattern gains its full associative strength on the occurrence of its first pairing with a response. Guthrie attempted to explain the subsequent improvement in learning, as a matter of refining the learning for each element of movement involved in a particular behaviour.

(8) - REPETITION LEARNING: The acquisition of skills or knowledge as a result of successive repetitions of a behavioural or perceptual pattern, is normally referred to as 'repetition-learning' or 'learning-by-doing', which could be thought of as a special case of trial & error learning, where there is no 'choice' of response for the animal.

(9) - ROTE LEARNING: This type of learning involves the memorizing of behavioural/cognitive patterns or response sequences and the subsequent display of such behaviours in appropriate situations (e.g., learning of nonsensical string of letters); in this type of learning the sound or the shape of stimulus is remembered and not its meaning.

(10) - SERIAL LEARNING: The extension of individual stimulus-response bonding to the sequential chains of such pairings, and the learning of associations in tasks involving such serial elements, are referred to as 'serial-learning', and was originally introduced by Ebbinghaus.

(11) - LATENT LEARNING: This type of learning, the study of which has been associated with Tolman, is the learning of spatial or other environmental information by animals, and although it takes place most of the time, does not exhibit itself under normal conditions of performance. Latent-Learning, is basically the association of indifferent stimuli or situations without a patent reward (e.g., the formation of memory-maps in normal foraging activities of animals).

(12) - INSIGHT LEARNING: Insight relates to the organizational aspects of learning, also, the perception of relationships. Insight Learning originally investigated by Kohler, is the solving of a problem by the reorganization of experiences or the sudden discovery of novel relationships.

(13) - PROBLEM SOLVING: Another category of insightful-learning, is normally associated with cognitive-sciences, and generally refers to solving problems by humans using semantic descriptive explanations, incorporating mathematical or logical arguments in the process.

(14) - VERBAL LEARNING: The learning of verbal information, in terms of associations based on ideas rather than the semantic lists of percepts.

(15) - CONCEPT LEARNING: The learning of conceptual relationships, by the discovery of list of common attributes in a context. Some of the methods involved in the research of this type of learning are:-

(a) - Oddity-Principle-Learning: Involves presenting several stimuli to a subject, the task being, to pick the one which is different from all the others in a particular attribute.

(b) - Paired-Associative-Learning: Subjects learn a list of discrete associations (e.g., pairs of symbols, one of the pair is used as a cue (stimulus) for the recall of the second (response)).

(c) - Learning-Sets or Learning-To-Learn: A specific concept or criteria is presented to the subject (e.g., the subject must extract the implicit conceptual-rule from a selection of objects). It has been noticed that higher animals subsequently tackle a similar type of problem much better, and possess an ability of 'how-to-learn'.

(d) - Free-Recall-Learning: Subjects after learning a list of items (serial or parallel), are tested for learning retention by being asked to recall the items in same order or randomly.

2.8 THE BRAIN AND NERVE-SYSTEM SCIENCES APPROACH TO ISSUES IN LEARNING

So far in this chapter we have looked at the manifestation of learning outside the body of an organism. The brief references to possible internal mechanisms have only been temporary digressions from the mainstream external approach, prevalent in both behavioural and cognitive sciences.

The behaviour of an animal in nature, and in particular those labelled as 'learned', can be explained in many levels of: social interactions; states of mind; changes in hormonal secretions or the levels of certain chemicals in the blood stream; levels of sensory inputs; the altered patterns of brain cells; new synaptic properties of a brain region; changes in protein synthesis or even the changes in the atoms composing the animal's body.

The immense complexity and diversity of the task of explaining any behaviour in absolute terms is obvious. However, since not all such information can be utilised all the time, the purpose for which a description is intended, dictates the level of explanation. The investigation at an inappropriate level of enquiry, not only will be redundant, but would not have the essential elements for the proper analysis of the problem.

The studies of the natural learning processes have, generally, been conducted in three major empirical levels: behavioural/executive, cognitive/mental and neural/mechanism. The neural level, admittedly, is the basis for both the behavioural and the cognitive levels. But, while in cognitive or behavioural psychology, normally, observations involve a single domain (i.e., perception or behaviour), in neuronal explanations a hierarchy of descriptive levels exist (i.e., system, single neuron, biochemical, biophysical, etc.).

2.8.1 SOME PHILOSOPHICAL ISSUES - HOLISM vs. REDUCTIONISM

An organism may exhibit complex and intricate patterns of activity, but, learning is said to take place only when previous experiences are utilised in the subsequent behaviours. The impressions of such behaviours at the neuronal level, perceived by the subjective reality of the animal's senses and transformed into electrical nerve signals, has raised many issues and controversies on subjects such as 'mind' or 'consciousness'. The legitimacy of such physical/mental relationships, and also the question whether this problem can be meaningfully tackled by man in a 'self-understanding' sense, has been the concern of philosophers and psychologists throughout the ages.

An important philosophical distinction, Holism vs. Reductionism has been made regarding the possible approaches to the study of the brain and the nervous-system. The Holist subscribes to the rationalist view that biological investigations of the brain should be carried out on the totality of the cells, and not on its elements or molecular components. The brain is considered to be 'more' than the sum of its interacting elements. The main emphasis in holism is the organizational aspects. The Reductionist, on the other hand, being from the school of empiricism, believes that the full understanding of the complete workings of the brain is only possible by the comprehension of its neuronal, molecular and sub-molecular components. The notion of 'mind' or any other mystical properties of the brain are rejected by the reductionist.

2.8.2 HISTORICAL BACKGROUND TO THE DEVELOPMENT OF BRAIN-SCIENCES

The localization of 'mind' or other psychological phenomena in the brain and nervous-system was only established from the late seventeenth century. The more ancient theories regarded the heart as the centre for 'soul', and the brain was considered as an organ simply for the cooling of blood. Later, the 'ventricular fluids' brain theories were introduced by Descartes who believed the soul was located at the brain-stem. For Descartes, the brain consisted of a ventricular system of canals and pistons which circulated the vital body fluids. The mechanically minded scientists of the eighteenth century replaced this model of the mechanisms of the brain with a clockwork model of cogs and wheels.

At the beginning of the 19th century, the science of 'Phrenology' was introduced, namely by Gall. The phrenologists were interested in mapping and labelling various areas of the cortex as responsible for particular mental faculties and moral values. Indeed, the issue of whether separate brain regions are responsible for specific functions, or whether they fulfill many roles; otherwise referred to as 'specifity vs. plasticity', has been a major point of controversy ever since.

The discovery of 'Electricity' and 'Animal-Electricity', led to findings on the neuro-physiological properties of animals' nervous-systems. Similarly, the use of low-power microscopes in biology facilitated the anatomical investigations of brain cells, and the discovery of 'neurons'. Much of the research which followed involved the electrical investigations of the brain and neurons, and the analysis of various portions of the brain. These investigations resulted in the introduction of a new class of models for the brain mechanisms, based on the electrical-networks of its elements. A popular analogy was the brain being a "telephone exchange".

The modern era of Brain-Sciences began with the invention of electronic recording and amplifying equipments, which have been progressively refined ever since their introduction. The discovery of the 'brain-waves' showed that, regular electrical activities could be recorded, with their frequency and strength depending on the various modes and levels of animal's activity, the patterns of such electrical recordings were rigorously investigated. Many other technological advances such as: the probing techniques of the brain using precision electrodes; the discovery of chemicals for staining or tracing of nerve fibers; and the discovery of electron-microscope have shed light on many anatomical and physiological features of the brain. More recently, also, the science of molecular-biology and neuro-chemistry have revealed many new chemical properties of the brain-cells.

The clinical study of different brain disorders and defects, both psychologically and biologically oriented, also the study of the interfering affects of various drugs on the workings of the nervous system, have been major contributing factors to a better understanding of brain mechanisms.

2.8.3 THEORIES OF LEARNING - NEUROLOGICAL APPROACH

Today, in learning sciences the behaviour of an animal is considered to be exclusively a product of its nervous system. Keeping this in mind, it is surprising that in the majority of definitions and theories of learning little reference is made to the neurological aspects. There may be possible historical or developmental reasons for this apparent neglect. For instance, until the early 1960's psychologists held the view that neuro-physiology had very little to offer which was relevant. However, with the explosion of neuro-physiological research within the past three decades, the gaps between the psychological and neuro-physiological explanations are narrowing.

Using the metaphor of "brain and computer" for analogy. Behaviour can be thought of as the output of the computer, in the form of data. On the other hand, the neurons and sub-cellular elements of the brain can be considered as the individual memory-cores, or the electrical charges in the electronic circuitry. The enormity of the task will be appreciated if we were asked to investigate the workings of a computer by only observing the movements of electrical charges within its memory storage. Conversely, it would be extremely difficult, if not impossible, to understand the underlying mechanisms of a computer from the analysis of its input/output data.

From the time of Pavlov, psychologists have speculated notions and theories regarding the nature of the physiological substrate of learning. In general, they hold the view that there are brain correlates of behaviour and learning; and have postulated that an organism has certain primitive neural mechanisms, with their specific functional rules stipulated by behavioural implications. However, not a great deal of interest is shown towards understanding the precise nature of the mechanisms of such 'engrams' or 'memory-traces'.

Neuro-physiologists, on the other hand, have been interested in the mechanisms that provide the animal with the diversity of behaviour. In their view the investigations of biological correlates of learning and meticulous probing of the hardware of the brain, will finally prove to be more rewarding than the behavioural investigations.

Another point to mention here is that, although, the neurological domain of investigation includes the sense organs, the muscles and the central-nervous-system (CNS). Yet, the main areas of interest in learning related biological sciences have been away from the peripheries, and mainly involving the examination of changes that occur in the CNS during a 'learning-process'.

2.8.4 METHODS AND TECHNIQUES OF OBSERVATION IN BRAIN SCIENCES

A study of the course of development of brain sciences highlights the point that technological advances play a major role in the understanding of brain-mechanisms. One of the first techniques used for the investigation of the brain was 'ablation', in which a portion of brain is destroyed or cut out and the resulting behavioural changes noticed. This technique was refined by the introduction of electrodes that could cause a small lesion at a more precise location of the brain. Another method, which does not entail the destruction of the nerve-cell is the stimulation technique, which involves using an external electrode or a permanently implanted electrode in the animal's brain (sometimes activated remotely by radio waves). Similarly, recordings of the brain's electrical activities can be made using sensitive electrodes which are capable of detecting the faint firings of a single neuron, or the wide-scale electrical activities of the outer skull (EEG). Computers are, also, used to filter out the background electrical noise present in the brain-wave patterns under observation.

The chemical stimulation (instead of electrical) and detection of nerve cell activity is also possible - whereby, delicate hypodermic needles are used. Neurally acting drugs under investigation can be injected in tiny amounts, and their effect on behavioural patterns or an adjoining neuron detected. Other methods of stimulation or lesion of brain cells are also utilised, such as laser-beam, cooling, etc.

Although, many important discoveries are made using these techniques the conceptual objections to methods such as ablation, lesion or stimulation are still valid. Specially, in view of our limited knowledge of the actual processes involved in the cellular levels of the brain, and the interference which these techniques might cause to the delicate mechanisms under scrutiny. Hence, these type of results should be viewed with caution and supported by the evidence of other experimental methods.

Another issue which is relevant to these types of experimental techniques is the subjectivity of the observation. The occasional experiments on human subjects, with the verbal accounts of the stimulated sensations, does not give a systematic and consistent picture. Conversely, animals can be systematically studied, but there is no verbal description of the induced sensations, and the inferences must be made from the behaviour of the animal.

It is, also, worth noting that the distinction in the 'holist' and 'reductionist' approach is also manifested in a dichotomy of methods used for the probing of the brain. The holist may scrutinize the patterns of brain-waves during a learning process, while the reductionist will look at the individual neuronal firings during such a process.

2.8.5 MODIFICATIONS OF THE BRAIN DURING LEARNING

The search for the 'engram', the neural substrate of learning experience has proven to be so far futile; the problem may well be due to the inappropriate way which the issue has been posed. The origins of speculative brain mechanisms of learning was embodied within the associative learning theories of physiological-psychology, as typified by Hebb's (1949) "reverberating circuits" explanation of the formation of neural associations. Such theories have been searching for neural mechanisms to support the notion that new neural connections are made as a result of learning. But in view of more recent evidence, it progressively looks like that a subtle change in the already existing network of connections, could be the only functional change which comes about as the result of learning.

The investigations in this area of learning research, has amassed a useful body of knowledge on the changes that some specific types of learning cause to the nervous-systems of mainly simple animals. However, no universally valid principles have been demonstrated. With the criteria used in such investigations, differing neuro-biological mechanisms have been observed in the cellular correlates of the phenomenon of habituation in cats and mollusks. This shows that to find a true globally applicable engram, either we have to increase our knowledge of the workings of the brain and nervous-systems, or approach the issue from a different perspective and use different criteria for observation.

The evidence of neural changes associated with learning are as follows:-

(i) - NEURO-ANATOMICAL CHANGES OF LEARNING

The prediction had been made in theories such as Hebb's 'reverberating circuits', that an enduring oscillatory loop between two neurons will bring about persisting structural changes (of unknown nature); one possible hypothesis was the anatomical growth of synapsis between the two neurons, which results in both neurons acting as a single unit.

The only positive finding regarding the induced neuro-anatomical change as a result of external behaviour, is the relative large numbers of the neuronal synapsis and the larger cortex size of the group of experimental animals which are reared in complex environments rich in sensory stimulus, when compared with a similar group of animals that have been reared in a dull environment. These anatomical variations primarily observed during the growth of animals but in certain cases also in the experiments involving the adult animals, are possibly attributable to some neuro-chemical secretions. Hence, the general consensus in this field is that experience as well as organising the existing neural synapsis in the brain, in a 'select and preserve' fashion, 'directs' or 'causes' the formation of new synapsis.

(ii) - NEURO-PHYSIOLOGICAL CHANGES OF LEARNING

The proposal was made by Pavlov that possibly some electric fields could be responsible for the associative bonding of two neurons, this was investigated later but no conclusive evidence for such mechanism were discovered.

A method preferred by some researchers has been to use simple neuronal preparations as a model for investigating various hypothesis regarding the physiological changes in learning; these preparations range from a single neuron to groups of identical neurons. By observing the levels of electrical neuronal firings, it has been established that many simple types of learning such as habituation, sensitization, classical conditioning, and instrumental conditioning can be demonstrated at the neuronal level in simple animals like sea-slugs or locusts.

When studying the physiological manifestations of learning in such simple animals, two major problems are evident: firstly, many problems arise when attempts are made to generalise the findings to the more complex animals or man; and secondly, when a neural preparation exhibits the properties of habituation or conditioning, it does not necessarily follow that they correspond to habituation or conditioning at the behavioural level.

At higher levels of investigation, the experiments are carried out on the intact nervous-systems of animals; the 'mirror-epilepsy' experiments have managed to achieve the conditioning of neurons on the one hemisphere of mammalian cerebral-cortex to the firing of neurons on the opposite hemisphere. Other neural conditioning experimenters have accomplished similar results with the conditioning of reflex actions such as the eyelid movement of cats. Also, in a class of experiments, the neuronal correlates of particular learning tasks in the active animals have been localised on the basis of the increased electrical activity of such neurons.

At the more holistic level, some researchers have demonstrated a shift in certain brain-wave patterns or other electrical (EEG) changes during learning, but generally speaking no clear explanation of the nature of the 'brain product' of learning has emerged.

(iii) - NEURO-CHEMICAL CHANGES OF LEARNING

In the past few decades, the possibility of the existence of a neuro-chemical substrate of learning has been rigorously investigated, often with controversial and sensational outbursts of claims regarding the chemical means of 'memory transfer' or dramatic enhancement of learning abilities by use of certain chemicals.

The role of 'Ribonucleic-Acid' (RNA) in the transfer or acquisition of learning is a subject under close scrutiny of many neuro-chemists. It is established that some modifications of the structure of the RNA takes place during learning. Also, the maze-running proficiency of experimental rats have been correlated with certain characteristics of the RNA molecules in their brain. Similarly, the administration of some drugs which inhibit the RNA synthesis slows the learning in rats; however, this inhibition of protein synthesis does not affect the contents of long-term memory.

By far the most dramatic claims have been regarding the transfer of memory from one animal to another by means of 'encode RNA'. In one set of such experiments, planaria (earth-worms) were fed on the remains of other 'trained' planaria; and the claim has been made that the behaviour of the recipient animals were modified as though they had gone through the original training procedure. In the absence of a plausible RNA memory hypothesis which can explain exactly how RNA and memory are related, these types of observations should be viewed cautiously. Many other chemicals and drugs are also shown to be able to interfere with or block the memory-trace in a definite and well documented manner.

2.8.6 A BRIEF LOOK AT THE MECHANISMS OF THE BRAIN AND NERVOUS-SYSTEM

The Brain is one of the most complex and challenging mysteries of nature, the great volume of knowledge amassed about this object of intense scientific enquiry has only accomplished a minimal understanding of its mechanisms and functions.

In human nervous systems millions of 'receptors' monitor internal and external changes and transmit the information to the brain. The 'moto-neurons' in various organs receive the commands sent from the brain, for the execution of muscular movements or internal regulations. In the brain itself billions of neurons, with the number of possible interconnections greater than the number of atoms in the universe (according to many texts), process the incoming information and act appropriately. In FIG.2.2 a general view of human brain with its major anatomical components, also a typical outline of 'neuron' and its action potential are shown.

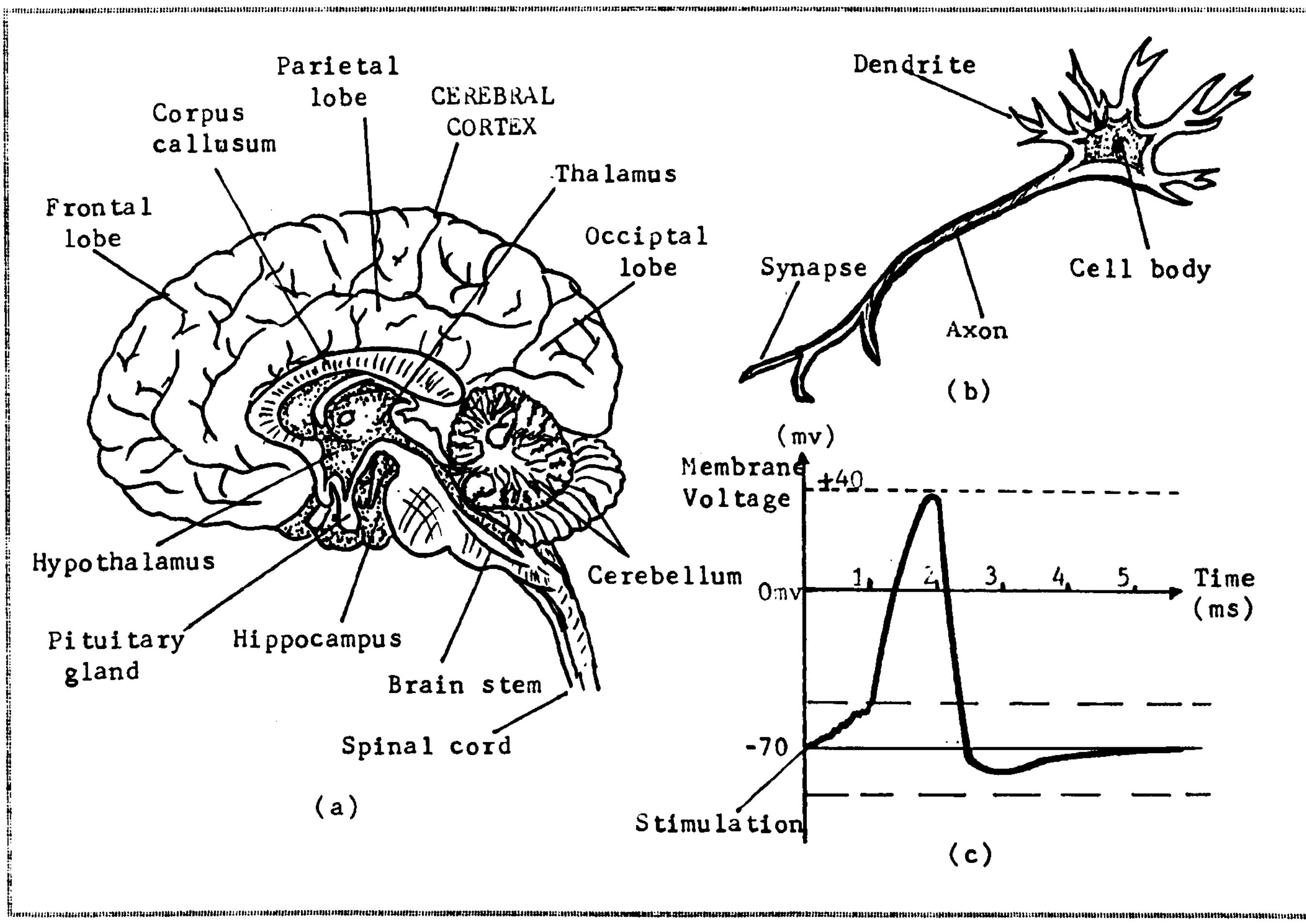


FIGURE 2.2. (a) - The brain with its principal regions
(b) - The outline of a typical neuron and its major components
(c) - The pattern of a typical neuronal electrical firing

(i) - THE NEURO-ANATOMY OF THE BRAIN

The brain is a mass of folded grey spongy tissue, made up of billions of living cells called 'neurons' (approximately 10^{10} cells in human brain). The brain anatomically consists of two different tissues. The outer cortex (grey-matter) and the inner (white) sub-cortical nuclei. The neurons in the cortex are relatively independent of their near neighbours, but through trans-cortical fibers, communicate with other distant neurons, and form a neural-network. The largest component of the brain of mammals and higher animals is the 'cerebral cortex', it is made up of two mirror copy hemispheres or lobes. The cortex is the region of the brain mainly associated with the non-automatic and adaptive behaviours of animals. The concept of the localization of functions or 'specificity', signifies that certain cortical areas are devoted exclusively or primarily to particular functions, such as: vision; audition; tactile sensing; initiation of muscular actions; establishing of associations; and speech. More locally, however, there appears to be an overlap of functions and the duplication of many activities in different regions of the cortex, this is referred to as the 'plasticity' of the brain.

The sub-cortical areas of the brain are mainly responsible for the mental life of animals. Their motivation, arousal, drive and attention, as well as some automatic reflexive actions of their behaviour are controlled from this region of the brain.

The units of the nervous-system, the neurons, come in a variety of shapes and sizes, in different species and also in different parts of the CNS. The neurons are made up of three major components: the'dendrites' (for receiving the information); the 'cell-body'; and the 'axons' (for conducting and transmitting of information from the cell-body). Each neuron is connected to many other neurons via its axons and dendrites (e.g., 10^4 connections in a typical human brain cell).

(ii) - THE NEURO-PHYSIOLOGY OF THE BRAIN

The nervous-system, as a vast network of receptors, neurons and effectors, can be considered to be an electrical maze of elements in continuous modes of activity. The means of communication between this network of neurons, is the electrical impulses that travel along the axons from one neuron to another. The neurons, typically, fire in interacting groups or modules, these modules, in turn, are the elements of larger integrated subsystems in the brain. There are three important principles of neuronal interactions: 'convergence' (many neurons influence the firing rate of a single neuron); 'divergence' (each neuron causes the firing of many other neurons); and 'feedback' (a neuron influences the firing rate of the neurons that affect them).

The nerve impulse is usually described in terms of travelling electrical charges or 'action-potential', this has an 'all-or-none' (on/off) characteristic. However, the sensitivity of the nerve-cell is not constant, and after an impulse the nerve is less sensitive to firing. The threshold of firing could also rise, as a result of 'adaptations' to a continuous stimulus, or 'accommodations' for a stronger stimulus.

The amplitudes of individual impulses do not vary and are of no value as far as the information transfer is concerned. The frequency of the firing of neurons, although significant at sense organ and effector level, is not of prime importance at the processing level, where the identity and the distribution of excited fibers, convey the real information.

Here, it is relevant to mention that the holistic approach to the electrical investigations of the brain has produced a mass of important results, involving the 'electroencephalogram' (EEG) recordings of the brain. Various types of EEG wave patterns are correlated with human mental states, and although, initially, strong claims were made regarding the significance of such recordings by their pioneering protagonists, their true composition is more uncertain today.

(iii) - THE NEURO-CHEMISTRY OF THE BRAIN

The neuronal discharge and the progression of the impulse in dendrites are generally considered in terms of the electrical properties of such phenomena, in much the same way as an electrical charge travels along a cable. However, the transmission of neural impulses can be considered in terms of the complex electro-chemical interactions that take place in the dendrites. The uneven distribution of charged 'ions' in the membrane of axons and dendrites create a potential gradient which facilitates the passing of the electrical signals through the cell. Also, other chemicals called 'neuro-transmitters' are responsible for the electrical impulse travelling between the synaptic junctions of two neurons.

As mentioned earlier, the investigations of 'ribonucleic acid' (RNA), which is a molecule in the nerve-cell body, and is responsible for protein synthesis within the neuron, has been another important area of neuro-chemistry. The production of RNA is shown, although not conclusively, to be related to certain aspects of neural organization.

Many hormones and chemicals have been shown to have blocking or stimulating effect on neuronal activity, some interfere with the transmission of impulses by loosening the synaptic connections between neurons. It has been recently established, that the brain also produces its own sensitizing morphine like chemicals ('endorphines') to suppress pain.

2.8.7 THE NEUROLOGICAL APPROACH TO SOME PSYCHOLOGICAL CONCEPTS OF LEARNING

The neuronal study of the brain has resulted in the discovery of mechanisms for the neural correlates of some psychological concepts. Some relevant issues to learning and performance namely, attention, arousal, motivation, drive, reward and punishment, will be discussed.

(i) - AROUSAL AND ATTENTION

The activities of a collection of diffusively arranged neurons, based at the core of the brain-stem, namely, the 'reticular formation', have been directly linked, with the phenomena of arousal, attention, alertness, awareness, sleep and coma. The process of attention or arousal, is the brain's way of limiting the intake of information. The learning animal must extract from its experiences the stimuli which are important for its survival and need to be remembered for future utilization. The role of the reticular formation in the attention mechanism has been clearly demonstrated ever since the work of Bremer in the 1930's.

The different levels of arousal, from alertness to deep sleep have been correlated with the patterns of EEG brain waves of animals. Also, the electrical stimulation of the reticular formation is seen to increase the attention level of an animal.

Animals with ablations or lesions to their reticular formation would fall into a deep sleep or coma. However, in more controlled and careful experiments, these animals if kept alive for a duration, would get some of their normal behaviour restored and come out of the coma. Again, the plasticity of the brain seems to have been demonstrated.

Studies have also shown, that a group of experimental primates that had learned to discriminate between two lights, and had their attention aroused by the electrical stimulation of the reticular formation, prior to the training, performed better than a similar controlled group. Hence, the increased attention had improved the learning. The nature of actual mechanisms involved in attention is not clear yet, but it is thought that the sensory input fibers to the cortex branch off at the brain-stem, and send their information to the reticular formation within which the level of alertness is evaluated and relayed to the cortex.

The reticular formation also plays a significant part in the sleep/wakefulness cycle. During sleep, only the important signals will be received by the brain. In the awake mode, the reticular formation is thought to exert some measure of control and modulation over the incoming sensory inputs. This modulation is usually inhibitory, attenuating a certain input while attending to some other stimulus.

In 'habituation' phenomenon, the sensory inputs are blocked or attenuated, so that the stimulus is no longer considered novel or significant. The concept of attention is of prime importance to learning, specially, in attaching 'meaning' to stimuli, and determining which inputs should be registered and which discarded.

Chemicals such as anesthetic drugs that produce unconsciousness, appear to act by depressing the reticular formation. Other drugs which increase alertness (e.g., amphetamines) and probably improve learning ability temporarily, also seem to exert their effect in this region of the brain.

(ii) - MOTIVATION, DRIVE, REWARD AND PUNISHMENT

Learning and motivation are highly dependant phenomena, as pointed out in the behavioural discussions of the subject. The role of drive in acquiring learning and performance is conclusively demonstrated.

The physiological investigations of the concept of drive and motivation, have revolved round the common biological needs, such as, food, water, etc. The mechanism of these drives have been rigorously examined. Various inputs of the sense organs, levels of certain chemical or hormonal secretions in the animal's body, collectively, determine the specific drive level of the animal and are also instrumental in regulating the appetitive actions.

The neuronal centre for the motivational aspects of the behaviour, has been found to be the 'hypothalamus' of the brain. Different areas of which are observed to be responsible for each type of drive or reward. The regions for the stop/start commands of an appetitive activity, are seen to be allocated for the reward/punishment of the same activity.

The ablation or lesion experiments have demonstrated that the removal or destruction of certain parts of the hypothalamus could make an animal overeat by large amounts (by interfering with the 'stop' mechanism), at the same time the drive of the animal seem to have been diminished. Similarly, lesions to other regions inhibits the 'start' mechanism, and if not force fed, the animal will die of starvation. The lesions if not extensive, will allow the animal to recover normal eating habits, after a period of time. This also, is in view of the plasticity of the brain cells involved.

The significance of the hypothalamus as a centre for a range of psychological variables, has been further emphasized by a line of research started in the past twenty years, initially by Olds and Milner, this work involved the mapping of reward and punishment centres in the brain. By using electrical stimulations of the hypothalamus, the 'pain' and 'pleasure' centres of the brain have been localized. An experimental animal would self-stimulate a pleasure-centre, to the point of exhaustion. Although, this type of self-stimulation at some brain sites, seem to depend on the animal's drive state. Some controversial experiments have reported similar studies on human subjects. The chemical investigations of drive mechanisms, has shown that, there are certain chemicals involved in facilitating or blocking of synaptic connections in drive areas of the hypothalamus. The evidence of the effect of many drugs commonly known as 'mood-altering', and also the discovery of endorphines (the internally made morphine), show a conclusive underlying chemical substrate to the machinery of reward and punishment.

2.8.8 THE NEUROLOGICAL BASIS OF MEMORY

In earlier sections of this chapter, the notion of memory, its different categorizations and various other relevant issues were discussed in detail. Here, we will only briefly elaborate the neurological correlates of memory.

The contemporary theories and speculations of brain mechanisms for memory-storage systems are somewhat confusing, and do not provide an adequate basis for solid explanations of this phenomena.

Memories are the stored records of an individual's experiences. In the physiological investigations of memory, it was established quite early that the mechanisms involved are not as originally thought a collection of neurons oscillating in a dynamic 'reverberating circuits'. Having rejected the purely functional neuro-physiological explanation, the 'consolidation' theories of memory-trace, based on some enduring physical, structural or biochemical changes in neurons, were introduced.

The experimental work on 'amnesia', induced in patients after accidents or brought about on animals by 'electroconvulsive-shock' (ECS), has demonstrated that, there appears to be two different memory modes of long-term-memory (LTM) and short-term-memory (STM).

2.9 THE EVOLUTIONARY AND DEVELOPMENTAL ASPECTS OF BRAIN AND LEARNING

In this section we will review the issues raised by the phylogenetic (the evolutionary relationships in species) and the ontogenetic (the developmental stages of individual's growth) studies of nervous-systems and learning.

These two outlooks are intricately dependant on each other, and, as is often said, the various stages of animal evolutionary developments are analogous to the developmental stages of an embryo into an adult mammal, which involves the transformation of a single-cell into a fish like organism then into an amphibian form and ending with a fury mammal.

The fact that animals and plants adapt to their environment was recognized long before the theories of evolution. The 'natural' classification scheme and the similarities of different species were the main supporting criteria. However, ever since the introduction of the systematic explanations of the process of evolution by Darwin, Lamarck and others, the scientific search of the mechanisms of this complex process has been the objective of researchers from many disciplines. A constant stream of evolutionary hypothesis, corroborated by empirical results or fossil records, have been contributing to the true understanding of the origins of life and species. A prime example was the finding of the mechanisms of heredity, and variations in animals, in the science of Genetics.

The first issue to be emphasized is that the biological evolution should be only looked at in the context of a general cosmic evolution, which encompasses: nuclear, geological, chemical, organic, behavioural, social and cultural evolutions, with a hierarchy of ordering. Life has evolved filling a particular 'niche' in the composite picture of cosmic evolution.

Biological evolution is defined as the process of adaptations seen within the living organisms. It has been the force which has helped to enhance the complexity of biological life. The process of 'natural selection', as introduced by Darwin, is generally thought to be the primary law of evolution. However, looking from alternate view points, the propagation/reproduction/nourishment of the genes have all been considered as the governing agents of the process of evolution.

The origin of life dates back perhaps 3.5 billion years, some 1.2 billion years after the solidification of earth. The simplest forms of 'living' organism, created by physio-chemical processes, were the organic molecular structures capable of simple chemical reactions. The feasibility of the synthesis of inorganic chemicals into complex organic molecules, such as that found in the living cells, has been experimentally demonstrated by numerous workers.

The earliest living organisms are envisaged to have been energy-producing rather than energy-utilizing entities, possible conceptual models for such mechanisms have also been put forward. The forces of natural selection acting on these simple life forms, which possessed the three vital properties of 'multiplication', 'variation' and 'heredity', created the complex and varied range of life seen today.

The process of biological evolution operating on all organisms, yet at varying degrees and speeds, has produced, on the one hand, the plants which are more or less homogeneous in their complexity; and, on the other hand, the animal species from simple bacteria to man. Although, it must be noted that the forces of evolution are active on all levels of life, and today's single-cell organisms are of much greater complexity than that of say two billion years ago.

2.9.1 THE EVOLUTION OF REACTION MECHANISMS (NERVOUS-SYSTEMS)

The 'nervous-system' is one of the most complex and intriguing products of the process of biological evolution. Life has evolved from the simple living molecules into the thinking and symbolizing animals, that are capable of questioning, observing and explaining in scientific terms their own nature and origins. In the science of biology, it is unanimously accepted that animals share a common ancestry, this is not only in view of their unique genetic coding mechanisms, but also because of the similarities in their neural transmission methods and other biophysical resemblances.

One of the most essential and fundamental properties of living matter is 'irritability', which is the capacity to respond to stimuli. The most primitive organisms are thought to have exhibited this irritability by performing simple chemical reactions, such as the process of photosynthesis. To call these organisms 'alive' or not is a matter of definition. Viruses which in size and complexity of activity approach the very largest of non living molecules (proteins), are normally considered an intermediary between living and non living matter.

(i) - UNICELLULAR SYSTEMS

Evolutionary adaptations transformed some of the primitive simple living entities into more complex single celled organisms. The present day unicellular animals exhibit a variety of 'behavioural properties' which could give an indication of the kind of forces active in the primitive life. Initially isolated masses of living material probably showed a general irritability, through the course of evolution gradually the property of irritability came to be localized and refined to special pathways, both in sub-cellular and acellular levels.

There are three rudimentary properties which are seen even in the simplest of the single cell organisms of today: Sensitivity, Reaction and Conductivity. In simple unicellular animals such as 'amoeba' and 'protozoa' (e.g., Paramecium), simple stimuli (e.g., food, heat, cold, touch, chemicals or light intensity) are sensed by specialized parts of the animals' body (e.g., photo-chemically sensitive molecules or sensory bristles). Reactions of these organisms consists of feeding or movements, which takes place by the reshaping of the body (amoeba) or the movements of whip-like 'cilia' (paramecium).

Even with the limited scope of these sense and reaction mechanisms these simple animals show a wide variety of behavioural patterns, mainly based on the attraction to or the aversion from different stimuli. The amoeba's entire behavioural repertoire seems to revolve on searching for food, capture and ingestion of food depends on the specific type of the food and the satiety levels of the amoeba, it invariably withdraws from all non-food stimuli. The 'conductivity' is also present in amoeba in the biochemical channels of communication within its cell body, stimulation at one point is followed by the appearance of finger-like projections at another point.

The paramecium also is in a spontaneous state of activity, characterized by the rhythmic and coordinated movements of its cilia, however, it switches to specialized patterns of behaviour (e.g., feeding, object-avoiding, etc.) when encountering a significant stimulus. The conducting system in these protozoa being the forerunner to the early nervous-systems, is made up of a network of thread like fine fibers connecting all the cilia together. In the absence of a neural mechanism for behaviour control, it is assumed that the genetic and other information processing is carried out biochemically within the animal's cell body.

The primitive life, on the one hand, evolved into more complex unicellular organisms; and on the other hand, evolved into associatively joined cells, then into tissues, tissues into organs, and finally organs into organisms, where the adaptations of each unit is not solely in terms of the individual unit's requirements, but in terms of the survival of the whole.

(ii) - MULTICELLULAR SYSTEMS

The next evolutionary step in the development of CNS was the transformation of the single cell organisms into the multi-cell animals. This process first involved the change into colonies of single cell units and later into true multicellular entities. The multicellular animals with their infinite possibility of combinations of cell units, attained complexities much higher than the unicellular organisms. The resulting much larger animals were made up of self contained cells, each with their individual metabolic and dynamic attributes, yet affecting each other in varying degrees. With the multicellular animals came the need for greater coordination and communication between individual cells. The cells in the body were grouped into specialized units and there was a division of labour in dealing with different changes (both inside and outside the body).

Specific sense organs, muscle systems, digestive systems, etc., were developed, which further enhanced the irritability of the animal.

The unicellular means of signaling, namely, chemical messenger systems (hormones), were too slow and inaccurate (were diffused randomly throughout the body), hence, were gradually replaced by a more rapid electrical conduction system. This system could sense, integrate and react to environmental information much more efficiently. The arrangements of irritability and conduction cells, that control the functions of the whole organism are called the 'nervous system'.

(iii) - SIMPLE NERVOUS SYSTEMS

Whether the nervous systems were developed after the emergence of its basic element the 'neuron', or were developed from equivalent biochemical systems is not clearly known. However, it is commonly accepted that, the first neural structures were direct connections between the 'receptors' and 'effectors', and that the evolution of the nervous systems has resulted in the development of less direct and more complicated linkages between the sense and response mechanisms.

A large group of animals whose ancestors possessed the primitive nervous systems are hydras, jelly fish, corals and the sea anemones. In these organisms, neurons and synapses are already fully developed.

Although, the physiology of the neuron is almost the same within all species, there is a large variation of the type of synapses both within the nerve-cells of a particular animal and between different species. In the simplest of the above group of animals, the hydra, reflexive behaviours, such as feeding or movements away from dangerous stimuli have been developed. To achieve the relatively high level of coordination required for these reflexive behaviours, basically three types of neural cells have been evolved: 'receptors', 'effectors' and 'conductors'. These cells are organized mainly near the surface of the animal's body close to the external stimuli. Three types of reaction mechanisms can be seen: (1) - effectors that can be directly and independently stimulated by environmental inputs; (2) - effectors that are stimulated directly by receptor cells; and (3) - receptor cells that are connected to the effector cells via the conducting cells, hence facilitating a more varied and complex range of reflexive behaviours.

In the colonial species of these groups of animals, a colonial nervous system as well as the individual unit's nervous system can be seen.

More advanced sense organs (e.g., light sensitive) were also developed in some species similar to jelly fish. The 'nerve net' conduction system of these simple animals, although appear to be a synaptic system, has two distinctive features, firstly, unlike synaptic systems they allow for bi-directional transmissions, secondly, the whole nerve net system is interconnected, and every activity seems to involve all of the nerve cells. The centralization of these nervous systems were severely restricted by the radial asymmetry of the animals' body. The ultimate in radial nervous systems is achieved in

'Echinoderms' (e.g., starfish), with their much higher versatility of movements and central structuring (though fairly superficially) of their nerve nets.

Almost from the beginning of the evolution of the nervous systems, two major trends have been evident, one towards the specialization of units, the other towards the integration of similar units. The overall effect has been towards centralization.

As the nerve nets condensed into nerve tracts and moved away from their location near the surface of the body of animals, the distances between receptors and effectors became greater, and the direct synapses between input and output mechanisms all but disappeared. The number of intervening neurons between effectors and receptors increased, and were progressed deeper into the animal's body. These developments and specializations allowed animals to make more refined and coordinated responses that could be utilised more or less independently (e.g., swimming, feeding).

(iv) - COMPLEX NERVOUS SYSTEMS (Invertebrates)

The next major evolutionary step was the formation of 'ganglionic system'. Worms and mollusks are examples of the types of organisms which developed the 'ganglia'. These species have a greater repertoire of behaviour, have more complex movements and unlike the simpler animals have a specific set of neuro-muscular connections. Another important characteristic of their nervous system is that, while in the nerve net cells were dispersed evenly throughout the organism's body, in worms and mollusks the inter-neurons are clustered together and cell bodies are formed into localized masses called 'ganglia'. Nerves arrive at the ganglion from sense organs or other ganglia and leave it for muscles or other ganglia. Ganglia possess three different types of nerve cells: receptor-neurons, motor-neurons and inter-neurons. Thus the ganglia having all the essential features of the central nervous systems, is the evolutionary link between nerves and brains.

The transmission of neural impulses in ganglionic systems are unidirectional and a much more precise and rapid control of muscular activity is possible. The ultimate in neural conduction speeds have been achieved by the development of giant nerve fibers in mollusks. These neural developments together with the bilateral configuration of the animals' bodies, evolved organisms, that had their centre of neural activity (the ganglia), located in their heads. Unlike more primitive animals, worm like creatures had a distinct 'head' and 'tail', and their movements were biased in the forward direction. The evolutionary selection forces developed sense organs of higher complexity, situated mainly in the anterior areas of the animal body.

Originally, the 'head ganglia' was a kind of relay centre for the sensory and motor neurons that stretched their axons into it. Later the messages arriving from the senses were modified before being passed on to the motor systems. Many inter-neurons were evolved, possibly independently, within the confines of the ganglion itself. The complexity and variety in structure was reflected in the functional richness; many areas in the ganglion were localized and specialized for specific tasks, and an hierarchical mechanism for the control of the activities of animal evolved.

The ultimate evolutionary products of ganglionic systems can be seen in today's arthropods, insects and mollusks. The nervous systems of such animals although controlled centrally, to some extent have a degree of autonomy within each individual functional group. A severed tentacle of an octopus will still function normally for a period of time. This diffused nature of ganglionic activity, together with the design and structural limitations imposed on the growth of the brains of such animals, has resulted in the development of highly successful and varied yet stimulus-bound species.

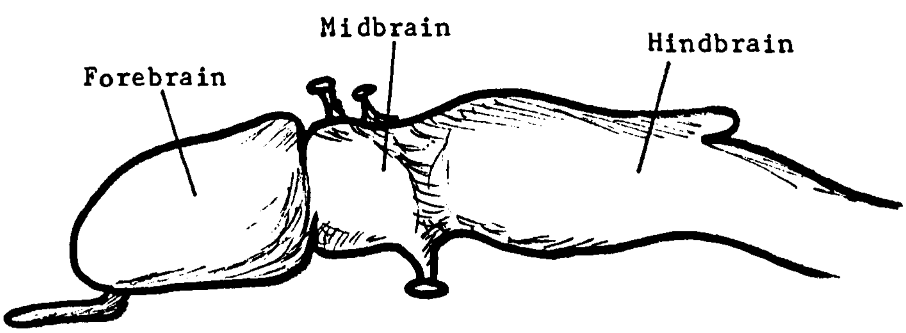
Insects and ganglionic animals with their complex vision, olfactory, orientation and chemical sensory capabilities, exhibit an intriguing array of instinctive behaviours and mimic many aspects of vertebrate life. An octopus on the other hand, with the largest invertebrate brain, if allowances are made for the peripheral ganglia, has only a rat size brain. Ganglionic brains seemingly do not have the adequate features which enable the evolution of powerful learning mechanisms.

(v) - CENTRAL NERVOUS SYSTEMS (Vertebrates)

A separate evolutionary development has led to the appearance of vertebrate nervous systems and brains. For this to have happened two major elaborations of the invertebrate nervous system was necessary, first the separation of the radial neural tract from the gut, and second the centralization of the nervous system in the brain.

The development of bony skeletons in fish-like primitive animals was the first ascending stage; the spinal cords developed within the solid tubular bones of these animals, but relatively speaking spinal cords have not evolved to a great extent during the course of time, however, they have lost their exclusive dominance of neural activities. The swelling at the head end of the nerve tubes of the primitive vertebrates were the forerunners for today's mammalian brains.

The vertebrate brain as well as the spinal cord has three basic areas: 'sensory', 'motor' and 'association'. The development of more and more complex receptors was one of the major forces involved in the growth of vertebrate brains. In terms of functional complexity, the more recently evolved parts of the nervous system, namely the anterior levels, are responsible for the more complex behaviours, and the more primitive posterior regions (e.g., spinal cords), are responsible for reflexive and stereotyped functions.



The primitive vertebrate brain (from: Rose, 1976)

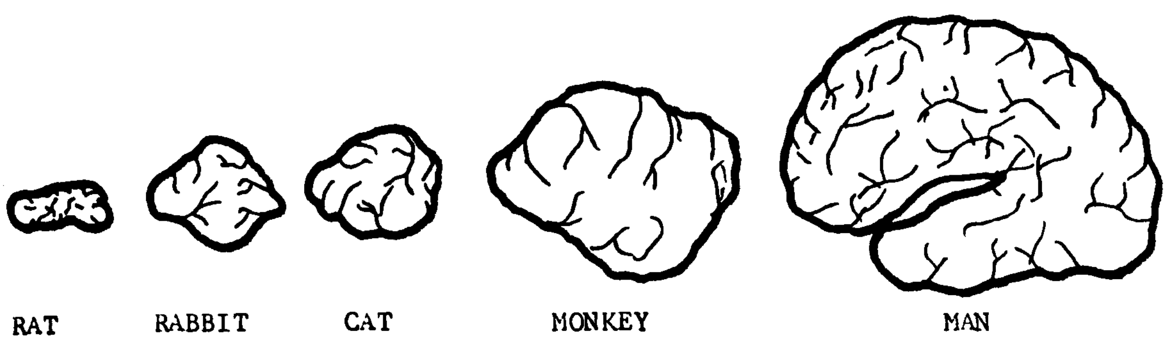


FIGURE 2.3. A diagram of primitive brain (3, regions) + a diagram showing relative sizes of brains of few animals (rats, rabbits, bird, man, etc.).

Generally, it is thought that ever since the formation of first vertebrate brains there have been three distinct areas present at the head end of the neural tube. These areas are the 'forebrain', the 'midbrain' and the 'hindbrain', respectively associated with the functions of smell, vision and equilibrium/vibration/balance (refer to FIG.2.3).

The development of these three regions of the brains and the relative dominance of different regions and hence their relative sizes in different species, is the prominent feature of the brain's evolutionary history. The primitive marine animals had the midbrain as the most utilized section of the brain. The 'cerebellum' started to feature as an important area in the amphibians, needing a greater degree of coordination of movement and body equilibrium. With land based animals the forebrain started to develope into 'cerebrum', which was used for the processing of the more complex olfactory environmental inputs. The 'thalamus' developed in the amphibians to deal with the coordination of different senses lost its dominance of this function in the land based animals, to a more advanced cerebrum. In birds, cerebellum is well developed, perhaps to deal with the coordination problems encountered in flight.

A variety of cortical developments can be seen in mammals. The smooth cortical surfaces are the feature of the cortex of the lower vertebrates and the folding twisted cortical surfaces are seen in the higher mammals, this presumably is to accommodate more volume of nerve cells within the brain. The latest addition to the mammalian brain is the 'neocortex', which is the only area in the mammalian brain that does not have a primitive equivalent in the reptile brain. Neocortex of the primitive mammal was used for motor functions, but in the higher animals this region of the brain is used as the 'association area', which is an essential region of the brain in learning animals.

Attempts to equate the intellectual powers of animals and the size of their association areas have failed, since the development of the brains depends both on the quantity and the complexity of the sensory inputs. Different areas in the brains of various species are dominant, and the same areas of the brain are sometimes used for different functions in two groups of animals.

In evolutionary terms there seems to be a steady increase in the size and complexity of cerebral cortex, humans are amongst the highest (cortical-neuron-number)/(body-weight) ratio animals (refer to Fig.2.3). A striking feature of human brain is that it has biologically evolved very little from the brains of Homo-Sapiens. Also, despite the similarity to the brains of some primates in shape and anatomy, human brain is of far greater volume - an indication of the much greater reasoning and language capabilities of humans.

2.9.2 THE EVOLUTION OF ADAPTIVE BEHAVIOUR

Parallel to the evolution of reaction mechanisms and the emergence of nervous-systems in animals, their behaviour has also been evolving and modifying into activities of much greater complexity, hence prolonging and increasing their chances of survival. The adaptations of behaviour started from the 'irritability' of crystal like living molecules and culminated into the diverse and intricate patterns of individual and collective behaviours of mammals.

According to the hypothesis of natural selection and the empirical observations concerning the inheritance of behavioural traits in successive generations, it is assumed that as well as purely physical variations between members of a species, behavioural traits and mutations also exist. These variations are passed on to the off-springs in the form of neural organizational patterns, and the preservation of such patterns governed by the laws of natural-selection, controls the development of behavioural evolution.

In the case of man, the language capability has initiated the cultural evolution which has accelerated the natural course of behavioural evolution by many times, the information and knowledge is not only passed on to the next generation in the genetic form but also by other symbolic and semantic means.

An alternate explanation of the evolutionary mechanism, is the intuitive notion that, animals pass on to their offsprings not only their genetically coded information which is present at birth, but also part of their acquired knowledge during their lifetime (possibly by some sort of genetic modification). This hypothesis commonly known as 'Lamarckian' explanation of the evolutionary process, has occupied the minds of many researchers since the 19th century.

Yet, in spite of numerous experiments all attempts to substantiate such a theory has been futile. Specially, since the rise of Darwinian evolutionary ideas there has not been a sympathetic scientific environment to promote Lamarckian ideas much further.

The interesting point is that the mechanisms for inheriting learned behavioural traits (hence accelerating the evolution of behaviour) evidently exists in the form of genetic-coding-machinery, which can accommodate very large amounts of coded information within each cell body of the organism.

Recent developments have shown some possible ways that the RNA and DNA synthesis could modify the genes and transform the acquired information of an animal to its next generation.

However, presently in the absence of any plausible solid explanation, the statistical effectiveness of large numbers of individuals and generations as promoted by 'natural selection' is the commonly accepted hypothesis, but it must be remembered that Darwinian explanation is not the only or the complete explanation for the process of evolution.

In modern psychology, the evolutionary aspects are seldom looked at in the topic of 'learning'. In general, the main emphasis in learning research (in geographical sense), have been the American Behaviourism, the Western Cognitivism and the Eastern-Block Pavlovianism.

The evolution of learning behaviour, from the simple habituation to complex conceptual problem-solving, has been a slow progression accompanying the development of nervous-system from the single-cell organism to man.

The process of evolution itself can be thought of as an elementary kind of slow learning process. A population of birds, under the influence of the forces of natural selection, genetically adapt to certain new environmental conditions (e.g., change in dietary supply), after a few generations of encounter with such conditions; while the individual birds show some adaptations (learn) to the same environmental conditions during the span of their lifetime.

A principal question is: when exactly in the course of evolution did associative memory and learning enter life and behaviour. Various evolutionary developments indicate that in the same way that the birds acquired wings before learning to fly, the mechanisms for learning must have been available before any specific type of learning was exhibited. The extent of learning ability has therefore been governed by the elaboration of the sense organs, the development and the concentration of the nervous-systems.

The process of mental evolution and the evolution of learning behaviour is characterised by many continuities and discontinuities. As species evolved into many diverse groups and phyla, variations in the rates of development of different branches of evolution started to appear. In some cases, an optimal plateau seems to have been reached, and with the type of nervous-system incorporated within such species, the optimal potentiality of a specific level of learning is also seemingly attained, such as the intricate innate behaviours of insects with relatively poor degree of adaptability.

The view that the evolution of learning encompasses definite leaps is not a universally accepted one, and although some novel types of learning such as linguistic-learning in man is acknowledged, a continuous transformation from instinctive reaction to conceptual thought is also generally envisaged. Every organism that is not solely dominated by innate modes of behaviour, is capable of learning; and according to Pavlov and his followers organisms with synaptic neural systems have the conditioning of reflexes as the basis for all adaptation and learning they exhibit. This argument has since been partially justified for the lower animals up to mammals, though again different learning processes seem to be involved in fish and birds.

In a comparative study of learning within species variations between individuals or other environmental factors are ignored, and simply the existence or the lack of a particular type of learning in a group of animals is considered under optimal conditions.

It must also be noted that certain physical commonalities superpose and encompass all types of behaviour in species, also indicating a continuity in the formation of various adaptive mechanisms and processes. The causality and the uniformity of physical events; the concept of time; spatial, visual, auditory, and other physio-chemical properties imbedded in perceptual information; are examples of factors which seem to have identical intrinsic values for all animals. Such issues although trivial in a sense, and normally taken for granted, should be considered and incorporated in any true model of the learning process.

The speculative method which was used to trace the development of nervous-systems throughout the evolution is even more apparent in the investigations of behavioural evolution, and in particular the evolution of learning. No direct fossil records of primitive animals' nerve-systems exist, hence predictions of their form and mechanisms are made using the present day animals of identical physiology. Similarly, in tracking the evolution of learning behaviour the same criterion is used. However, it must be said that this hypothesis is even more speculative in the case of behavioural evolution and can never be fully verified.

2.9.3 EVOLUTIONARY STAGES OF ADAPTIVE AND LEARNING BEHAVIOURS

The most trivial types of adaptive behaviour in evolutionary terms are the 'taxes' (orientation responses to stimuli), 'reflexes' and 'instincts' which govern the majority of the actions of lower animals. Yet, the most primitive true "learning" is probably in 'habituation', which manifests itself in the protozoa, man, and even in the single neuron; the mechanism of habituation has remained virtually unchanged throughout the evolution.

Habituation occurs when an organism no longer attends to repeated arousable stimuli. This action of nervous system (apparently based on the actions of neuro-transmitter chemicals in the active neurons) ensures that only the relevant information reaches the animal's brain. Habituation also implies that a sort of record should be produced in the form of a memory-image, since the habituated organism partially retains such response in future occurrences of the stimulus. The investigations of habituation in lower animals gives a clue as to where and how this memory-image is formed. Evolutionary kinship of habituation and associative learning is apparent from the many similarities of their properties.

Sensitization is the next developmental step in evolutionary terms, it is the opposite notion to habituation, where an animal is aroused or alarmed and the probability of its response to a stimulus increases, whereby it reacts more readily to the repeated stimuli. Sensitization seemingly was not present at the most primitive types of life and is only evident in the pre-vertebrates such as worms and early vertebrates.

Learning of associations is the next major evolutionary category, an association can be 'conditioned' in an animal based on the bounding of various stimuli and responses, mediated by certain inhibitory or reinforcing influences. There are generally thought to be three types of conditioning, which in their ascending evolutionary order are:-

(a) - Inhibitory (punishment) Conditioning
(b) - Classical (Pavlovian) Conditioning
(c) - Reinforcement (reward/operant) Conditioning.

Inhibitory conditioning or punishment learning can be observed throughout the primitive metazoic animals. This process has evolved as a more efficient replacement for habituation. In evolutionary terms punishment mechanisms were evolved prior to reinforcement (reward) mechanisms; hence, these two notions are not symmetrically opposite in the developmental sense.

The classical conditioning or simple associative learning is the core of conditioning and the most typical manifestation of simple learning behaviour. This type of conditioning can be readily seen in simple animals such as worms and mollusks, and at the highest level can be observed in the conditioning of the reflexive responses of man. Classical conditioning has two basic varieties of 'aversive' and 'appetitive', having different retention durations for the conditioned response.

The reinforcement (reward) conditioning is the next stage of the evolutionary ladder of development. The presence of reward systems in animals of more advanced nervous-systems is apparent from both physiological and psychological observations. The simple mechanistic relationship between reward and stimulus-response learning is the basis for governing the behaviour of simpler animals. The emergence of reinforcement conditioning in animals can be correlated to the appearance of spinal cords, brain stems, and hippocampal reward centres

Some higher forms of conditioning govern the more complex learning behaviours of higher (perhaps more intelligent) animals. Unlike the lower animal forms, conditioning in the more complex animals is not so much stimulus-bound and mainly originates from within the organism itself. Guided by the hypothesis derived from experience, the animal attempts to reproduce the rewarding and pleasant experiences by acting on the basis of 'expectations' rather than inputs, and in the process may use complex symbolic means such as the use of language in man. These higher conditioning forms have three basic evolutionary types:-

(i) - Sensory Preconditioning - (e.g., learning of associations between sensory data).

(ii) - Integrative Conditioning - (e.g., latent learning).

(iii) - Predictive Conditioning - (e.g., insight learning).

Even higher types of linguistic learning is prevalent in man, in the form of symbolic learning which can also be categorised into following:-

(a) - learning of simple words;

(b) - learning of simple predications;

(c) - learning of propositional connections.

Although as mentioned earlier the evolutionary hypothesis of the development of learning is largely speculative and is based on tentative criteria. However, several important principles have been postulated using the evidence accumulated in this field Razran (1971):-

(1) - Higher levels of learning arise from the lower levels, and bring new laws of learning manifestation.

(2) - Complex modes of behaviour are added to simpler ones and finally replace them.

(3) - Lower levels of learning continue as subsystems within the higher systems, contributing to their processes.

(4) - At each level of learning new behavioural patterns emerge.

(5) - Higher levels of learning are more efficient but the lower levels are more universal and less disreputable.

(6) - Normally higher levels of learning have control over lower levels, but under certain conditions the lower levels are predominant.

(7) - The behaviour of simpler animals can not be accounted for in terms of higher levels of learning.

(8) - Similarly, the behaviour of more complex animals can not be accounted for by reducing it to simpler components of lower level learning such as reflexes.

Some empirical observations justifying the hierarchy of the evolutionary levels of learning are: Coelenterates (hydra, sea-anemone) can be habituated but not conditioned; some spinal mammals habituate, become sensitized but are not conditionable; gill extension in some fish can be modified by punishment and classical conditioning, but not by reward conditioning; other fish and lower vertebrates can be fully conditioned in all three types; the classical conditioning is prominent in most invertebrates, but higher invertebrates can also master reward conditioning; the sensory pre-conditioning and integrative conditioning is possible in birds and mammals; predictive conditioning can be seen in some birds (crows) and higher mammals (dogs, cats); higher forms of learning are in general attributed to primates and the symbolic learning is an exclusive achievement of man.

2.9.4 LEARNING ABILITIES IN DIFFERENT SPECIES

The relative dominance of various types of adaptive behaviours within major species groups is outlined in FIG.2.5.

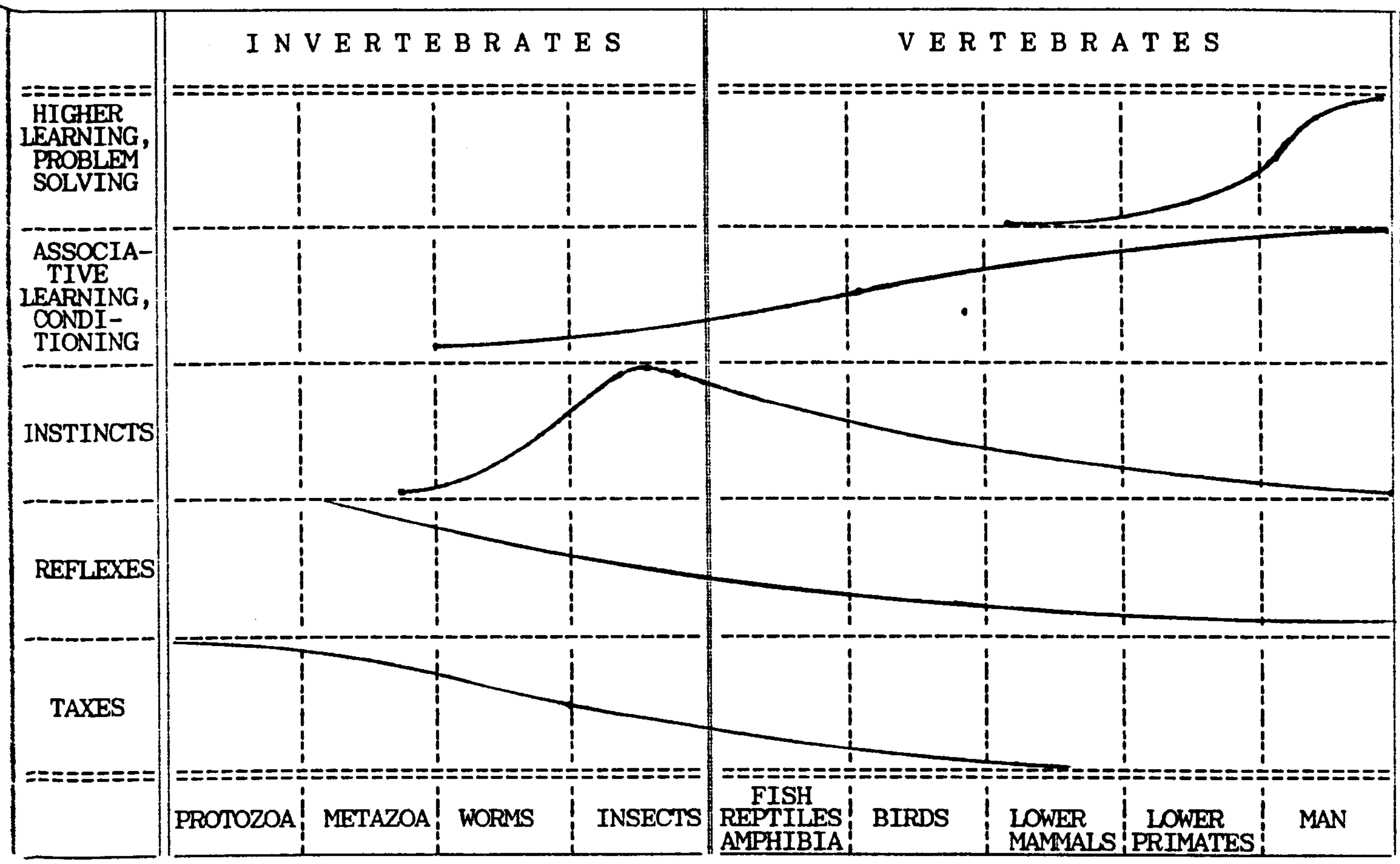


FIGURE 2.4. An illustration of relative contributions of various types of adaptive behaviours in the major groups of animal species, in the ascending order of their evolutionary development (from Dethier and Stellar, 1970).

In the following each major animal phyla will be briefly examined for their adaptive and learning potentialities.

(a) - LEARNING IN PROTOZOA: Protozoa of today represent animals of much greater complexity than the primitive single-celled organisms, they range from simple 'amoeba' to the organisms such as 'parameciums' with their elaborate moving cilia. Protozoa are observed to have innate patterns of behaviour (mainly taxes), such as the 'avoiding reaction' of a paramecium when it collides with an obstacle. Habituation has also been demonstrated, but less conclusively, since the habituated response diminishes very quickly; hence, the temporary sensory change, damage or sensitization cannot be completely ruled out. Associative learning has also been reported in these animals, but again the evidence has not been universally accepted or substantiated, basically on the same grounds as habituation.

(b) - **LEARNING IN METAZOA:** Animals in this group such as 'hydra', 'sea-anemone', and 'star-fish' have a range of simple to complex modes of stereotyped behaviours (e.g., feeding, locomotion, protection from noxious stimuli, etc.). Higher metazoa such as star-fish, characterised by more freedom and dexterity of movement, have developed relatively more complex range of behaviours, with faster and more coordinated actions. Efforts to establish habituation and associative learning in this group of animals have also met with many controversies, based on the same critical argument as in the case of protozoa. However, there is enough experimental evidence to suggest that habituation and simple associative learning can be displayed according to certain criteria, which have to be chosen carefully in view of the limited perceptual and sensory domain of such animals.

(c) - **LEARNING IN WORMS AND MOLLUSKS:** These animals are still very much stimulus-response systems. However, here habituation and associative learning can be clearly demonstrated, even in the simple examples of this group (e.g., flatworm). The increased variety of activity, allows experiments such as the learning of 'T' mazes to be conducted on these organisms. The development of 'ganglia' within this phyla, is the major factor which has enabled the more advanced mollusks such as octopuses to perform relatively complex learning and discrimination tasks, and solve 'detour' problems. The dynamics of ganglionic cells have also been investigated in isolation for the presence of any adaptive characteristics, and properties analogous to habituation, classical and operant conditioning have been demonstrated within a single cell.

(d) - **LEARNING IN ARTHROPODS AND INSECTS:** This phyla includes more species than all other species combined, and the diversity and elaboration of form and behaviour is immense. The development of acute sense organs, enabled the insects to perform highly specialised individual and social patterns of innate behaviour. Although, learning plays a small role in the life of arthropods which are basically stimulus-bound organisms, a variety of different types of associative learning have been observed amongst specific groups of arthropods; examples are: habituation, classical and operant conditioning, trial and error learning, maze learning, latent learning, imprinting, and even some primitive forms of tool using.

(e) - LEARNING IN FISH, REPTILES AND AMPHIBIA: These animals generally possess all adaptive capabilities of the previous groups, fish have been trained for various detour, maze solving, discrimination, and other problem solving tasks. The maze solving abilities of these early vertebrates are found to be more rapid and consistent than the invertebrates (with the possible exception of ants).

(d) - LEARNING IN BIRDS: Birds show yet more facility in learning of more complex problems, and as well as other types of lower learning, there are instances where the insightful behaviour has been demonstrated, concept of numbers, territorial learning, complex imprinting, tool usage, and other interesting facilities have been empirically observed.

(e) - LEARNING IN MAMMALS, PRIMATES AND MAN: Learning ability is enhanced to its optimal limit in this phyla of animals. The innate behaviours of mammals are not as elaborated as other species, because the process of evolution facilitated higher forms of learning, and freed these animals from the dominance of simple stimulus-response and trial and error type learning. Mammals are able to use their experiences more and more in the modifications of their behaviour; in higher mammals and primates the conceptual learning aids the solving of complex problems; and in man, the ultimate in learning, the symbolic type learning is utilised for the rapid and extensive analysis and processing of a diverse range of problems.

2.10 GROWTH AND DEVELOPMENTAL PERSPECTIVE OF THE BRAIN AND LEARNING

Finally, in this chapter we look at the ontological aspects involved in the formation of adaptive behaviours. The various stages of the development of a complete adult animals' nervous system are normally initiated from a single egg-cell. Through an accurate process of duplication and division, and using the internal genetic instructions, the cell transforms into a complete animal.

The blueprint available to the developing foetus, designates the spatial organization of various organs. The brain starts to develop by the formation of a 'neural-tube' that is capable of producing large amounts of simple nerve cells; these cells in tern, by using the chemical tags attached to them, travel to their allocated region or make the appropriate connections; hence, developing the wiring diagram of the brain. Extending axons search for their destined targets, and after they are found the secondary dendrites are formed in a similar manner.

The degree of the brain development at birth (both physiologically and biochemically), varies amongst different species. In humans, the actual production of neurons all but ceases at birth; the subsequent increase in size of the brain and the developments which occur in the nervous system only involve the restructuring and the growth of further neuronal connections.

Throughout the animal's life, its neurons become less modifiable functionally, and a small percentage of them are destroyed, but unlike other cells in their body are not replaced. The variations of the modifiability of neurons can be observed in the apparent ease with which children learn languages.

Many sociological and psychological evidence have pointed out the factors which can affect the development of the nervous systems; examples are: nutrition, experiences, and environmental parameters that can enhance or disrupt the quality of neural functioning. The behaviour of an animal or certain biochemical/physiological aspects of the brain may be modified during a critical time of its life, referred to as the 'sensitive period'.

Using the criterion that performance and learning are supported by appropriate neural mechanisms, then the behaviour of an animal should develop from the very simple reflexive and automatic functions. This type of behaviour in fact can be evoked prenatally, and has been observed in human foetus which display a reflexive avoidance mechanism.

At birth most of the activities present are non-cortex dependant, and are normally controlled by the spinal cord or the lower brain regions. Later motor capabilities are developed followed by sensory processing capabilities. Increasing interactions of the animal with its environment, causes the brain, the perception, and the behaviour to develop in parallel; this is seen by the increase in the brain mass and connectivity, and also by the modifications of behaviour into the more complex adaptive forms.

In the case of humans, there have been many scientific investigations into the changes which the learning, intellectual and perceptual capabilities go through, during the early years of their life; most prominent in this area being Piaget's (1977) extensive research into the origins of intelligence.

During the span of behavioural and neural development, two criteria are prevalent, 'specifity' and 'plasticity'.

(1) - SPECIFITY: This property can be observed in the way which certain neural mechanisms have their implicit coding instructions dedicated to specific tasks, and these mechanisms can not be modified to do a different function (e.g., severed or damaged nerve tracts that regenerate in a specific manner). Similarly, the specifity of behaviour can be seen in the fixed patterns of reflexive or instinctive behaviours.

(2) - PLASTICITY: This property has been experimentally demonstrated by the formation of certain neural connections in the brain which only become functional as the system matures; it is shown that during the sensitive period of animal's life, the development of these connections can be influenced by experiences (e.g., in the formation of visual cortex of cats). Behavioural plasticity can be seen in all animals which display adaptive activities (i.e., by using feedback from environmental sensory inputs, they can modify their perceptual abilities). The highest manifestation of the plasticity of the nervous systems is the learning behaviour; in the most primitive form, such as imprinting or the formation of parental attachments, a sensitive period during the early life is crucial; but for other types of learning, the capability persists throughout the life, with varying degrees.

CHAPTER 3

TOOLS & TECHNIQUES FOR MODELLING LEARNING

3.1 INTRODUCTION

In the previous chapter we covered the broad spectrum of theories, concepts, and mechanisms of learning from various scientific outlooks. Such observations had emerged as the result of extensive and diverse research on animal and human learning. The emphasis was on understanding and explaining the learning phenomenon by use of empirical methods of investigation. Formalization and theoretical analysis of such observations was not dealt with to a great extent.

In understanding the nature of learning, the immense volume of data available will not be of any use unless supported by appropriate theoretical hypothesis. Similarly, to verify a theoretical model we need the empirical data and proofs - the best models are always those based on experimental results.

Many psychologists or brain scientists are also model-builders, and construct analytical representations of their subjects of interest. Various examples can be cited from diverse areas of research on learning behaviours or learning mechanisms of the brain. Similarly, workers from other disciplines have attempted to model the process of learning, in terms of vaguer non-empirical notions of their own paradigms. These analytical models of cyberneticians, A.I. scientists and brain-theorists should all, in a sense, be regarded as an attempt to close the gap between the behaviourist's and the neuro-physiologist's models.

A model in psychology or brain-sciences is, normally, based on a specific criterion observed in animals. It often has rigid boundaries for descriptions of a natural phenomena, in the form of a range for variables or sets of possible elements. No great attempt is made to associate the concepts under scrutiny with other notions outside the requirements of experimental situation. The problems are looked at in isolation and the holistic view is not generally favoured.

In the modelling of a phenomenon as complex as learning, with many intrinsic problems (as outlined in the previous chapter regarding the variety and the hierarchy of levels of explanations and investigations), an assortment of diverse approaches have been utilised. At the cellular level, the brain-theorists concerned with the mechanisms of learning in nature, have devised models based on the primary units or elements (i.e., nerve cells) of the brain.

At the higher functional and behavioural levels, some psychologists (cognitive and behavioural) and workers in the information-processing sciences have concentrated on the organizational and executive aspects involved in learning, the models designed by such theorists do not generally stress the details of their internal physical construction, in other words, a 'learning' system is not believed to be the product of 'unique' internal interconnections. Also, in the view of some cognitive or information-processing workers the brain should be studied holistically, since they believe that the summation of the isolated explanations of parts will not yield a true explanation of the whole.

'Modelling' and 'simulation', are different from simple replication, are normally used as tools for: better understanding; condensing of results in coherent forms; and communicating ideas to others. In the modelling of a psycho-physiological notion such as learning an initial description of the system is necessary, the essential features involved should be identified and understood as much as possible. A learning animal itself can be thought of as a complex model-builder; in a constant process of creating models of its percepts, testing various hypothesis about such models, and modifying the models into more efficient ones. The brain of a newly born animal can, also, be analogised to a system of genetically primed units of model-making elements.

A model is the simplified representation of a physical reality, which can be used to draw conclusions about that reality without direct reference to it. The model retains the essential features of the original, and together with the embedded inferential structure of rules can be used for comprehension, explanation or experimentation. It can be one of the three types: 'visual', 'symbolic', or 'physical'.

The nervous system is an example of how the real world can be modelled using neural networks and connections, problems can be solved using a combined physical and symbolic neuronal domain rather than manipulating the real world. However, it must be remembered that irrespective of the complexity or the accuracy of a model it is not the reality but only a representation of it.

The type and the complexity of a model or simulation is stipulated by the particular usage requirements. Simulation is the dynamic execution or manipulation of a model, in simulation predictive methods are used to constantly validate a model against its reality. Three major factors have to be decided in simulation:-
(a) - the 'usage' level, (b) - the 'domain', and (c) - the 'approach'.

In constructing the model of a natural learning process, the subjectivity of the perception and the limitations of the senses of an animal in representing the reality are prime considerations. The animal, using its past experiences, constructs a predictive internal model and conceptualises the reality. Hence, to design an accurate representation of such a process, the intrinsic continuous (space and time) nature of the environment should be incorporated in the model, and the ordering and the scale of various realities maintained. Also, in trying to create an autonomous 'learning' system, the philosophical objection of 'infinite regression' is raised regarding the true objective value of descriptions.

The goal of model-builders in learning sciences is to construct models or machines (physical models) that exhibit the properties of natural learning, hence enabling the discovery of new facts about biological learning. However, to build true models of natural learning, much more should be known about the workings of brain and the way knowledge is represented in nervous systems. The question exists, whether the physiological experiments on animals or neuro-tissue cultures are the quickest and best methods for the understanding of the learning phenomenon, or whether the study of the external effects of learning processes is the more fruitful approach. On the other hand, the objective of some workers in fields such as artificial intelligence is to construct models and machines that outperform humans in specific tasks.

3.2 MODELLING AND SIMULATION

In this section we will briefly discuss the issues involved in modelling and simulation. Also, different types of models and tools used for the representation of models will be outlined.

As mentioned earlier a model is a simple representation of an original entity; it can be used for explanation, experimentation or better understanding. We can only say a model is a 'good' or 'true' model of the original, if its specific functional behaviour follows closely that of the original within a particular contextual use. The specific purpose of an experiment determines the type and the choice of essential parameters and variables to be included in the model.

The concept of 'system' is a useful notion in studying most scientific objects which are distinguished from their 'environment' (though boundaries are arbitrarily defined). Examples of such systems can be found in: biology, physics, mathematics, psychology, cybernetics, economics, etc. Modelling can be thought of as a definite kind of similarity between systems. In general this similarity can one be of two types:-

(1) - Models of behaviour: similarity between the behaviour of two systems.

(2) - Models of mechanisms: similarity between the structure and mechanisms of two system.

For animals, a model of mechanism could be one which displays the essential features of the organization of nerve cells and networks, and a behavioural model could be one that represents a specific behavioural pattern.

The well established theories of systems are able to analyze the similarities of behaviour or structure between two formal systems, and based on the isomorphic properties of systems, judge if one is a good model of the other. However, in the psychological and biological sciences major obstacles are encountered in isolating and identifying the essential features relevant to a system; whereby, detailed models can be devised which suffer from extreme vagueness or imprecision.

In constructing a model, the objective is to represent the original in a simple form; the initial model can either be a 'complex' type, where gradually the non-essential features are excluded, or it could be a 'simple' type, having the minimal properties of the original, where more features are supplemented. Following the identification of the essential features, the task of a model-builder is to discover the various relationships that might exist between such features and devise an inferential structure of rules.

The choice of the medium used for modelling is determined by the particular use it is intended for. A 'qualitative' view of the original can be conveyed in graphical or descriptive models. The 'quantitative' forms of representations involve mathematical, physical, electronic, and computer-based (digital and analogue) models. Also, an 'analogue' of a system can be constructed, which is essentially different from a model, it simply duplicates the behaviour of the original using entirely dissimilar physical properties.

An important phase in model-building is the procedure of 'validation'. The behaviour of the model is matched and tested against that of the original, and various adjustments made for a better correlation. A model can become a substitute for the original for the purposes of experimentation or testing. The exploration of a working model is known as 'simulation', which is used for the testing of various hypothesis about the model. In the course of simulation additional properties might emerge about the nature of the original, these properties can be consequently validated or rejected. A 'good' model should be able to predict the original's behaviour for different variable values in diverse experimental situations.

Finally, the results obtained from simulation and modelling should be translated to the original domain, by making conclusions about the original's behaviour. Care should be taken in the generalization of observations or the introduction of new properties, if too much commitment is made to a particular model many incorrect results can easily emerge.

3.3 MODELLING AND SIMULATION OF INTELLIGENCE AND LEARNING

The mind/brain dichotomy, a dominant feature of many scientific fields until early this century, imposed the belief that even the most detailed understanding of the workings of the brain would not give an insight into the mental activities of animals. The erosion of such views and the development of more mechanistic concepts of mind and brain, has introduced the idea that in principle models and machines that can perform mental processes are conceivable.

The principal supporting evidence for the mechanization of thoughts has been the discovery of various neuro-physiological mechanisms and structures of the brain. These findings have shown that specific physiological units could be correlated with certain mental concepts. The mechanistic view prevalent today is that if you can understand something, you can build a model or a machine to imitate it.

In the explanations of the nature of intelligence, and one of its most fundamental attributes learning, various approaches have been undertaken:-

(a) - The non-biological approach: involves criteria outside the body and the brain.

(b) - The cellular approach: involves the investigations of the properties of nerve cells and their interconnections.

(c) - The structural approach: involves the organizational aspects, and the total patterns of electrical activities of the brain.

(d) - The cognitive approach: sees the brain as a deterministic machine, and analyses the functional aspects of parts of such machines.

3.3.1 MACHINES AND ROBOTS

Historically, the earliest types of modelling of intelligent behaviour can be associated to what is known as 'robotology', or the mimicking of various aspects of intelligent behaviour by constructing machines and models that would appear to have some 'mental' faculty. There is a long history of automated machines made up of gears and pulleys, attempting to copy the living organism in crude ways. Examples of the more advanced generation of such automata were Ashby's Homeostat (1952) and Walter's Machina Speculatrix (1953). These machines were able to show to some extent a modifiability of response under different environmental conditions. The objective in building such physical models was to simulate a behaviour isomorphic with natural learning behaviour.

The technological advances and the techniques available for simulation and modelling, have been some of the major influencing factors in the development of 'learning' models; many of the theoretical models seem to have evolved around the technology present at the time.

The abstract concepts used in the design of a 'learning' machine should in principle be similar to those involved in the design and function of living organism; robots or machines constructed with such concepts in mind, would only be a physical embodiment of these theories, and the type of hardware used should be an irrelevant issue. However, in reality there is an intricate relationship between hardware and theoretical development, and many theories seem to have derived from hardware considerations. Hence, in some instances, seemingly, the problem is approached from the wrong direction.

In spite of this paradox, the machines and robots that have been constructed to demonstrate specific aspects of learning behaviour, such as the early simple conditioned reflex machines or conditional probability computers and the later more sophisticated electronically oriented machines, are not only interesting curiosities of the 'state of art' technology, but are genuine attempts to discover the nature of adaptive behaviour; each generation having more and more functional components identical to the natural learning process.

3.3.2 ANALYTICAL 'LEARNING' MODELS

The earliest types of quantitative models of learning, originating from the science of behavioural psychology, involved the mathematical modelling of various aspects of learning. The most prominent researcher in this field was Estes (1967) who introduced the probabilistic Stimulus-Sampling-Theory based on experimental results on animals. Other mathematical models of learning and conditioning have, also, been developed in stimulus-response studies, and some have been used in teaching and educational sciences.

Another early trend in the modelling of learning was the system and automaton approach in conjunction with logical decision making techniques, examples of which are the neural nets and the self-organising systems; the development of these fields contributed to the introduction of disciplines such as 'pattern-recognition' and 'control system theory'. The analytical researchers of the learning phenomenon, are eager to transfer the problems encountered in learning sciences to a mathematical domain, where they seem to be more skilled and confident in manipulating the criteria and extracting deductions.

3.3.3 COMPUTERS

The introduction of digital computers was a major landmark in the development of modelling techniques. The rapid progress of computer technology, presently in its fifth generation, has given rise to countless discussions and arguments regarding the relation of such potentially powerful machines and the brain. Questions such as: - can computers and brains be equated in any functional or structural sense? - what are the relative similarities? - will one eventually exert a control over the other? - can intelligence be a property of a machine? are raised, and many philosophical and scientific controversies have emerged.

In the early computer simulations of learning, such as Newell and Simon's (1972) 'learning' models, computers were used as tools for making deductions.

Basically, the task of computer simulation of learning in principle involves two stages. Firstly, the discovery of what is going on during the natural learning process; and secondly, how to make computers go through the same process. It is not surprising that in view of our fragmented and limited knowledge of the former the progress in devising successful 'learning' programs has been slow.

A secondary and more recent objective of computer modelers of learning has been to discover methods and techniques which will enable a machine to learn a specific task much faster than humans, yet the process need not have any resemblance to natural learning.

3.3.4 BRAINS vs. 'INTELLIGENT' MACHINES

Brains resemble machines in numerous but generally primitive ways. Yet, many functions of the brain seem to be purposive and directive in nature, with seemingly unique living qualities. A concept such as 'intelligence' must involve all aspects and activities of the brain as a whole. Leaving aside 'animal qualities', to make computers which are functionally identical to the brain, we must make machines with the same input/output behavioural characteristics and the same capability for adaptive modifications.

According to Arbib (1972): "To be intelligent is to perceive the elements of a situation beyond raw sensations." Hence, a complex machine that blindly follows a deterministic set of rules, with no true element of goal-directiveness, would not qualify for this definition of intelligence. It is generally accepted that the term 'intelligence' is a matter of definition, and saying a machine is intelligent does not mean that it has to learn to be intelligent.

Turing's (1950) test for 'intelligence', is a criterion which can be used to establish whether an entity possesses such an attribute; the test involves interrogating the machine in question through a channel of communication, hence, not noticing the obvious physical differences; and if at the end of exhaustive examinations we still cannot establish whether there is a human or machine at the other end, then, we can attribute the machine with 'intelligence'. However, the subjective nature of this criterion must be pointed out, whereby, we can never be sure if the machine or the person under interrogation has any sense of 'understanding' or 'purposiveness'.

3.3.5 CYBERNETICS AND SYSTEMS THEORY

Before advancing to the discussion of information processing and A.I. models of learning, it is relevant here to mention the role and the influence of the science of Cybernetics in the development of 'learning' models.

Cybernetics was defined as: "the science of control and communications in man and machine," by its principal proponent Wiener (1948). The basic characteristic of a cybernetic model is its generality of use for various systems of organization. The concept of 'system' is a primary prerequisite of any cybernetic model, various notions of general systems theory are used in the cybernetic modelling and simulation, such as: input, output, variables, states, operations, transitions, fields, etc. Also, the idea of 'feedback' is a dominant feature of most cybernetic models, especially in designing of automatic control systems.

In modelling a natural learning behaviour or mechanism, the relation of parts are seen as the important factor rather than the analysis of each part in isolation. Furthermore, the animal is seen as a deterministic machine, whose behaviour is the product of the totality of its parts, and with such criteria the supposition is made that an artificial device or system can be made, that like the brain will be able to develop adaptations in its behaviour.

Although, cybernetics has been suffering from diversity and generality of its concepts it has been applied successfully to more specific problems in engineering, social sciences and economics. General system theories and Automata theories are other fields which have benefited from the growth of cybernetics.

3.3.6 INFORMATION PROCESSING

The merger of various notions from fields of cybernetics, system theory, information theory, communication theory, and the methodology of computer simulation has culminated in the introduction of the information processing concepts and models of learning. The computer is used as a machine for expressing and developing different cognitive and behavioural theories. In simulation, the computer language is used to devise a sequential program for actions; the hardware of the computer is of no interest. In information processing models, the transmission of information is viewed using various engineering notions such as: coding, decoding, signal, noise, and channels of communications.

The earliest types of information processing 'learning' models, involved symbolic concept oriented learning, and included many applied A.I. and pattern-recognition models which solved conceptual problems by means of creating goals and sub-goals for the task.

Cognitive modelling is another area of the science of information processing, models of various functional properties of the brain such as the formation and search of memory are simulated on computers. Modelling of skills, knowledge, goals, motivation are some other examples of the tasks undertaken by cognitive scientists. Various 'production-rules', 'strategies' or 'propositional-beliefs' are used in the design of such models.

3.3.7 ARTIFICIAL INTELLIGENCE

Artificial intelligence (A.I.) is a specialised area of information processing sciences, the principal difference of A.I. models from other general system models is the specificity of application in A.I. The goals of A.I. are to design and build intelligent machines that perform tasks normally requiring human intelligence, and to simulate and study the natural learning phenomena. Learning in A.I. is defined as the adaptive changes in a system that enables the system to perform a task or a similar task more efficiently and effectively next time.

The theoretical analysis of learning in A.I. provides a means of exploring various possible methods of learning, since, although through the process of evolution human learning is possibly close to optimally efficient, by no means is the only form of learning.

In comparing natural learning with machine 'learning' we can see some distinct differences: (a) - the natural learning can be a very slow process for certain tasks, such as the acquisition of skills or knowledge, while, in A.I. once a 'learning' model is successful the performance will quickly improve and 'learning' can be easily duplicated; (b) - the perception of a machine is quite poor compared to that of an animal; (c) - the two 'learning' systems may accomplish the same results but in differing ways; (d) - certain mental attributes cannot be modelled successfully on machines, examples are 'consciousness', 'awareness' and 'insight', although attempts have been made to define these notions in machine terms.

A.I. can aid cognitive psychology by means of computer simulation and investigation of certain aspects of natural learning such as: perception, problem-solving, verbal-learning, concept-formation, thinking, memory, meaning, knowledge-representation and understanding natural languages. Also concepts in A.I. can be used to devise computer assisted teaching methods.

There are basically two main forms of 'learning' models in A.I. Firstly, the skill refinement 'learning' models which involve the devising of 'learning' strategies for a particular domain of application, the various 'learning' strategies include: 'rote-learning', 'learning by instruction', 'learning by analogy', 'learning by example', 'learning from observation' and 'learning from discovery'; in these task oriented models, the performance of a functioning system is improved in a more or less automatic manner, at a sub-conscious level.

Secondly, the knowledge acquisition 'learning' models, again involving various 'learning' strategies but in this case within specific domains of knowledge; the main functions of these models are the obtaining of new descriptive information, the perceiving of relationships, and the understanding of the meaning and usage of input data. Recently, the knowledge intensive models involving specific expert-systems have been the dominant area of A.I. research.

A different categorization of A.I. 'learning' models could be made as follows:-

(a) - The models designed for specific task oriented analysis.

(b) - The models used for the simulation of natural cognitive processes.

(c) - The models devised for the general theoretical investigations of all possible learning methods independent of their application.

3.4 PRINCIPAL CONSIDERATIONS AND ELEMENTS OF A GENERAL 'LEARNING MODEL'

Taking the mechanistic view that certain aspects of mind and brain are deterministic and in principle can be modeled, then, it is clear that the main problem of the design of a 'learning' model is one of programming rather than the hardware construction of equivalent elements.

In devising such models of learning, certain criteria and considerations should be applied. A choice can be made whether the model is to be 'initiallised' with a preformed framework of knowledge about the learning task, or more or less start the learning from zero like a child, the level and complexity of the child-model should also be considered. The domain of the 'learning' machine or model could be either a specific fixed domain or a more general changing one, in which case the model should be designed to be more critical in separating the significant experiences.

The relative contribution of 'instructed' or 'experienced' inputs to the model should be assessed in the 'learning' strategies involved. Unless the 'learning' model is based on purely abstract notions which do not directly relate to any natural phenomena, then natural concepts such as punishment and reward or other mental and physiological parameters are normally incorporated. Also, a random element can be included in the model to enable the discovery of novel situations. Other important considerations could be: the criterion for the self-realization of the model; and the ultimate goal of the model, which could be based on notions of seeking novelty, change, stability, etc. The basic elements of a general, simple 'learning' model can be enumerated as follows:-

(1) - Filtering element: to limit or focus the relevant data from environment for utilization, this element can be an instructor or a teacher.

(2) - Memory element: to store in dual form of short and long term the essential features and information extracted from various experiences.

(3) - Evaluation element: to compare with a standard, or ascertain the achievement of goals, usually manifested by reward/punishment centers.

(4) - Learning element: to modify and improve the response by translation of the evaluation element's outputs.

(5) - Executive element: to select the response for each stimuli.

(6) - Generalization element: to extract specific concepts from the mass of stored experiences in memory.

(7) - Discovery element: to distinguish novel or significant experiences.

(8) - Planning element: to devise strategies and theories for future actions.

3.5 "NATURAL" LEARNING MODELS

Historically, the first attempts at the modelling of aspects of learning involved the 'natural' approach to the subject, these models exemplified various qualities of natural learning using empirical observations. The 'natural' model builders do not explore the possible alternate learning processes or methods, but are mainly interested in the simple representation of the psycho-physiological concepts in learning. The models of 'natural' learning have two distinct types: (a) - the models of natural learning behaviours, and (b) - the models of natural learning mechanisms.

3.5.1 MODELS OF NATURAL LEARNING BEHAVIOUR

The emergence of analytical models of learning behaviour can be traced to the work of experimental psychologists who were involved in the empirical studies of learning. The earliest learning theories, proposed models which were mainly based on a descriptive explanation of processes involved. The pioneering adherents of such theories namely Ebbinghaus, Thorndike, Pavlov, Gutherie, and Skinner did not give particular quantified versions of their criteria.

However, some quantitative methods were established for the recording or describing of learning behaviour; the strength or latency of a stimulus or response could be measured in terms of its amplitude or relative frequency, attempts were also made to fit the results of the experimental work into specific mathematical functions. For example, mathematical 'learning functions' such as hyperbola, exponential growth or other functions were proposed which could depict certain aspects of learning experiments such as forgetting or performance changes over successive experiments. Yet, none of these learning functions involved any axiomatic considerations about learning. Gradually, the trend inclined more towards the precise analytical models of learning behaviour, in an attempt to explain and predict the exact details of experimental results.

The early programmatic theories of conditioning and learning processes were supplemented by Hull's (1943) analytical theories which strongly emphasised the need for the development of quantitative learning theories. He postulated sets of both descriptive criteria and quantitative mathematical equations, mainly based on generalized empirical results; despite the rich analytical flavour of Hull's writings, his most important contribution to learning research was the comprehensive and clarified nature of his 'qualitative' work. Hull's formal knowledge base for learning processes, also the influence of other disciplines such as 'information theory', were instrumental in the introduction of mathematical learning theories from the late 1940's, these theories included the stochastic as well as the deterministic views of behaviour, and generally involved the estimation of some parameters.

Mathematical learning theories and models have on the main concentrated on the associationist approach and the experimental situations which were exploited by Hull, Skinner and other S-R researchers, but generally for simplification purposes do not incorporate the immense range of variables and parameters which influence animal learning, many factors such as the 'drive level' or the 'stimulus intensity' are taken to be fixed throughout an experiment.

The range of behavioural phenomena investigated by mathematical learning theories include: 'classical' and 'operant' conditioning, 'generalization', 'reinforcement', 'extinction', 'serial-learning', 'discrimination-learning', 'paired-associate-learning', 'memory', 'imitation-learning', 'concept-formation', and many other aspects of learning and conditioning.

Such theories and models have also been applied in the optimization of educational teaching programs; for the optimal determination of methods of presentation, amounts of teaching material, or various other features of classroom organization. The formulation of these theories if done correctly can result in a very accurate model of the learning situation and a vast amount of useful information can be gathered from the simulation and analysis of results.

More recently some cognitive hypothesis have also been represented in mathematical forms. In the following we will briefly discuss Estes' (1967) 'Stimulus-Sampling-Theory' (SST), one of the most prominent and best developed mathematical models of learning behaviour. Many of the ideas in SST have also been translated into theoretical notions in other disciplines such as cognitive-modelling or information-processing.

Stimulus-sampling-theory started as a formalization of S-R associationist ideas. The basic assumption is that an organism 'learns' by attaching new adaptive behaviours to stimulus situations which formally did not have an appropriate response. The effectiveness of an adaptive response is evaluated by the 'reinforcing' outcome. This use of reinforcement as the essence of the learning process, as referred to previously, is the basis for explanations incorporating the empirical 'law of effect'. According to SST a subject experiencing a reinforcing outcome (O) as a consequence of response (R) to stimulus (S) forms a S-R-O sequence in internal representations and learns associations of all pairs of elements S-R, R-O, and S-O.

Stimulus-sampling-theory considers the behaviour to be a probabilistic or stochastic process. The choice that a learning animal makes in any situation (e.g., learning a maze) is looked at in terms of the probabilities of responses, according to criteria based on groups of animals or numbers of successive trials. The behaviour is thought to be causally determined, but in view of the large number of possible factors involved such as: genetic make up, motivation levels, sensory differences, fluctuations in recalling memory, variety of environmental stimuli, and many other un-measurable factors, the best prediction of behaviour is considered to be a probabilistic one. Although, it is assumed that if all such causal factors were precisely known then the behaviour could be completely deterministic.

The main variable of a statistical learning theory is the probability of response of a subject at any point of time,the probability of a correct response should approach 1.00 as learning progresses. Many inferences and predictions can be made from the simulations of statistical 'learning' models such as those in SST, but care should be taken to avoid the pitfall of fictitious deductions. An SST model is normally based on a small sample of individuals, from which conclusions are made about the populations of such individuals. The performance variations between the members of a population, such as differing response tendencies or learning rates can be incorporated within the model by the inclusion of a 'variance' factor.

One of the basic assumptions of SST is that a learning organism that is confronted by the environmental stimuli, only has a limited sample of N elements (normally fixed) available for processing, the sample may vary due to environmental or the subject's fluctuating circumstances. And it is postulated that each element in the sample can be assigned a particular probability for each trial. Another assumption is made about the connection of responses with the sampled stimuli, a deterministic criteria is used whereby a response for a particular stimulus is fixed at any specific time and the 'state' of the 'learning' model is the listing of such stimulus-response pairs, these connections and hence the state of the system changes from trial to trial. Although, the probability of a response associated with a particular stimulus is 1.00, at any one time there is a probability attached to the occurrence of a particular response within a sampled set of stimuli, this is due to the probabilistic nature of the stimuli.

The 'performance' is judged and reinforced by evaluating the 'reinforcing outcome' after a subject responds to a particular stimulus sample. This reinforcing outcome changes the conditional connections of the elements sampled in the trial and updates the model. By utilising this general principle, successive trials results in the attachment of 'correct' or 'rewarded' responses to increasingly more stimuli elements and the gradual detachment of inappropriate responses. The phenomenon of 'extinction' is also embodied within such process, whereby the stimulus-response connections which are not rewarded frequently are slowly phased out.

Using such basic criteria on stimulus-samples, linear learning-equations are formulated for various learning experiments or situations, and by following the progress of different S-R connections global inferences or predictions are made about many aspects of the populations of such samples. Of special interest are the small element samples, for which SST can be applied much more consistently and precisely. Hence the models which are based on simple conditioning or learning situations can predict the behaviour much more accurately. One by-product of such investigations has been to propose systems of basic building blocks of learning elements which could be used in synthesising much more complex learning processes. More recently, the stimulus-sampling-theories have shifted from the purely response oriented associative views to the more cognitive event and memory formation ideas.

As well as SST, other mathematical methods such as Bush and Mosteller's (1955) have been developed to model specific aspects of natural learning behaviour. A different class of models of natural learning behaviour, those that strongly emphasise or depend on hardware will be dealt with in later sections - the distinguishing feature of the design of these models is the particular attention which is given to the intricacies of their physical construct.

Simulations and models of learning behaviour involving the use of computing machinery as tools for the implementation of logical deductions will be investigated later, these models mainly attempt to duplicate the human's higher mental processes such as thinking and problem-solving. Other non-computer based models of various aspects of mental activity have been devised by cognitive-modelers; generally, in the cognitive type models, the inclination is away from the simpler aspects of learning and conditioning. Simple natural learning situations such as conditioning or trial and error learning can be easily simulated by a non-analytical programmatic computer model, displaying the appropriate behaviour and characteristics of the process, yet it should be recognised that to investigate any novel aspects of such learning situation, it will be necessary to dispense with the model and create a more complex and elaborate one. The new model could be further explored for other crucial assumptions. The problem envisaged with such computer simulations is that although a good fit to empirical results can be achieved and interesting data obtained, no important theorems in the accepted mathematical sense can be derived or proven, since the emphasis is heavily on the simulation aspects of the learning process.

There is a gap between the computer simulations of learning behaviour, such as the A.I. programs, and the simulations of learning in experimental psychology, the principal reason being the difference in background and approach. In psychological simulation the model tries to mimic step by step the process which a learning organism goes through, while in A.I. the modeler is interested in efficient programs that perform a task regardless of how a living organism might achieve the same task.

3.5.2 MODELS OF NATURAL LEARNING MECHANISMS

The mechanism by which the learning process is manifested within an organism's nervous system has been a prime subject for the model-builders of many scientific fields, some of the diverse processes and devices that have been employed in such models include: the flow of fluids in pipes, optical vibrations, electro-chemistry, animal electricity, wires and conductors, electrical and electronic circuits, and digital and analogue computers.

In this section we are basically interested in the models of learning mechanisms more or less based on the structural features of the natural processes of learning. Models involving functional equivalences of natural learning with no regard for the underlying biological mechanisms will be dealt with later, such models mainly incorporate the information flow and processing aspects of a learning animal rather than the specific correspondence of elements in physiological learning systems.

The principle question in the modelling of natural learning mechanism is the physiological basis of association. Although, the physiology of learning can be modelled in many different hierarchical levels (from biochemical to total electrical activities of brain regions), generally the most fruitful investigations of the learning mechanisms have been carried out by the modelling and simulation of nerve-cell action and nerve-network activities.

In the modelling of the behaviour of a single nerve-cell various characteristics and functional parameters of the nerve-cell activity are duplicated in conceptual, mathematical or abstract forms. The degree of elaboration is determined by what parameters are incorporated, some examples of such cell characteristics are: excitation and inhibition thresholds, synaptic integration, coding and decoding of pulses, pulse generation, resting and active potentials, conduction; and many other features of nerve-cell activity. Such individual nerve models have also been constructed in hardware using physical or electrical components. The objective of devising such models is to develope networks made up of these elements, also to test and discover specific functional characteristics of individual elements and networks of cell-models, hence predict and understand the behaviour of real neurons.

Simple idealized neuronal models were devised as a simplified tool for such investigations. The earliest and the most notable examples were the "formal neurons" introduced by McCulloch and Pitts (1943), initially these models were abstract configurations with several inputs and one output, which could have been excited or inhibited at particular threshold values, later models were developed to resemble the natural nerve-cell more closely.

The introduction of such abstract models of the neurons allowed the simulation and manipulation of networks of such elements using formal logic notions and mathematical operations. The model could also be used as a 'computing element' in the engineering sense and many experimental physiological observations could be simulated on computers, any new properties discovered during such simulations could be further verified or refuted by experimentation, hence contributing to the better understanding and explanation of neuronal activities. Even the simple abstract single neuron model can be made to display a kind of simple learning. By changing the weighting or the threshold parameters of the model according to a specific algorithm, the model can learn to distinguish between stimuli received more frequently from the less frequent ones.

The collective behaviour of nerve-cells have been also extensively modelled. Firstly, by the modelling of biological nerve-networks; these types of models have been mainly based on the peripheral sensory or motor neurons, since these neurons are easier to isolate and observe as a system (e.g., cat's visual system). These models provide a theoretical framework for the investigation of activities of groups of neurons, using which deductions or inferences can be made about the natural nerve-networks. Initially, the models were represented in a simple conceptual or pictorial form but later more abstract mathematical and computer-based models were developed which contributed to the establishment of new disciplines such as 'pattern-recognition', 'robotics', 'self-organising systems' etc. An important discovery which has emerged from these nerve-net models is the significant role of the process of 'reciprocal inhibition' in nerve activity, whereby the excitation of a cell is accompanied by the inhibition of other cells.

Secondly, the modelling of the nerve networks is carried out using groups of abstract single nerve-cell models developed previously. The behaviour of such networks being composed of more precise analytical units, can be defined and analyzed much more readily. The problems encountered are analogues to many problems in logic nets and the mathematical notions and operations of Boolean Algebra could be applied to such models. In comparing this type of models with the natural neural-networks, we can see that to have the same degree of efficient and reliable operation we need to incorporate into the model a high degree of redundancy which is prevalent in all natural systems, hence duplicate and cross over networks and extensive inter-connectivity is the feature of most networks of model neurons.

Various reflex conditioning experiments can be demonstrated using networks of abstract models of neurons, even more complex discrimination and performance capabilities can be shown by the appropriate manifestation of such networks. Using a criteria for 'success' or 'failure' (punishment/reward) and devising feedback channels from the environment, autonomous 'learning' systems have been developed that are able to adaptively modify their behaviour under various situations. Rosenblatt (1958,1962) showed that in a randomly interconnected network of model neurons an orderly discriminatative performance can be produced, and more recently other elaborate and complex neural models have been developed that are able to simulate interesting features of conditioning and learning.

Although, the workers in this field do not claim that such automata created using abstract models of neurons have any direct similarity to the natural processes involved, nevertheless they contend that by the use of a simplified model of neuron they are able to show some of the capabilities of a learning organism.

A criticism of the majority of the neural-models is that they favour stimulus-substitution notion in the formation of associations, and the idea of goal-directiveness is not generally featured, it is assumed that stimuli and responses have corresponding neural-pools and the learning of stimulus-response patterns is accompanied by the formation of the neural equivalent of such associations.

In the physiological sense we know that the temporary proximate activity of two nerve-cells causes them to become associated in a primitive or basic manner, so that one can now excite the other while previously was unable to.

A possible explanation of the mechanisms involved in the formation of associations during learning was put forward by Hebb (1949), in his so called "reverberating circuit" theories, he postulated that synapses modify or activate during the formation of the associative memory which plays an important role in the process of learning; according to Hebb the cell assemblies are formed as a result of increased efficiency of 'excitation' of one neuron by another neuron. Later, Milner (1957) postulated that in addition the increase in the 'inhibitory' synaptic efficiency is also an important factor. The investigation of these and other similar theories using the neural modelling techniques and various computer-based simulations has shown that indeed cell assemblies can be formed around each block of input cells, the excitatory synapses become strong and the inhibitory synapses become weak, however the inhibitory synapses are shown to be more dominant. The neural activity can propagate in the form of oscillations, either in rhythmic unison or bursts of activity. These observations loosely correspond to the characteristics of the EEG recordings that are made from the brain activity of an animal engaged in the learning of a task.

Although, no direct physiological proofs have been found for Hebb's fundamental postulates, and clearly memory and learning are much more persistent than initially implied by the purely functional changes of Hebb's theories, still many adaptive 'learning' and self-organising models have been devised using such criteria. A follow up to this line of work has been a considerable amount of research into the process of memory formation, and of particular interest to learning sciences, the findings on the organizational and functional aspects of the associative memory formation within the LTM system.

3.6 "ARTIFICIAL" 'LEARNING' MODELS

The animal learning mechanisms and processes have been studied and modelled intensively in disciplines rooted in biological or psychological sciences. In spite of having the same final objectives as the 'natural' modelers 'artificial' models of learning have, also, been developed by workers from many diverse fields of science unrelated to biology or physiology. In such areas the problems are usually tackled by attempting to develop 'simple' models of natural learning-processes which incorporate some abstract criteria not readily seen in nature. Subsequently, these models are analyzed and improved in an effort to approach the complexities of biological learning-processes. The emphasis in most such 'artificial' models are on the underlying techniques, theories, media or the physical hardware used in the model construction, and not on the empirical observations of natural learning phenomena.

Scientists in the more theoretically inclined sciences of Cybernetics, System-Theory, Automata-Theory and other analytical subjects, have developed models based on the notion that learning is in fact a universal phenomenon applicable to all types of living and non-living systems capable of directive behaviour, the animal learning being a special realization of this phenomenon in nature. Using various stochastic or deterministic metaphors, adaptive 'learning' systems are devised. On the one hand, general and abstract cybernetic systems are developed encompassing all types of adaptive processes, and on the other hand, specific and precise systems with well defined boundaries are designed in fields such as adaptive-control-systems, dealing with particular real-life problems.

A different approach in the development of artificial 'learning' models has been to construct hardware realizations of different learning theories in the form of physical models. The technological advances and tools available to such model-builders have been the major factors in determining the direction, the complexity and the accuracy of these models. The building of 'robots' that can exhibit true learning has been the ultimate objective of researchers in the science of Robotics; other workers have devised mechanical, electrical or electronic models which demonstrate some features of conditioning or a learning-process. Although, these hardware models have been mainly used to simulate or test theories previously developed in the more abstract learning sciences, in certain instances, the actual physical characteristics of a specific hardware device is used as the fundamental criterion and the starting point for subsequent developments. The hardware oriented models will be discussed later and their contribution to the learning science analyzed.

Finally, in the past 30 years with the advent of computers and the explosive development of related sciences, 'learning' models based on the information-processing notions have been devised, whereby, the flow of information during the learning process is depicted on computers in the form of programs. Workers in the field of Artificial-Intelligence (A.I.) or Cognitive-Psychology design 'learning' models which can be simulated on computers, displaying various behavioural or cognitive aspects of learning-processes without necessarily following the 'natural' way. Elaborate systems have been constructed that can 'learn' specific tasks in particular domains, but they are far from the definition of versatile animal or human learning systems. The computer in such sciences is seen as simply a tool for the simulation of the model, but the implicit dominant role of computers in the design and development of these artificial 'learning' models must be emphasised.

3.6.1 'LEARNING' MODELS IN CYBERNETICS

In this section we will briefly sketch the historical background and the development of the science of Cybernetics, and describe briefly some of the underlying objectives of constructing cybernetic-models.

The origins of the name 'Cybernetics' can be traced to the ancient Greek word of 'KYBERNETES', meaning the 'steersman' in the navigational sense. However, philosophers like Plato used this term to denote "the art of steering" in the governing of various activities, the word 'KYBERNETIKOS' hence implied a knowledge of a subject and the ability to remain in command in order to reach a goal. The next landmark in the use of word 'Cybernetics' was the introduction of this term by the 19th century mathematician and physicist Ampere, he used "CYBERNETIQUE" in the context of 'statesmanship' with reference to politics.

The introduction of the modern terminology of 'cybernetics' is attributed to Wiener (1948), generally considered to be the founder of this new branch of science. In an attempt to unify subjects from a variety of fields such as biology, mathematics, physiology, psychology, electrical engineering, control and communication engineering, Wiener defined the distinctly new discipline of 'Cybernetics' as:-

"The science of control and communication in the animal and the machine."

The development of cybernetics has been highlighted by many collaborations with different disciplines of science, followed by the specializations in the form of newly created independent field of study. The main reason for divergence of such specialized subjects away from the mainstream cybernetic research, has been the lack of a formal 'cybernetic' structure for unifying various concepts.

During the late 1940's with the appearance of digital computers, devices that were capable of fast and complex symbolic-manipulations, cybernetic ideas were tied with many mathematical notions in formal-logic, the whole area of computational machinery became a major area of interest for cyberneticians but gradually the trend of such work drifted away from cybernetics towards the more specialized subject of 'computing'. As well as logic, other mathematical theories and concepts have had brief or more permanent and deep coalitions with cybernetics.

Cybernetics can be seen as a science which cuts across many other established sciences, it is primarily interested in the interactions and the functional aspects of an entity, the energy and physical aspects are of only secondary importance. Although, the machines and mathematical worlds, were from the beginning the principal domains of enquiry for the researchers in cybernetics, the biological and physiological mechanisms that could be modeled or scrutinized analytically were of strong concern too. For some workers (specially in the eastern-block countries) cybernetics is used even in a more broad sense, encompassing many mathematical or statistical socio-economic theories.

As evident from Wiener's definition of cybernetics, a principal metaphor in cybernetics is that an 'animal' is considered to be a kind of 'machine', in particular, the brains and the nervous-systems are equated with computer-like processing machines. The effort of some workers in cybernetics has been directed towards the design of machines, that in principle can emulate some intelligent behaviours or brain-functions. The biological results are analyzed and reconsidered in mathematical terms, these quantitative frameworks allow much higher degrees of precisians and manipulations of results than the ordinary linguistic descriptive models. Although, the gap between organism and machine is immense, effective cybernetic theories and models have been constructed to show some similarities with living phenomena, but the resemblances have been primarily at performance level.

Wiener's formal definition of the science of Cybernetics does not expressly state the topics and the range of criteria that can be investigated by this discipline, it specifies: (a) - the types of objects for which the cyberneticians could formulate their hypothesis about (man and machine); and (b) - the points of view with which the problems are approached, defined and analyzed (control and communication).

The evolution of cybernetics and its applications have shown that, possibly, Wiener's definition is too general to account for the distinctive yet diverse views that have been developed, and the type of the problems that have been tackled in cybernetics; other definitions of cybernetics emphasising the 'system', the 'control', the 'information' or the 'structural' and 'organizational' aspects have also been devised.

The principal aims of the science of Cybernetics according to George (1973), who also considers "the pursuit of artificial intelligence" as the central feature of cybernetics, are:-

(1) - To construct effective theories, with or without actual hardware, which realize the principal functions of humans.

(2) - To simulate the functions of human behaviour by the same logical means used in human beings.

(3) - To produce models which are constructed from the same colloidal chemicals as are used in human beings.

The approaches to the above objectives have ranged from 'theoretical', 'experimental' to 'engineering'; many overlaps with other disciplines have created 'applied-cybernetic' subjects in areas such as economics, neuro-physiology, mathematics, etc.

3.6.2 ANALYTICAL TOOLS AND TECHNIQUES USED FOR MODELLING OF LEARNING

In the following we will briefly look at some important methodologies, concepts and analytical tools that have been incorporated in various 'learning' and 'adaptive' models - in particular, many cybernetic models.

(i) - SYSTEMS AND CONTROL SYSTEMS

The concept of 'system' has been widely used for the description and analysis of the learning process. The framework of terminologies and theories developed in 'system-theory', have helped the formalization of many processes in the living world as well as the inanimate. We will enumerate some of the principal elements of system-theory and other notions relevant to our discussion of 'learning-models'.

A 'system' in general terms can be defined as: 'a set of interrelated elements'; in the case of 'learning' systems, the concepts of 'change' and 'variety' are of special concern, hence for our purposes the system can be considered as 'a set of variables' - a variable here is any measurable quantity. In 'physical-systems' everything not included within the system can be considered as the 'environment' of the system, but usually a more restricted form of environment is chosen which only includes the elements of the 'universe' relevant to the design of the system.

The system and the environment must not be looked at in isolation, whereby the environment is only seen as the external forces which act on the organism from the outside. The environment and the system must be considered as mutually interactive, or in the case of the more abstract systems as mutually interrelated.

Every real physical organism can be attributed with an infinite number of variables, a major task for the scientist devising a system is the 'identification' of the appropriate selection of variables from the indefinite list of possibilities. A 'good' system representation is one that only includes variables which have a bearing to the 'context' and the 'resolution-level' of the problem.

The 'state' of a system is defined as the set of values which the variables take at any instance of time, 'inputs' or 'stimuli' to the system denote all the parameters of the environment that effect the system, similarly, 'outputs' or 'responses' of the system are all effects of the system on the environment. A schematic diagram of the main system components are outlined in FIG.3.1.

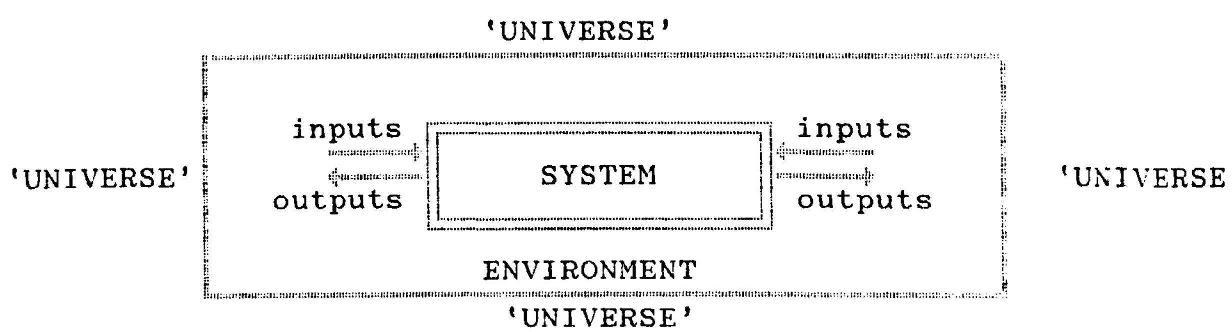


FIGURE 3.1. A general system configuration with its main elements.

The properties of a system can be divided into two different basic groups of: (1) - **behavioural**, and (2) - **structural.**

The behaviour of any general system can be completely and uniquely defined by: the 'state-transition-functions', which determine how the state will change under the influence of various inputs; and the 'output-functions', which determine what the system outputs will be for different inputs, given a specified state. The 'behaviour' of the system can be represented by a 'line of behaviour' or a 'phase-space', whereby the transitions from an 'initial-state' are shown respectively as successive states drawn against time intervals or as 'vector' representations, a 'field' is a phase-space containing all the possible lines of behaviour.

The structure of a system describes, on the one hand, the organizational aspects or the 'couplings' between the elements of the system, and on the other hand, the interactions between these elements. The coupling between the elements of a system can be achieved by 'series', 'parallel' and 'feedback' methods, as shown in FIG.3.2.

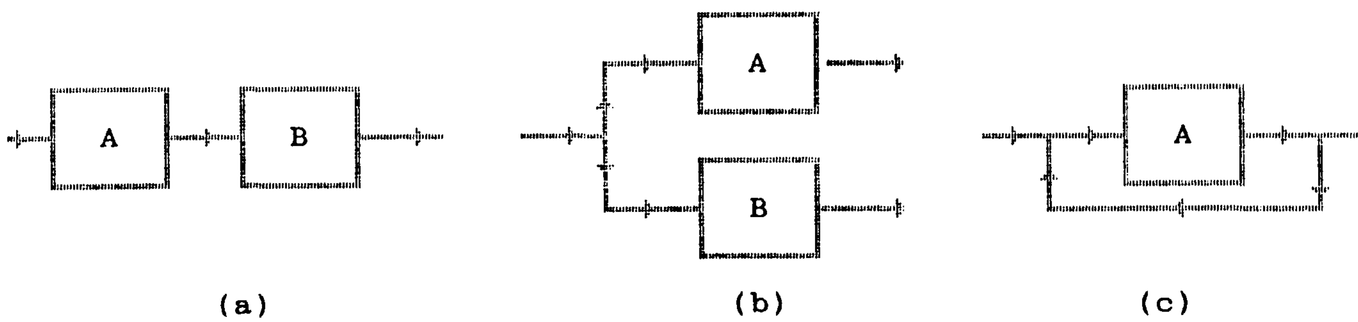


FIGURE 3.2. The three types of coupling between the elements of a system: (a)-series, (b)-parallel, (c)-feedback.

The notion of 'feedback' is an important feature of most systems, an organism or a system is said to have 'feedback' if its outputs have some bearing on its inputs. A special class of systems the 'servomechanisms' or 'negative-feedback-systems' have a special significance for learning modelers. These systems use the feedback loop to incorporate an error reducing element which enables the system to achieve a 'desired' state or performance. It was the analysis of this type of systems that influenced the thoughts of the founders of modern cybernetic theory; for example, Wiener's work on the aircraft tracking servo-systems.

Systems have also been classified into distinct types and groupings such as: 'Physical' against 'Abstract'; 'Continuous' against 'Discrete'; 'Artificial' against 'Natural'; 'Open' against'Closed', where in open systems all possible interactions between an organism and environment are considered, while in closed systems (or partially closed systems) no interaction (or limited paths of interactions) are considered; 'Deterministic' against 'Stochastic', where the response and the new state of a deterministic system can be uniquely determined by its previous state and the present stimulus, while, the behaviour of the stochastic-system for any given input-state pair is only predicted in terms of the probabilities of outputs and new states.

As mentioned earlier the two major properties of a system are its 'structure' and 'behaviour'. The types of problems that system-theory can be applied to, invariably depend on the investigation or the discovery of one or both of these fundamental properties. A specified system behaviour can be either the result of a unique system structure or a class of structures. In principle, the problems are classified as follows:-

(a) - The 'synthesis' of a system: a specified behaviour is known, a structure is to be designed so that it can exhibit the prescribed behaviour.

(b) - The 'analysis' of a system: the structure of a system is known, the task is to determine the behaviour of such a system.

(c) - The 'black-box' problem: a partial or no knowledge of the structure exists, the problem consists of determining the behaviour by empirical experimentation and inferring a hypothesis about its structure.

It must be noted that learning sciences utilise all three approaches (in varying degrees) in the design of 'adaptive learning models'.

A special class of systems are those called the 'control-systems'; the word 'control' is normally understood to mean 'regulate', 'direct', or 'command. The criterion of control in systems is applied to 'dependant' variables or elements that interact and influence each other's behaviour. Control, is a subjective and relative notion, depending on various contextual limits imposed on the performance of a system, the actual physical connections of variables are of no great consequence in the definitions of this concept.

In the most general and abstract sense every physical or non-physical object can be considered as a control system, and for every system an arbitrary number of control elements defined. However, in science, normally, the notion of control system is defined on more objective basis as: systems that can 'actively' regulate, command, or direct; hence excluding the thermodynamic-equilibrium-seeking systems. These control systems are sometimes referred to as 'purposive', 'teleological' or 'goal-directed' systems.

Control systems are normally classified into two catagories: 'open-loop' and 'closed-loop' or 'feedback' control systems; in open-loop control systems, the control-action is independent of the output; while, in closed-loop control systems, the control-action is influenced by the output. However, the arbitrary and subjective nature of the definitions of control-action and other elements of systems must be remembered. The general scheme of a feedback control system is outlined in FIG.3.3.

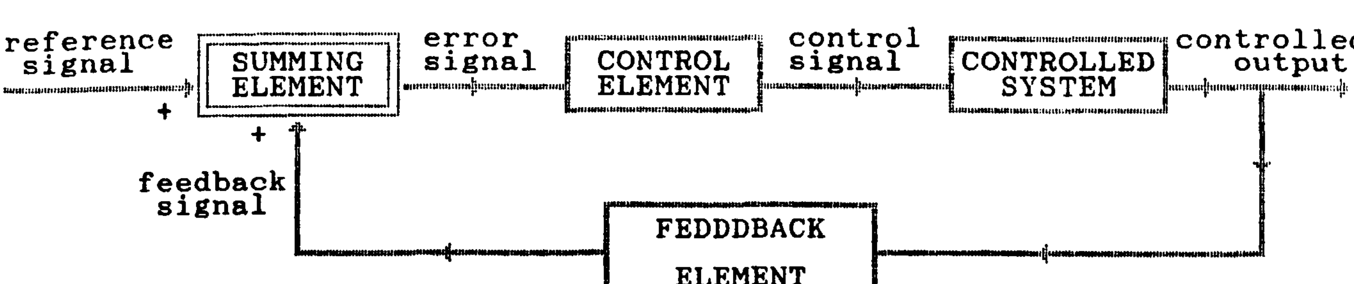


FIGURE 3.3. A feedback control system: control is established using the error signal which is a summation or the difference between the reference input and a factor of the actual output (the feedback signal).

(ii) - OTHER MATHEMATICAL METHODOLOGIES USED IN MODELLING OF LEARNING

The use of mathematical techniques in the modelling of learning has helped to devise formal systems that by utilising accurate criteria are able to describe, ask questions and provide answers about specific problems. Mathematics is a precise language with a structured axiomatic framework where 'consistency' and 'completeness' are prime concerns.

Of course, the researcher in learning is not interested only in 'pure' abstractions in mathematical domains, but also is involved in the design of some 'applied' mathematical techniques for realization of his models, predominantly by a digital computer. In the following a selection of mathematical topics of interest will be briefly discussed.

(a) - SET THEORY

A 'set' can be defined as simply a collection of elements, the set-theory has a well formed axiomatic framework of representations and functions; these notions have been employed in models where the representation of 'class' of objects or elements have been of concern. Ashby's (1952) formal representation of the property of Homeostasis (regulation) in organisms, and many other models of human thinking and memory, are examples where the set-theory notions are utilised.

(b) - THEORY OF LOGIC

Logic is generally considered to be the foundation of mathematics, logic has been called the 'grammar of reason' because of its intuitive similarity with the process of mental reasoning and thinking. But unlike the ordinary language which has an implicit vagueness in its construct, logic has been abstracted on orderly and objective criteria. Logic has been extensively applied in the understanding of cybernetic systems, it has been used to perform 'deductions' (generalizations), 'inductions' (proving statements) and finally 'abstractions' within such systems.

In logic variables stand for 'classes' of objects; an algebra for classes (the 'Boolean Algebra') has been devised, one interpretation of Boolean algebra is when it is performed on classes of propositions or statements. The Boolean operations of 'and', 'or', 'negation', 'class-inclusion', 'class-exclusion' and 'equivalence' preserve their intuitive linguistic meanings when acting upon propositions. For this type of symbolic-logic, the thoories of 'propositional-calculus' have been developed, which by using simple arithmetics of propositions allow the combinations or manipulations of statements. Hence, the validity of relationships between statements in a particular mathematical system can be examined and various theorems proved.

An extension of propositional-calculus is the methodologies and theories of 'predicate-calculus', predicate-calculus is utilised in the majority of 'theorem-proving' research of today. The predicate calculus is concerned with a more detailed analysis of propositions, allowing for some inner structural aspects of propositions as well. A 'predicate' is a function that maps terms onto truth values 'T' or 'F' (conventional notions of 'true' or 'false').

(c) - PROBABILITY THEORY

'Probability' is the single notion of mathematics most widely used in the learning related sciences. Probability theory is the formal foundation of statistical inferences, sampling and survey of problems, and the design of many experiments. The probability of an event occurring is calculated on the basis of the three criteria:-

(1) - The logical and semantic analysis of possible outcomes.
(2) - The frequency of occurrence of events.
(3) - The speculative judgements of possible future outcomes of unique or novel events.

Generally, a language or a sequence of symbols which characterises events and has probabilities associated with each symbol is called a 'stochastic-process'. The notion of stochastic process and the accompanying mathematical techniques have been applied to many non-deterministic 'learning' systems, in particular, to natural systems where there are no certainties associated with outcomes and events. A special class of stochastic-processes are the 'Markovian-Processes', they have the added property that the probabilities of symbols in a so called 'Markov-Chain' depend on a finite number of previous symbols, these ideas have also found important applications in the theory of communication.

(d) - GAMES THEORY

The theory of games as founded by von-Neumann (1947) has been used as a mathematical tool in some 'learning' models. The game theory methods can be applied to many 'goal-directed' systems where 'wining' optimal strategies can be achieved by probability estimations. The notions of game theory which involve chance situations between two or more opponents, can be further extended to games where one opponent is the environment, hence 'games-against-nature' could be devised. Similarly, the subject of 'dynamic-programming' of special interest to cybernetics and A.I., has been a development arising from the field of game-theory.

(e) - STABILITY THEORY

The formal concepts of 'stability', 'steady-state' and 'equilibrium' have been incorporated in the design of some cybernetic type models of adaptive-processes, in particular, those based on biological systems. An organism or system is said to be seeking-stability, if its behaviour is seen to be directed towards a state of equilibrium. Many artificial or natural systems can be observed that have this 'regulating' property, they react to modest disturbances from an equilibrium-state and normally by the use of feedback loop try to regain the steady-state. The process of 'homeostasis' is the biological counterpart of the stability concept; some examples of this process are: the physio-chemical cellular regulations, the automatic regulations of individual anatomical organs, or the central-nervous-system/hormonal regulations of totality of animal.

(f) - INFORMATION THEORY, COMMUNICATION THEORY

The concept of 'information' may be considered as one of the central issues in science, the implicit relation of information to learning is analogous to the kind of kinship that the concept of numbers have to the subject of mathematics. The word 'information' is used to signify a quantifiable variable which can be conveyed by a variety of physical, symbolic or other means. In the most general sense, information can be defined as 'the measure of the amount of organization'.

Shannon and Weaver (1949) laid the foundation of what is now known as 'statistical-communication-theory' or 'information-theory' during the late 1940's. The theory of communication according to Weaver (1949) can be regarded in the following three distinct levels:-

(1) - The 'technical' or 'syntactical' level, interested in the accurate transmission of symbols.

(2) - The 'meaning' or 'semantic' level, concerned with the conveying of precise inferences about the transmitted symbols.

(3) - The 'effectiveness' or 'pragmatic' level, concerned with the degree a message is able to influence the receiver's behaviour.

The classical theories of communication only deal at the technical level of problems. However, in learning related sciences all three levels are investigated. The communication aspects of a system are looked at in a broad sense, the total interactions of a system and its environment are taken into account, and also attention is paid to the 'meaning' and 'effect' as well as the quantitative measures of the message.

The fundamental components of a general communications system are shown in FIG.3.4. The messages that originate from the 'source' are 'coded' and the resulting 'signals' are 'transmitted' through a 'channel' of communication which may be distorted by 'noise', at the other end of the channel, the signals are 'received' and after being 'decoded' to their original form reach the 'destination'.

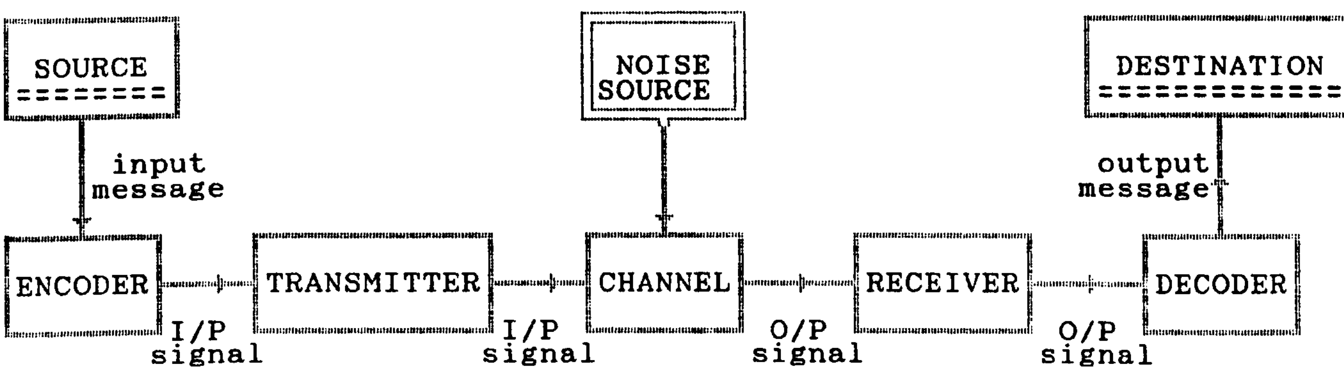


FIGURE 3.4. A schematic representation of a typical communication-system and its principal elements.

In the science of information-theory, a statistical view is favoured in dealing with the problem of message or symbol transmission, hence communication is seen as a kind of 'stochastic process'.

Various quantitative measures and theories have been developed for different aspects of communication, both in the engineering oriented discipline of communications and also independently in the mathematical field of probability theory. The amount of information ('bits') that could pass through a channel can be measured; the 'capacity' of a channel of communication can be determined; the 'uncertainty' or the 'redundancy' of a communication system can be defined in terms of the 'entropy' of a message, the term entropy is derived from a related measure used in thermodynamics, entropy is a measure of disorganization, an increase of information corresponds to a decrease of entropy, and vice versa; also efficient methods for 'coding' and 'error-correcting-codes' have been devised. The law of 'requisite variety' as introduced by Ashby (1952), has pointed out the various limitations that exist on the amount of control exerted by a system over a specific channel of communication.

Many technical findings in the science of communications have resulted in the widespread use of information-theory concepts in the modelling of learning. Although the engineering notions of 'communication' are used in a much wider and general sense. Parallels of the engineering oriented terminologies are drawn for the elements of natural information-processing systems, and many questions regarding the influence and the extent of interactive forces in such systems answered.

(g) - TURING MACHINES

The work of mathematician A.M. Turing during 1930's on the theories of 'computability' and 'computing-machines', gave rise to the notion of 'Turing Machines' and laid the foundations of 'Automata Theory'. Turing as a mathematician was interested in finding out a formal basis to deal with the solvability of problems; hence, the theory of computable functions were applied to an abstract simple construct called the 'Turing Machine'. Turing machines, in spite of having very basic components are capable of representing all computable functions.

The principal underlying concepts in Turing machines, are the ideas of 'effective procedures' or 'algorithms' for carrying out a given class of computations. An algorithm is a purely mechanical procedure, which starting from a particular initial data, will allow us to uniquely attain a defined goal by a step-by-step following of fixed rules. An algorithm need not necessarily be a terminating one, but it will in any case guarantee a result. The 'theory of algorithms' has introduced the concept of 'potential-realizability' or 'computability', which is not concerned with the 'limits' but only with the 'existence' of algorithms. Turing machines are capable of realizing all algorithms that can be completely specified by an ordered collection of logical states or machine-tables. An implication of Turing machine concept is that if a theory can be translated into a blueprint for a particular Turing machine, then such theory can also be realized in numerous hardware forms (e.g., physical, chemical, computer-simulation, etc.).

Although, Turing machines are very trivial in hardware sense and simple to construct (with some limitations), they possess no practical value, since they will be extremely slow and cumbersome in carrying out computations and solving problems. The principal applications of Turing machines have been in the theoretical domains of the 'theory of recursive functions' and the 'theory of computability'.

An important viewpoint that has emerged from the study of concepts of computability and the theory of machine is that: human and animal behaviour or mental activity can be achieved by machines, provided good enough descriptions of internal processes are available. The question of hardware is separated from the understanding of the behaviour of an entity; consequently, any natural phenomenon such as the learning-process could be duplicated by a variety of mechanisms and non-natural artifacts. Arguments developed from this standpoint, have generally involved two distinct approaches as far as the interactions of mathematics or logic and physiological processes are concerned; firstly, it is proposed that artificial machines can be constructed using natural observations; and secondly, it is contended that the analysis of artificial abstract machines can be beneficial to our understanding of natural organisms.

Turing (1950) in his analysis of the "fundamental problem of artificial intelligence" suggests a procedure for testing an organism for intelligence. Turing's test involves the interrogation, experimentation, and observation of the entity under question, and the comparisons of its performance with that of humans. However, at the conclusion of such tests, the question still remains whether the artifact only 'mimics' intelligence, or does actually 'possess' it.

In the most general and indirect sense an analogy exists between the logical states of a Turing machine and the mental states of humans, also between the structural states of a Turing machine and the physiological states of the brain; yet, a great deal of 'analysis' of human and animal activity is required before we can confidently embark on the 'synthesis' of such subjects in machine terms. Many unmistakably 'human' mental processes such as self-awareness or consciousness, point to an intuitive gap that exists between 'physical' and 'mental', or between 'machine' and 'animal'.

Our observations of many higher mental faculties suggest that activities such as problem-solving or thinking do not have a fixed algorithmic nature, and in many instances a 'heuristic' metaphor is more suitable, whereby the 'strategy' or the 'rule of thumb' gives the appropriate explanation. Whether the heuristic approach only differs from the algorithmic approach in degree, or if it truly represents the actual mental processes involved, is not clear yet, but seemingly, the animal's behaviour is the result of both algorithmic (deductive) and heuristic (inductive) type processes.

In the following, the main components of Turing machines and the different types Turing machines will be briefly outlined. Turing machine is an abstract concept, but the 'physical' characterization of its major components as shown in FIG.3.5 is comprised of: a control unit device, a read/write head, and an infinite length tape.

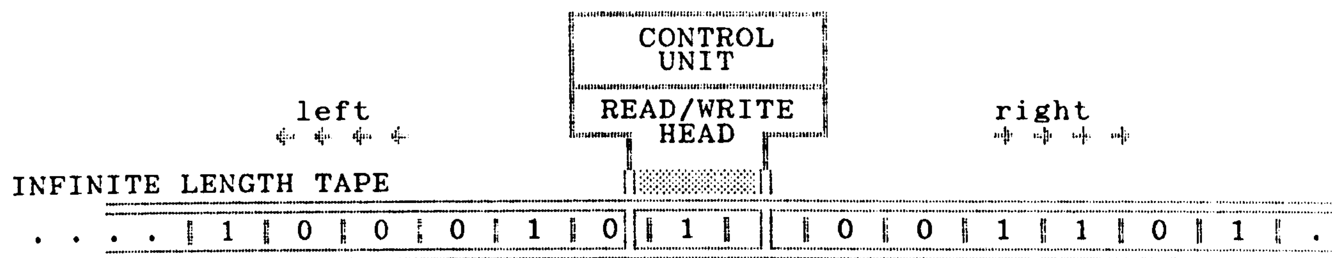


FIGURE 3.5. The basic components of a Turing machine.

The control unit can assume one of a finite set of symbols, it can move the tape a single discrete unit to left or right, it can 'halt' the machine, and also has a 'blank' symbol as one of the elements of its state-space. The read/write head scans the tape and transmits the information to and from the control unit to the tape.

The tape can be ideally infinite, but it is realistically thought of as finite but 'potentially infinite' or 'indefinitely extendable'; the tape itself is marked out into distinct squares lengthwise, and contains a blank or symbols from a finite set of symbols which can be read or written on one square at a time; the tape can be thought of as an external storage medium for information.

The behaviour of a Turing machine during successive discrete time steps can be determined by: the change in the internal-state of the control unit; the change in the scanned symbols of the tape; and the motion of the tape to left, right, or a stop. A complete Turing machine can be described by the three 'input', 'output' and 'internal-state' sets, together with the state and output mapping functions. Starting from an initial state, given the exact description of a Turing machine, a unique sequence of operations will take place. A symbol will be read from the tape; the state of the machine changed according to the state transition table; the output associated with this transition decoded by the control unit; and finally, a symbol may be written in the scanned square of the tape, or the tape moved along to left, right or stopped.

The 'algorithms' based on these simple operations can be represented by tabulated sets of symbols; the 'programs' of a Turing machine will hence define its actions for various state-symbol combinations; a 'computation' will denote a set of actions starting from a particular data point and ending with a halt. Turing was able to show that every algorithm could be represented by such a machine, furthermore, he was able to demonstrate that a so called 'Universal Turing Machine' (UTM) was able to perform any computation performed by any other machine, provided we had a description of its operations.

The behaviour of a UTM is determined by the program it reads from the tape rather than the specific computations prescribed by a fixed control unit. The fundamentals of modern digital computers can be described by universal Turing machines, and although the practical computers are not built as Turing machines, they can be simulated by the much simpler mechanisms of Turing machines, and many computational theories can be verified or discovered.

The original Turing machine has been modified into many alternate versions using different operating considerations, but in each case it has been shown that equivalent Turing machines could be devised which in mathematical terms have identical operational capabilities; however, despite the functional equivalence, the distinctive features of different Turing machines are of interest to cyberneticians. Some examples of these modified Turing machines are:-

(1) - One-Ended Tape Machines: the tape can only move in one direction.

(2) - Post-Davis Machines: cannot change both the symbol and move at the same step.

(3) - Paper-Tape Machines: a blank square can written on but only once.

(4) - Multi-Head Machines: more than one head per tape.

(5) - Multi-Tape Machines: more than one tape, each with its own head.

(6) - Multi-Dimensional Machines: tapes with more than a single dimension.

(7) - Two-Symbol or Wang Machines: having only two symbols of 0 (blank) and 1, but a large number of states.

(8) - Two-State Machines: having only two states (0 and 1), but the number of symbols may be large.

Shannon (1956) also made the important discovery that Turing machines with just two state or symbols can be constructed which are equivalent to any other Turing machine, these ideas being specially appealing to computer scientists because of the binary connotations of digital computers.

(h) - AUTOMATA THEORY

Many natural or artificial 'systems' can be defined by imposing arbitrary boundaries between a collection of entities and their environment. The concepts of 'Automata Theory' could be applied to the dynamics of any such system which has the three principal identifiable components of 'input', 'output' and 'internal-state'. 'Automata', generally refers to the abstract devices of finite size at any particular time, their behaviour is completely defined in terms of relations between their three basic elements (i.e., inputs, outputs, internal-state); whereby, the internal-state and output at any specific time is determined by the previous internal-state and input. Automata theory has developed within the science of mathematics as an independent abstract discipline; yet, many interpretations have been imposed on its concepts when applied to other scientific fields such as computer science, nervous-system networks, control-systems and biological or behavioural systems. Automata theory is used as a formal descriptive language to characterize the information processing and the behaviour of many automata.

Established behavioural and cognitive theories have been expressed by the conceptual framework of automata theory, these theories are seen to be logically identical to their automata theory representations. Adequate mathematical techniques such as 'logical nets' have evolved to represent the features of the external real world into the internal states of an automaton; many 'learning' neuron-like network models have been developed using the automata theory notions, some incorporate neuro-biological phenomena, others have a more abstract formal nature.

The problems of 'synthesis' and 'analysis' of systems represented by automata also arises. The analysis problem is to determine models and formulae that can efficiently represent the behaviour of a given automata, this can be achieved by devising:-

(1) - Mathematical descriptions: expressed by 'transition-functions', 'transition-tables', and 'transition-diagrams; examples of each are outlined in FIG.3.6.

$M = \langle I, O, S, G, F \rangle$

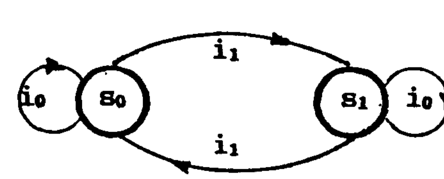
I is the input set.
O is the output set.
S is the internal-stat set.
$G : S \times I \longrightarrow S$
is the next state function.
$F : S \times I \longrightarrow O$
is the next output function.

(i)

(S)

(I) G	s_0	s_1
i_0	s_0	s_1
i_1	s_1	s_0

(ii)



(iii)

FIGURE 3.6. The representation of the behaviour of an automaton by three common methods:-
(i) State-Transition-Function,
(ii) State-Transition-Table,
(iii) State-Transition-Diagram.

(2) - Physical descriptions: such as circuit diagrams, indicating how the automata could be built using basic constructable elements.

(3) - Formal language descriptions: such as 'regular events' and 'regular expressions', which are notions devised primarily for the investigation of the nature of classes of realizable events, or the limitations of an automaton in recognizing classes of sequences.

The synthesis of an automaton is generally the more important problem; given a specific formulation for the behaviour of an automaton, the task is to design such an automaton if possible. In some cases these problems are not algorithmically solvable and 'heuristic' methods must be used to abstract the desired automaton, and then 'decision-theory' methods employed to evaluate how closely the end result satisfies our criteria.

An important question which has been tackled by automata theorists, such as Kleene (1956), is to identify the simplest forms of any given automaton. Kleene has shown that most natural or artificial events can be characterized by an equivalent automaton using nets of simple logic elements of 'and', 'or','not', and 'delay' or 'memory'.

Automata have been categorized into many different classes. 'Growth' automata as opposed to 'fixed' are those which can get arbitrarily large in size, they can be considered to be potentially infinite; Turing machines are interpreted as growth automata. 'Partial-growth-automata' or 'growing' automata, can grow but have a limit to their growth, this type of automata are effectively identical to the fixed ones in their computational capabilities.

'Discrete' automata are those acting over specified time intervals, their inputs, outputs and states are only considered at those instantaneous descriptive moments; while, 'continuous' automata are considered during an entire span of time interval; 'analogue-computers' represent a kind of continuous automata. However, it is established that most continuous systems can be effectively approximated by an equivalent discrete systems. Another distinction is made between 'synchronous' and 'non-synchronous' automata, the former having elements which act in unison.

A 'deterministic' automaton is one in which the present state and output are the inevitable consequence of previous state-input pair, the behaviour of this type of automata is precisely defined by its transition functions. A more random device, where each state-input pair could lead to more than one state-output pair, is called a 'non-deterministic' automaton. A class of non-deterministic automata which has been studied rigorously is the 'probabilistic' automata, these automata have specific probabilities attached to each alternate transition.

In principle, the probabilistic machines are capable of no more than the deterministic ones, but can often do things much more economically. Probabilistic or stochastic automata, have been formulated as models for systems that have unreliable components effected by various 'noise' factors, they have been used for the modelling of adaptive 'learning' systems, and also in aiding the simulation of a psychological phenomenon such as learning.

'Finite Automata', are a category of automata used widely in the modelling of real objects, they are assumed to be fixed, synchronous, discrete, deterministic, finite-input, finite-output, and finite-state automata. A universal Turing machine embodies a finite automaton, but having a potentially infinite memory, is not a practical concept for the modelling of physical entities such as digital computers or other artificial and natural systems.

Finite automata can be considered to be Turing type tape machines, with the following provisions: (a) - they have two separate input and output finite tapes both moving one square in one direction at each interval of time; (b) - only the symbols on the output tape may be changed. Other intermediate forms of tape machines have been designed, having applications in more complex computational situations, these tape machines can have many different tape or read/write-head specifications.

Although, finite automata are much simpler than Turing machines in construct, computationally they are not as powerful, this is because they do not possess the external memory storage facilities of Turing machines. Many systems with finite constitution of discrete elements, have been defined using the notions of finite automata. Of special interest to the modelers of learning behaviour and mechanism, are the 'neural network' representations as finite automata. The theories of logical nets, based on the mathematical treatments of systems composed of logic elements, have been applied widely to the biologically oriented neural network models.

The most extensive use of logical nets in the modelling of nerve systems can be seen in the 'modular' type networks introduced in 1943 by McCulloch and Pitts (1947). Their concepts involved the fusion of ideas from Boolean algebra with networks of idealised neurons. A feature of this type of neural network was the ease of their construction in hardware using sequential electronic switching circuits, implicating the possibility of their use in digital computers.

Neural network models were discussed earlier in this chapter, but their significance to modelling of learning is in their use of mathematical concepts and formulae in representing aspects of behaviour of a natural entity.

The original 'idealised' neuron models of McCulloch and Pitts, depicted the actions of a nerve-cell at a very simple level. A typical configuration for such nerve model is shown in FIG.3.7; the neuro-physiological notions of 'inhibitory' (I) and 'excitatory' (E) input, 'refractory-period', and 'threshold' value, were incorporated in these models. The idealised neuron is activated (fired) according to various criteria based on: the summation of excitatory and inhibitory inputs, the total number of excitatory inputs, or the majority of excitatory and inhibitory inputs. Also, in later developments the delay element was embodied in neural models, hence giving a memory or storage facility.

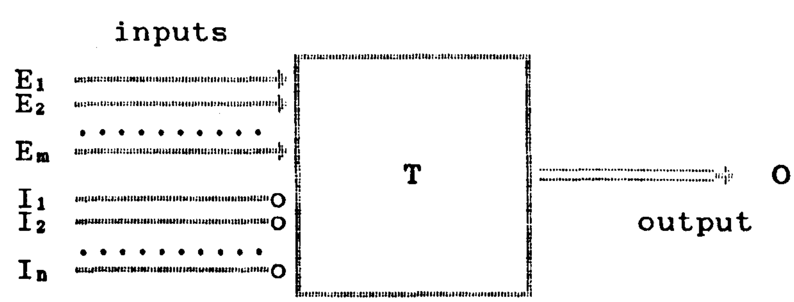


FIGURE 3.7. A representation of McCulloch and Pitts type neuron model, (E_i) are the excitatory inputs, (I_i) are the inhibitory inputs, (T) is the threshold value for the firing of the neuron, and (O) is the output.

The concepts of 'threshold-logic' and other mathematical theories were used to devise more precise models of neurons and networks of such neurons, both as ways of simulating physiological processes and also as tools for fresh lines of research. Effective procedures had been designed so that networks similar to any given behavioural pattern could be described.

These ideas have found many domains of application in 'learning' systems, in particular, in the field of pattern-recognition. The adaptive modifications of nerve-network automata have been investigated, both as the structural changes of hardware and also as the programmatic changes of behaviour. The hierarchical cognitive processes are also a convenient and widely used subject for modelling. Many aspects of perception, memory, motivation, generalization, classification of inputs and outputs, and even emotion have been modelled by neural-networks. Examples are the automata designed by George (1961), Stewart (1967), or Culbertson (1963) which display some features of cognitive classification of input data.

A different branch of automata-theory is the study of 'cellular' type automata, in finite or infinite form. This is the relatively more recent area of automata-theory, which is concerned with the behaviour of systems made up of regular arrays of identical interconnecting elements. The theoretical subjects covered in this field include the study of the computational powers and the limitations of such automata; and the applications include the design of computers and cellular type machines.

The behaviour of a simple cellular automaton can be investigated by the examination of the activity of networks made up of small number of elements that follow a set of specific transitional rules. Different initial configurations of states, normally, result in one of the following patterns of behaviour:-

(a) - the activity dies out after a while,
(b) - the activity reaches a stable state,
(c) - a repeating pattern of behaviour emerges (oscillations),
(d) - the behaviour seemingly changes indefinitely.

Von-Neumann (1956) was interested in the behaviour of cellular automata that had elements which did not perform ideally, and had a small probability of malfunction; he was basically concerned with the role of error in logic and in automaton. The main problem was to find the probability of the malfunction of large networks which were composed of cells, each with a small probability of breakdown. Von Neumann showed that if the individual probabilities of the cells were small, then the total probability for malfunction need not be very large. These considerations are part of the so called 'reliability' studies of the systems of inaccurate cells; an obvious choice for this type of automaton was the first generation valve computer. A basic method for improving the reliability is to incorporate 'redundancy' in the system; there are two distinct ways of doing this: (a) - having redundant components that act continuously, and (b) - having redundant components that are kept as spares. Von Neumann used some of the parallel features of the brain and nervous systems, which are able to function in a highly reliable fashion, in the design of nerve-networks. More reliable cellular automata were devised using elements with multiple inputs; later other methods such as 'quaded-logic' and various techniques for the testing of reliability of automata were also introduced.

The idea of a universal cellular type machine similar to the UTM was also proposed by von Neumann, this automaton, as well as being computationally and constructionally universal, was also able to 'self-reproduce'. Other complex cellular automata such as Arbib's machine (1969) have been devised.

3.6.3 SOME UNDERLYING PRINCIPLES INVOLVED IN 'LEARNING' SYSTEMS

The study of adaptive 'learning' systems within analytical subjects, such as cybernetics, has revealed or resulted in the redefinition of some basic underlying principles of adaptive and learning processes. In this section we will discuss the nature of the criteria that are, normally, deemed to govern or trigger the adaptive modifications of systems. The main concern will be the forces which seem to activate the learning process, or the holistic interpretations that can be imposed on such processes. The specific components of mechanisms involved, such as the memory or the reward/punishment centres, will not be elaborated here.

Some of the attributes to be discussed in this section have equivalent psycho-physiological counterparts, while others are only by-products of non-biological abstract theories. Another point which should be emphasised, is that there is no unanimity amongst various researchers regarding the precise meaning of some terminologies, and many different interpretations of the same phenomenon exists, depending on the approach and the background of the individual worker.

(a) - GOAL-DIRECTEDNESS

Notions of 'goal', 'intention', 'purpose', 'drive', and 'motivation' have been discussed earlier in their biological or psychological contexts. Thorpe (1963) described the two concepts of 'directiveness' and 'purposiveness' as the central features of all adaptive behaviours. Similarly, Tolman's (1959) learning theories emphasised the importance of the ideas of 'purposiveness' and 'goal-directedness' in learning.

Many animal activities are clearly goal-directed, hence it is a prerequisite of any adaptive or 'learning' system that purports to be animal like, to also be goal-directed - by displaying a selective or purposive behaviour. A dilemma is anticipated here, since the intuitive notion of 'animal' purposiveness is extended to the domain of 'inanimate' objects. The absence of intrinsic intentionality in a machine or abstract system means that when we say an artifact is goal-seeking, we are speaking in purely subjective terms, there is no comprehension of the notion of 'goal' - and in the case of simple goal-directed automata, such as the physical equilibrium-seeking systems, not even a memory of the goal event.

Goals and sub-goals are defined for adaptive systems either by having specifically defined target goals to be reached or maintained, or by a set of imprecise general improvement type goals; great difficulty is encountered in conveying the meaning and the rationale of achievements (e.g., 'winning' in a chess game) in objective terms.

Even the highly adaptive artificial 'learning' systems seem to only mimic the goal-directive behaviour by following a predetermined set of error-reducing actions stipulated by a hierarchy of goals - where the sense of achievement is normally determined by 'hedonic' principles of reward and punishment. The problem is further complicated since we can also look at a natural goal-seeking behaviour in 'drive-reduction' terms, where an animal seeking food can be equally construed as trying to reduce its hunger. This duality of explanation is equivalent to the dilemma of interpreting goal-directedness as error-reduction in artificial systems.

Despite such epistemological and other philosophical objections, the property of goal-directedness is seen by many workers as a universal phenomenon, applicable to all types of adaptations. The structural interconnections of a system is what determines the goal-directed behaviour; hence, goal-directedness is seen as a system phenomenon; a temporary swaying of behaviour from the obvious features of directiveness is allowed, provided the total behaviour is pointed towards a goal.

'Teleological' systems are those that show a goal-directed behaviour, and appear to have a purpose. The problem has been to formalise in mathematical terms the behaviour of such systems. The difficulties involved in the precise definition of purposive type behaviour has also resulted in many controversies in their analytical characterization.

Rosenblueth, Wiener and Biggelow (1943) distinguished clearly the structural studies and the behavioural studies of systems, they believed that goal-directedness was a behavioural property of systems that dynamically interacted with their environment to maintain certain constancies, the behaviour was directed by the goals it was trying to seek or maintain; while the simple goal events could easily be observed in both biological and inanimate systems, the more complex hierarchical goals were difficult to manifest in artificial systems.

Ashby (1952) proposed an objective criterion for the explanation of teleological behaviour; he defined goals in terms of the objectives of an equilibrium-seeking system. Some examples for such goal-directed behaviour were described as: the behaviour of stable systems around their state of equilibrium, and other regulatory type systems such as servo-systems. These and other biological systems that steered toward a goal under various external disturbances were thought to be equivalent in principle; the natural phenomenon of 'homeostasis' was depicted in his stability-seeking automaton (Homeostat), which appeared to behave purposively.

Sommerhoff (1969,1974) has analyzed the functional relationships between the variables of teleological systems, and has given a formal definition to what is meant by purposiveness in biology and psychology. Additionally, the two notions of 'subjective-purposiveness' and 'objective-goal-directedness' are distinguished (he calls the latter "directive-correlation"); also, by the introduction of the condition of 'orthogonality' (acting independently), has excluded the thermodynamic equilibrium-seeking systems from the class of living teleological systems. However, artificial adaptive systems, such as the servo-systems are considered to be in the same class as the natural goal-directed systems, provided they are considered in their broad context of application.

Sommerhoff envisages the difficulty of applying his state-determined approach to the more complex hierarchical goals of natural systems, and devises an integration scheme for goals and sub-goals. He also emphasises the goal-directedness rather than the goal-achievement aspects in his work, and mainly deals with behaviour rather than its means or ends.

Another argument for distinguishing the thermodynamic equilibrium-seeking systems from other teleological systems was put forward from the information-theory sciences, the former is seen to be closed to information flow, while the latter is considered to be open to transfer of information from its environment.

(ii) - SOME OTHER PROPERTIES OF TELEOLOGICAL SYSTEMS

While goal-directedness explains the nature of the behaviour of a teleological system at executive level, it does not account for the reasons behind the appearance of such activity in organisms, in other words, it does not tell us why such a pattern of behaviour should come about. This question, of vital importance to the survival of all organisms, has brought about varied and sometimes contradictory explanations.

The widespread opinion is that an organism is 'seeking-stability', hence in Ashby's terms, is trying to keep the essential variables within acceptable limits. The seeking of uniformity, stability, or equilibrium in biological life is an intuitive notion; however, some interesting observations such as the behaviour of animals that are apparently satiated, or experiments on human subjects that have had all their sensory inputs blocked, show that such 'stable' situations are not ultimately desireable, and a need for 'change' is possibly an equally powerful criterion of life. Therefore, some explanations have been devised based on 'instability-seeking' or 'change-seeking' concepts. Another argument is that the living organism seeks a 'concise' or 'economical' representation of its internal information, from which the successful features are extracted.

The degree of organization within an entity is also sometimes seen as the desireable cause for the initiation of goal-directed behaviour; such systems are thought to be 'negative-entropy-seeking', since the entropy (of communication theory) is a measure of uncertainty in systems, and the lower the entropy value the more organised the system is, hence the system must have a negative rate of entropy change to show purposiveness towards more organised states, rather than move towards randomness and chaos.

Finally, more recently the concept of 'autopoiesis' has been introduced by Maturana and Varela (1974), to characterise some of the most fundamental properties of living organisms. This abstract new insight into the organization of living animals refers to the capacity of the living organism to develop and maintain or more loosely put to 'self-copy' their identical organizational properties. The emphasis is on 'circularity' in the living organism; the holistic unitary view of the organism is adopted, hence the interest is in the interactions or the relations of the components of complex networks, rather than the analysis of individual components (such as reproduction or adaptive centers). Also, a clear distinction is made between the 'structure' and the essential 'organization' of a system. Autopoiesis is seen as a property of the organization of systems with 'internal variables' which can give rise to other similar type systems; all living organisms are considered to belong to such class of systems.

Cellular automata have been constructed as models to display this primary function of life. Autopoetic-systems also have a strong bearing on 'self-organising' studies.

A paradox emerges when attempts are made to abstract definitions or criteria for living organisms; once processes such as directive-correlation or autopoesis are defined to a certain degree of precision, essentially 'artificial' models or machines could be constructed to display such properties. Therefore, to have a consistent definition for the distinguishing features of living from the non-living, we must devise the concepts in terms of some biological functions, hence excluding all artificially constructable systems.

(iii) - SELF-ORGANISING SYSTEMS

The notion of 'self-organization' is also an intuitive concept, introduced principally as the result of observations based on the animal nervous-systems which seemed to possess similar properties to self-organization. This process is considered as an important aspect of goal-seeking systems. The peak of interest in self-organising systems and machines which were closely associated with learning and adaptive systems was during 1950's and early 1960's.

Self-organising systems, according to Glorioso (1975), can be defined as: "adaptive or learning systems in which the initial states are unknown, random, or unimportant;" they are able to change their internal states and hence the reactions to specific stimuli; they should also be able to act autonomously, whereby no external adaptive element (e.g., a teacher) should be involved, this type of 'self-modification' can be achieved in principle by incorporating hedonic centres in the system.

The question of control is also an important consideration, which is to determine how to keep the system within viable regions of environment, for this reason most self-organising systems have two 'operational' and 'learning' components.

The process of self-organization according to Ashby can come about in variety of ways; one possibility is to have independently acting separated parts, which become joined and organised, an example of this type is the process of nerve cell growth in embryo; another method envisaged is to have a system of loosely connected parts which become better organised, this method is more akin to the process of learning and adaptation.

One of the major problems with self-organising systems is their subjective and observer-dependent nature. A self-organising system does not mean much unless defined in relation to its environment, it can only display organization if there is an external source of order in environment. The law of 'requisite variety' postulated by Ashby (1956), describes this aspect of the interaction of system and environment, by stating "only variety can destroy variety."

Various abstract theories and formalization of self-organising concepts have been developed by cyberneticians, using information theory notions such as redundancy or entropy, or feedback control concepts. Models of different classes of self-organising automata ranging from the networks of neuron-like information processing elements to more complex economic systems have been constructed, most such models show a degree of self-organization, but on the whole do not display any efficient learning or adaptation. Mainly because of this reason, during the past two decades, the work in mainstream artificial-intelligence research has not involved a great deal of self-organising concepts; and has generally depended on the heuristic techniques which rely on an external teacher or instructor to direct the organization of a system.

(iv) - SELF-REPRODUCING SYSTEMS

The process of 'self-reproduction' is indisputably one of the most fundamental properties of living organisms, for it can be seen even at the lowest levels of life, such as cells, viruses and molecules. Self-reproduction involves the processing of matter and information, but our main interest in here is from the cybernetic organizational point of view.

Systems have been devised to satisfy the basic criteria of self-production,; at the lowest level, some simple mechanical automata have devised which depict the 'self-replicating' properties of some physical processes such as the formation of crystals, for example, the self-reproducing mechanical configuration of Penrose (1959), although, possibly the low level of information in this type of structure does not warrant their definition as true self-reproducing, as only the identical copies of the original are made.

The pioneers of cybernetics considered the process of self-reproduction, in both man and machine, as part of the vital processes of 'self-maintenance' and 'self-regulation', which in their totality determines the goal-directedness in the organism. The manifestation of this process in machines is not seen as the machine-tool type operations that can produce other machines, but the reproduction of similar structural and organizational intricacies.

The automata of today are only in their primeval evolutionary stage of development, and it can be envisaged that they will be able to reproduce like artifacts in future. Already some computers are used to design new generations of computers.

Von-Neumann (1951) had also proposed two approaches for the analysis of the problem of self-reproduction, a "kinematic" (mechanical) and a "logical" approach. His work having a strong biological flavour was based on the McCulloch and Pitts type networks of idealised neurons. He showed the theoretical possibility of constructing machines that could self-reproduce other machines as copies of itself or other similar machines. He abstracted a two dimensional cellular automaton which constituted an originally inert lattice for the analytical machine, such an environment would provide the 'matter' for the machine to self-reproduce; three types of element were incorporated in the automaton to perform 'logical-control', 'transmission', and 'muscular' functions in the abstract sense. The notion of a genetic-blueprint was utilised within the machine in the form of a 'tail' (borrowed from genetics terminology) which carried the information about what it is to build, this tail would always be copied in the new machines hence giving the off-springs also the ability to be self-reproducing.

This type of automata being a special class of growing-automata, have been studied and investigated by other researchers; 'computational-organs' and other evolutionary or ontogenetic aspects of natural self-reproduction have been depicted in the abstract form.

(v) - SELF-REGULATING SYSTEMS

The notions of 'stability', 'regulation', 'equilibrium', and 'homeostasis' were discussed earlier in this chapter. Some mental, emotional and biological processes can be seen to be clearly 'self-regulatory', the body as a whole can be considered to be also a self-regulating system. On a much larger scale, some theorists have applied the concept of self-regulation to the whole functioning of the plant earth - the 'Gaia' hypothesis. Some physiological organs such as the 'hypothalamus' have been associated with biological regulatory processes; similarly, in psychology, self-regulation is observed in processes involving various goal-setting, motivational and reinforcing elements.

In general system terms, the notion of regulation is intricately related to control, and various aspects of the amount and the degree of regulation can be defined by the use of Ashby's (1956) 'Law of Requisite Variety'. The principal component of a self-regulating system is its 'feed-back' loop, and the main feature of the design of this class of systems is their ability to internally set their own goals and objectives, rather than have such criteria determined by external interventions.

PAGINATION
ERROR

CHAPTER 4

APPROACHES TO MODELLING OF LEARNING – PART-I

4.1 INTRODUCTION

'Learning Systems' in some literature exclusively refer to the formal manifestations of learning in control systems, or to the general systems-theory view of the realization of the learning process in systems. But, in this chapter, 'learning systems' will refer to all physical artifacts, abstract theories, or computer programs which can be defined independently from their environment, and can display some kind of 'learning' or 'adaptive' qualities. A broad definition of learning in systems according to H. Simon (1983) is: "Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and effectively the next time."

All four aspects of 'simulation', 'analysis', 'synthesis', and 'black-box problem' can be seen in research involving learning-system modelling. But, the main issue is the design of systems which use information obtained during one interaction with environment to improve performance during a future interaction; the basic criterion used could be some form of generalization, averaging, or selection of 'ideal' (typical) elements of the set of past records. Andreae (1977) specifies three principal qualities for learning (teachable) systems, they are: 'potential', 'aptitude', and 'resilience'; respectively referring to: the highest levels of task complexity achieved, the speed and ease of learning, and the ability to deal with new knowledge or 'noisy' information.

The terms 'learning', 'adaptation', 'concept-formation', 'induction', 'regulation', 'self-modification', 'self-organization', and 'self-repair' have all been used in the context of 'learning system' research. The goals of learning-system scientists are to construct machines or systems that can perform such processes. Many workers in subjects such as Cybernetics, Robotics, Artificial-Intelligence, Computer-Sciences, Pattern-Recognition, Adaptive-Control-Theory, Automata-Theory, and Cognitive-Psychology have been intensively involved in pursuit of these goals.

There are basically three outlooks in learning-system sciences: 'task-oriented studies' (or engineering approach), 'cognitive-simulation', and 'theoretical analysis', each involving structural (knowledge) based and/or behavioural (skill) based problems. Although, workers in this field initially aim towards one of the above three distinct objectives, their research often permeates into different areas and, in turn, creates new lines of investigation. The majority of the significant achievements have involved the 'task-specific' type research; work implicating 'general' principles of learning, such as the finding of efficient global adaptive methods and theories, has been progressing at a much slower pace.

Many taxonomies are possible in the field of 'learning systems'. A classification based on the underlying strategies or inferences can be made, ranging from the simple rote-learning of programmed learning to the complex learning involved in discovery and observation; the relative role of the 'teacher' is of prime concern here. Other classifications are also possible, based on: the methods of representation of systems (e.g., hardware, computer program, abstract, etc.); the types of acquired knowledge or skills (e.g., parameters in mathematical expressions, classifications, formal or grammatical rules, etc.); the domains of applications (e.g., robotics, education, natural-language-processing, etc.); or the forms of alterations involved in systems (e.g., parameter estimation, structural changes, changes in assertions, etc.). Another important consideration, which can be used as a basis for classification, is the developmental feature of a 'learning system', whereby, the systems that interact with their environment, and go through a specific training phase, are distinguished from those that continuously interact, and whose performance and training stages are not sharply differentiated.

In this chapter we will attempt to outline a number of examples of 'learning systems', and discuss their characteristics. The particular taxonomy used will be based on the branches of science involved in this area, each class is characterised by a commonality of approach in dealing with the modelling problems of learning-systems. Most researchers in such fields, while ultimately seek to compare and relate the results of their work to natural learning processes, do not, usually, attempt to replicate the mechanics of natural learning phenomena; and generally use alternate artificial means for manifesting learning and adaptation in systems. The distinctions made here are by no means exhaustive or clear cut, but simply represent the research fields that share a common methodology. Many overlaps and similarities, as well as differing points of view, may arise; furthermore, research under different headings could in fact be referring to the same concept, and conversely, results which satisfy the requirements of one discipline may not be suitable for application in a different discipline.

4.2 ADAPTIVE AND 'LEARNING' CONTROL SYSTEMS APPROACH

The scientific use of terms 'adaptation' and 'adaptive' originate from within the sciences of biology and psychology - referring to 'success' in a changing environment. However, these terms have taken special meanings in different disciplines.

Wiener (1948) defined learning in a broad all-encompassing form, to include both 'ontogenetic' and 'phylogenetic' adaptations; while, on the other hand, Shannon (1953) gave a more selective definition for the term learning, based on a time dependent measure of 'success' or 'adaptation' to the environment, hence restricting the attention to the ontogenetic aspects of learning. Andrew (1967) distinguishes learning as the more elaborate manifestation of self-improvement than adaptation. In formal control systems, generally, learning is seen as a special form of adaptation which is related to time, and hence its effectiveness is of issue.

An engineer, normally, equates learning and adaptation and often uses these terms interchangeably; but, a psychologist makes a clear distinction; and a biologist has an in between view point. In engineering a 'machine' is considered to be adaptive if it has one or both properties of 'stability' and 'reliability', in other words, is able to remain within the constraints of prescribed bounds and/or is able to repair its actual failed machine parts. In such disciplines 'adaptiveness' is seen as a property of a particular class of system, whose desired (almost predetermined) 'adaptive-states' can be reached or maintained by a series of estimations or error-reducing algorithmic techniques, less emphasis is given to the further improvement of the system beyond its initially specified goals - the main issue is the attaining of stable and reliable performance. On the contrary, in natural sciences the developmental view is adopted; adaptation is seen as the manifestation of a constantly improving system which interacts with its environment, with the end result being largely unknown.

The basic elements and relationships of a learning or adaptive system and its environment, as seen by the engineering control scientists, are outlined in FIG.4.1. The system is deemed to be 'adaptive' if the 'critic's' output remains within prescribed bounds as the 'environment' and/or the 'machine' change (a simple servomechanism is a trivial example of such systems). Similarly, the system is said to be 'learning' (in engineering terms) if the critic's efficiency is increased during a specified period of time following a change in its environment. 'Self-repair' type systems are also defined as those which can improve their performance after a failure or change in some internal components.

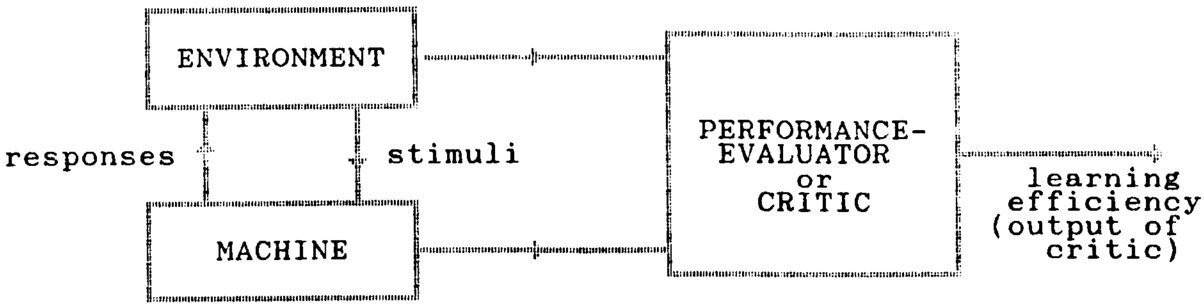


FIGURE 4.1. Schematic representation of an engineering 'learning system'.

'Feedback' is seen as an important and convenient way of manifesting adaptation, but is not a necessary or sufficient precondition for all adaptive behaviour (i.e., a well programmed machine with no feedback loop may appear to behave adaptively); although most adaptive control systems are 'closed-loop' and use feedback, in particular, those systems whose inputs do not reflect the important environmental or structural fluctuations. Here, a simple hierarchical classification of systems in the ascending order of machine-environment interaction complexity could be considered: 'random systems', 'directive open-loop systems', 'adaptive closed-loop systems', and 'learning closed-loop systems'.

The notion of 'self-organization', and its manifestation by the 'self-organizing control systems', has also been frequently touched upon in this area of research, normally, referring to and emphasising the internal structural changeability of an adaptive system.

Adaptive or 'Learning' Control Theories are an extension of 'Automatic Control Theories', and a significant amount of research has been carried out in this subject since the middle 1950's. Application areas have included the control of aerospace and industrial processes as well as man-machine and socio-economic systems.

The goal of the designers of adaptive or 'learning' control systems is to construct effective mathematical models (in the more advance systems, normally, formulated in stochastic and non-linear terms), from which they are able to find or learn the values of certain parameters which optimise some criterion in the system.

In deterministic or stochastic control systems with known structural and/or behavioural (input/output) details, if the exact nature of performance criterion and its relation to different system elements were known, then the various standard analytical methods, such as error-control techniques or other formalizations provided by system theory, can be used to determine the appropriate optimal behaviour of a system, and keep its performance within satisfied constraints - the control and modification of the system would be done on the bases of its known internal structure and external behaviour. However, in the case of unknown, or incompletely known, criteria or parameters of a system, the techniques of adaptive or 'learning' control are utilised to reach or maintain the optimal performance; since, the observable changing parameters do not provide sufficient data to enable us using the existing optimising techniques of control theory. If the living organisms are looked at in this vain, as Ashby (1956), then we can say that "learning" is evolved to tackle the adaptive control problems of life.

An adaptive control system can, usually, be functionally decomposed into a quickly changing component, loosely termed as 'the state' of the system, plus a slowly changing component which includes the adaptation elements. This distinction is vaguely analogous to the bimodal representation of the experiences of an organism, by the 'short-term' and 'long-term' components of its memory, as specified in the behavioural sciences. More formally, adaptive control systems are characterised by its two major components: a 'plant' which is to be controlled, and a 'controller'. This is shown in FIG.4.2.

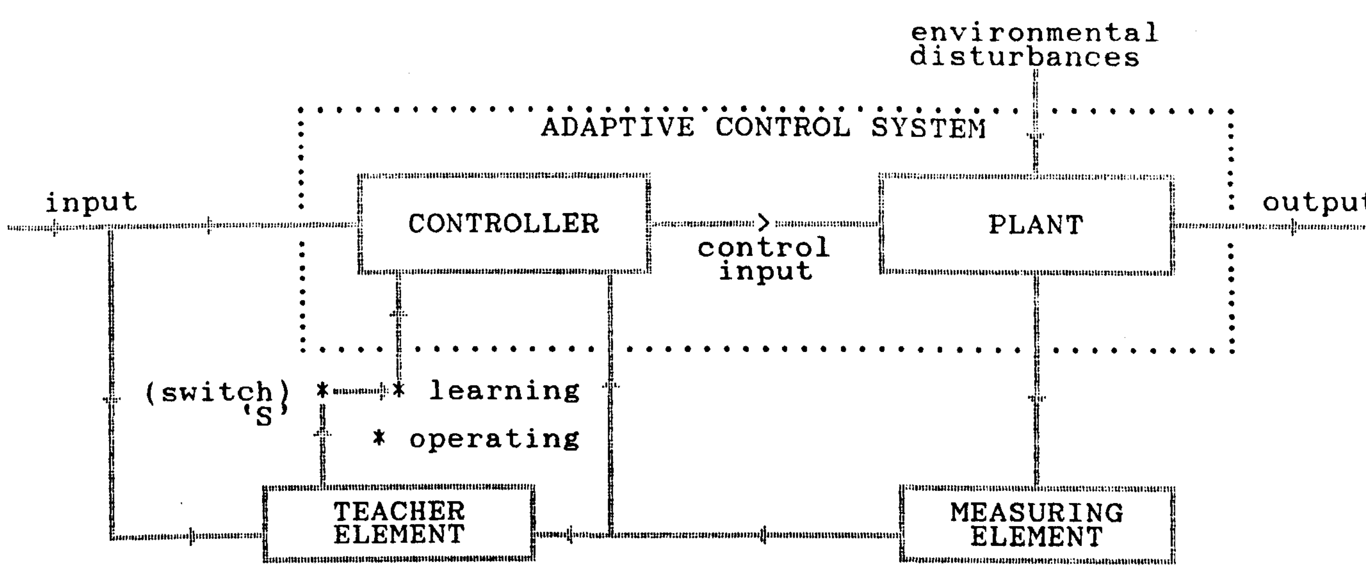


FIGURE 4.2. A general schematic functional diagram of an adaptive or 'learning' control system.

4.2.1 ADAPTIVE CONTROLLERS

An adaptive controller functions by constantly monitoring the performance of a system relative to some desired performance criterion, and subsequently making modifications, by closed-loop action, to the system structure and/or behaviour, so as to approach such a criterion.

There are two principal types of adaptive controllers: 'passive' and 'active'. The designers of passive controllers only use real-time information, that is to say, they use parameters which are already available at the outset of an experiment - the unknown information is either ignored or given a tentative value. On the other hand, the designers of active controllers incorporate the possible future availability of further information into their systems; hence, such systems, as well as real-time data, can estimate other new unknown parameters during their operation; this type of controllers are based on the so called 'dual control theory'.

Passive adaptive controllers, in general, are the simpler yet the more inefficient and sub-optimal controllers, however, they have been investigated much more extensively. Active adaptive controllers are able to make estimates of the values or the forms of various parameters, and gradually improve the system's performance. Active adaptive controllers should ideally approach an optimal controller, the final performance of which should be on a par with the performance of a controller designed with all the required information known in advance. Due to the similarity of this second type of adaptive control action to the process of natural learning (as far as their performance improvement characteristics are concerned), this class of adaptive systems are more readily referred to as 'learning control systems'.

4.2.2 ADAPTIVE CONTROL PROCEDURES

Referring back to FIG.4.2, if the exact 'input' to 'output' transformation functions of a system were known, then, as mentioned earlier, the control theory methods could determine the course of the optimal behaviour of the system. However, in some cases there are very rapid or large variations of certain system parameters under environmental or internal influences; or alternately, we do not know enough about the details of the system or its operation; in either case, the techniques of adaptive control, by means of successive approximation or estimation, are able to predict the unknown parameters or functions, and achieve the desired behaviour of the system. The brain is an example of such a system, whereby, it has to cope with the changes in its structure through growth and decay, and also make sense of a variety of changing or novel perceptual experiences.

The 'input' in FIG.4.2 refers to some 'desired' or 'reference' conditions, and is normally not known a priori. The adaptive control system functions by choosing a tentative input which generates an output from the plant; the 'measuring element' directly or indirectly measures some aspects of the output and applies it to the 'teacher element'; the 'teacher' or the 'evaluation element' compares such measurements with the input of the system, and makes adjustments to the state variables of the controller. The system's overall performance gradually improves by the updating of the controller, and also the revising of control and performance criteria.

4.2.3 THE TEACHER ELEMENT

The 'teacher' or 'evaluation' element in an adaptive or 'learning' control system has one of two forms, either it is 'external' (i.e., supervised learning, off-line learning, or training) or it is 'internal' (i.e., non-supervised learning, on-line learning, or self-evaluation); in both cases, the function of the teacher element is to oversee the overall performance, and evaluate and direct the process of learning or adaptation. In the externally supervised systems the exact desired answers are normally known in advance, but, in the non-supervised systems, where the 'teacher element' is an integral part of the adaptive controller, the optimal answers are gradually approached by using some built-in performance criterion. It must also be noted here that the external teacher, on the whole, will be largely ignorant of the workings of the system, although it might be able to predict some aspects of the behaviour of the system.

The 'teacher element' is universally observed in all types of learning and adaptive systems - both natural and artificial. In nature, there are abundant examples of 'learning systems' which distinctively have non-supervised characteristics; but, in artificial systems such a categorization is not so clear-cut, since, the majority of adaptive or 'learning systems' which seemingly operate in autonomous and self-learning mode have their performance criteria or goals specified by their designers or external teachers prior to their operation. Although, some artificial systems are genuinely capable of setting their own independent goals and sub-goals during their operation.

It can be argued that no precise classification of internal vs. external 'teacher' can be made, since in both natural and artificial systems it is virtually impossible to isolate the internal teaching or evaluation element as an independent functioning entity - either the external events directly or indirectly influence the evaluation criterion or the initial design concepts (inborn genetic information in the case of animals), to some extent, guide the future goal setting aspects of the system. However, for practical purposes such a distinction is made, specifying if during the course of the operation of an adaptive system an external supervisor is involved or not. The external supervision, normally, resulting in the relatively quicker rate of learning. Some adaptive control systems are designed with both types of teacher element.

The switch 'S' incorporated in FIG.4.2 could be used to run the system in two phases of 'training' and 'operating'. Once the system has been trained to an acceptable level of performance and modified accordingly, then the system can be switched to its operating mode, further training becomes only necessary if the performance criterion (the same or a different one) is no longer satisfied.

4.2.4 IDENTIFICATION, DECISION, AND MODIFICATION

The adaptive control law, as defined by most authors in this field (e.g., Mendel and McLaren, 1970), has three major functional elements of: 'identification', 'decision', and 'modification'. The process of identification characterises the various constituents of a plant or environment. The decision element determines how and which aspects of the system's actual performance should be related to the desired performance conditions. And finally, the modification element changes the system parameters in accordance with the findings of the identification and decision processes - by updating the system towards the optimal performance setting.

The 'identification problem', as posed by Arbib (1972), is: "To use repeated experiments upon the input-output behaviour of a system to build up a state-variable description for a system which yields similar behaviour." This problem which involves the measuring and the estimating of system's significant features is of crucial importance to control scientists. In cognitive terms, identification is loosely synonymous with the process of 'feature extraction', an important aspect of the problem of 'pattern-recognition'. Identification procedures could be used in two different situations: (a) - to identify or recognise some static parameters, and (b) - to identify or estimate the time-varying parameters of a dynamic process.

A further classification of identification procedures can be made for the 'parameter adaptive control systems', which is: 'explicit identification' against 'implicit identification'. Both procedures are used prominently in the 'passive' type adaptive control systems; the explicit techniques, although less general, are usually simpler and have the advantage of being able to adapt more rapidly to changes in environment.

The explicit, or sometimes called 'indirect', identification schemes are based on the observation of the behaviour of the plant, and allow the updating of the systems's state equations. This approach was developed to utilise the existing control techniques which required exact plant descriptions. The alternate method, which avoids the specific design of individual plant's controller, is the implicit or 'direct' identification. In this case, a general purpose controller is utilised which can accommodate the degree of the complexity of the particular system; the control parameters themselves are adjusted to improve the performance of the system without actually determining the exact parameters of the plant. One way of implementing this second type of identification procedure is in the 'model reference control systems', in which the inputs to the controller also drive a predetermined model in parallel to the main system, the output of this model is continuously compared with the plant's output, and the controller's parameters modified in accordance.

4.2.5 STABILITY ANALYSIS AND SEARCH TECHNIQUES

Broadly speaking, two techniques are utilised in the identification processes of adaptive control systems, they are: 'stability analysis' and 'search techniques'. By using these methods a control scientist is able to establish the values of various adaptation parameters and also their convergence characteristics (i.e., whether it converges to an optimal, and if so at what rate).

The stability analysis of partially known or unknown system parameters (or performance functions) can produce adaptive control algorithms which yield estimated optimal values for such parameters. These algorithms, if found, are normally 'asymptotically stable' (i.e., converge towards a unique origin); the most common method used in this area is 'Lyapunov's Stability Criterion' (Mendel and McLaren, 1970).

Search techniques are the other important class of analytical tools, widely used in a variety of subjects. In the discrete combinatorial system structures of problems in A.I., the 'heuristic' type search procedures are used much more extensively. On the other hand, if the area of search is a continuous lattice, such as in most real-time control systems, then predominantly algorithmic search procedures are employed. Recently, many computer based techniques have been developed using both algorithmic and heuristic search, and have found applications in diverse areas such as: operational research, economics, engineering design, pattern-recognition, etc.

The problem of seeking the optimal (minimum or maximum) of a function can be applied either to systems whose parameters have to be optimised; or to the more complicated systems with unknown parameters, in which case, a measure of effectiveness (performance index) is optimised. In adaptive and 'learning' control systems when estimating the parameters of the plant or the controller search techniques are basically used to minimise some error measurement function.

In some cases, such as 'statistical decision theory' or 'dynamic programming', the probability distribution or the density functions of some system or environment parameters are to be estimated (or 'learned'); here, stochastic techniques such as 'Bayesian estimation' are employed. Another class of search techniques which are designed to operate in 'noisy' situations are the 'stochastic approximation methods'.

Many classifications of search methods have been devised (McMurtry, 1970); some of these distinctions, without much elaboration, are: 'uni-modal' (examples of uni-modal search techniques are: 'Fabonacci Search', 'gradient search', and 'steepest ascent or descent search') against 'multi-modal' ('random search' techniques have been devised to cope with multi-modal problems); 'single-dimensional' against 'multi-dimensional'; 'deterministic' against 'stochastic'; 'discrete' against 'continuous'; 'simultaneous' against 'sequential' (in sequential search the results obtained at each step are used in the future steps). Other considerations are the 'stopping rules' for the termination of search, and also the 'step size' adjustments of a particular algorithms.

The algorithmic, and more recently some heuristic, techniques which have been developed to identify and estimate the values of various parameters or functions of a system provide a powerful tool-kit for many specialised applications. However, these techniques are mostly limited to the environment and system parameters which change at a relatively slow rate when compared to the identification procedure itself; they also, generally, have specific domains of application with defined boundaries (in view of various stability considerations). The designer of a complex adaptive system has the task of dividing the system into significant yet simple subsystems for which adequate parameter estimation techniques are already available.

4.2.6 SELF-ORGANIZING CONTROL SYSTEMS

The concepts of 'self-organization' was discussed in the previous chapter. This notion refers to processes which cause a system to change from a poorly organised or non-organised state into an organised state; or alternately, from separated parts to joined parts. An obvious example being the growth of the nervous systems.

There are many controversies regarding the necessity for defining such terminologies, and whether such concepts can already be accommodated by the notions of adaptation and learning.

The earlier argument raised regarding the objectivity of defining an independent internal evaluation element also applies to the concept of self-organization. Von Forester and Ashby have both investigated self-organizing systems, and introduced some analytical methods for examining such systems. Ashby's 'Law of Requisite Variety', stated loosely as 'only variety destroys variety', has formalised the above argument by stating that a system can only be self-organizing if it is defined with respect to some external source of order, hence, implying the implicit need for an 'external' teaching or training element in any non-random system. The use of information theory concepts such as 'uncertainty', 'redundancy', and 'entropy' have made it possible to establish many important mathematical notions about self-organizing systems and their environments.

Self-organizing control systems, aside from being intrinsically non-supervised, are according to Glorioso (1975) defined as: "adaptive or 'learning systems' in which the initial state is either unknown, a random variable, or a 'don't care'." These additional conditions imposed on adaptive and 'learning systems' could be considered to be quite arbitrary, however, many self-organizing controllers have been designed in control sciences. Such controllers initially have no information regarding the nature of correct control actions; a tentative control signal is generated by the controller, and its results assessed; if there is some correlation with performance improvement, the same class of actions is pursued, otherwise, a different randomly chosen action is tested. An interesting extension of such controllers are the 'multiple-input multiple-output' control systems where the inputs are correlated with outputs in a specific or unknown manner; these ideas have had many applications, particularly in the science of 'pattern-recognition'.

4.2.7 'LEARNING' CONTROL SYSTEMS

Adaptive control systems using passive controllers were based on real-time (instantaneous) performance measurements, or alternately, on averaged performance measurements for a short preceding interval. On the other hand, control systems can be designed with active controllers, which continuously monitor and renew the 'control-law' at each step by using the previously estimated parameters and states. This second more complex category of adaptive control systems are, normally, called 'learning control systems'; they have been devised to deal with rapidly changing or very poorly defined environments where adaptive control methods cannot be utilised.

Two basic additional concepts are featured in the design of 'learning' control systems. Firstly, the controllers for such systems classify inputs, outputs, or states into classes of 'control situations', and learn the best 'control action' for each class through specific algorithms; this is necessary because, normally, the individual inputs, outputs, or states are too numerous to be stored or are ambiguously defined. Secondly, the system is provide with a 'long-term-memory' element, this provision is made to make a more extensive use of the results of previous measurements for the computation of new parameters. Often, 'learning' control systems are viewed as adaptive control systems with memory.

The time constraint on the system's performance which, as defined in section 4.2, was the main distinguishing feature of 'learning systems' is incorporated implicitly in the design specifications of any 'learning control system' - 'learning' occurs if only the average of performance over a specified number of trials shows a trend towards improvement. The performance of a 'learning control system' is gradually improved with time (at an acceptable rate), due to the identification or estimation of some parameters of the system; the size of the memory element determining the extent to which past experiences are used.

In designing 'learning' controllers, two approaches are normally used: 'parallel' and 'serial'. In parallel approach, all possible input-output combinations are considered, and the control choice is made on such basis. In serial approach (or 'performance feedback' approach), the 'learned' information is considered as the experience of the controller, and if similar situations recur such experiences will be used to update and improve the associated estimates of control or plant parameters.

The serial approach has been adapted in many control systems because of its similarity to the natural course of the learning process. The studies of learning phenomenon in psychology, physiology, and other subjects have resulted in the formulation of various 'learning' models and theories; mathematical models, in particular, have been widely used in the synthesis of control systems.

The importance of the primary concept of 'the law of reinforcement' has been recognised by many control engineers who have used this principle as the underlying criterion of the design of many 'learning' control systems. This class of systems, normally, referred to as 'reinforcement learning control systems', mainly, use the probabilistic models of reinforcement learning which was developed by learning theorists.

Reinforcements in such models are of two types of 'positive reinforcements' ('rewards') or 'negative reinforcements' ('punishments'). Rewards and punishments are used to change response tendencies; rewards represent the favourable reactions which are strengthened, and punishment is used to suppress the associated unfavourable reactions.

These two reinforcement processes normally function simultaneously, thus, fortifying the favourable responses and weakening the alternatives. In stochastic (noisy) environments, rewards are also used to increase the 'extinction threshold' of a response - making it more immune to extinction. 'Secondary reinforcements' or 'acquired/learned rewards' can also occur, whereby, the 'primary goals' or 'primary rewards/punishments' are replaced by 'sub-goals' or 'secondary rewards/punishments', this is useful if the primary goals are difficult to achieve or are not well-defined.

Another feature of such models is the way stimuli or responses are generalised into classes of similar elements. Furthermore, some secondary rewards could be used as a generalised version of a class of primary rewards. A clear distinction is also made between the concepts of learning and performance in such theories, the notion of performance referring to a measure of learning.

'Learning control law' has all three elements of adaptive control law, namely, identification, decision, and modification. Additionally, a 'memory' element is also included. In particular, the decision and modification functions are accomplished by using reinforcement principles.

A 'learning' control system operates in situations where neither environment nor plant details are known. The system, in a self-organizing fashion (i.e., with no external supervision), is able to make changes to its control actions during its real-time operation. The best control actions are evaluated at each instant of time in the absence of complete information; the basis of this evaluations is the results of previous control choices, made according to prescribed reinforcement criteria. In general, if after a control action there is a marked improvement of the performance then such control action is rewarded (strengthened). On the contrary, if there is a decline in the performance then the previous control action is punished (weakened). All such changes are stored in the memory element and used for future evaluations of control actions.

Mendel and McLaren (1970) have analyzed the above interpretation of 'learning' control procedures in a more precise form, and described four basic notions of 'mapping', 'control situations', 'sub-goals', and 'memory' as the essence of such systems. Mappings refer to transformations from points in control-choice space to points in plant-environment/state space of a system. Control situations refer to regions in plant-environment/state space which are associated with a single control-choice. Sub-goals are some intermediate type goals which are used to direct the learning process towards the optimal solution; sub-goals should be consistent with the primary goal at each separate decision stage. Finally, the memory, normally with two components of 'short-term' and 'long-term', is a separate compartment for storing pertinent information about control situations. Based on such primary notions, a heuristic 'reinforcement 'learning' control algorithm' is introduced which can be used in the synthesis of various 'learning' control systems (e.g., precise control of orbiting satellites).

Although, 'learning' control systems are, generally, not used for the purposes of investigation or simulation of biological learning, many of the concepts developed for such systems are clearly depicted from the notions of natural learning processes. Furthermore, particular parallels have been established between the elements of 'learning' control and the elements of mathematical learning theory. 'Control-choice' is equated to 'response alternative'; 'control situations' are equated to 'events'; probability distribution of various control actions is equated to the probability distribution of response alternatives; and also, the reinforcement methods of such probabilities are paralleled.

Fu (1970) describes how the concepts of statistical learning theory can be applied to the design of a linear reinforcement 'learning' control system. The learning in such a system is achieved by methodically improving the performance by the use of the control action (response) probability reinforcement. The 'quality' of control actions (responses) for different control situations (events) or, in other words, the performance of the controller is evaluated by using the outputs of the plant (outcomes). Various algorithmic methods and equations (incorporating Markov processes) are described for measuring response-probabilities, performance-indices, and evaluating the performance. A classification procedure for control situations is also outlined; this type of classification can be thought of as some sort of 'generalization' of data for parsimony reasons.

4.2.8 OTHER ISSUES IN ADAPTIVE AND 'LEARNING' CONTROL SYSTEMS

Another class of control systems, which are specially related to the higher more complex forms of learning, are the 'hierarchical multilevel systems'. In these systems separate adaptive or 'learning' control systems (or algorithms) are interconnected, and the learning at each level is influenced by the learning at other (one or more) levels. The overall goal of such systems is to coordinate and control the actions of all levels, either in a subordinate manner or in a distributed fashion.

The parameters which are estimated using 'learning' control algorithms, in general, converge asymptotically to their true optimal value as the number of trials approach infinity. However, in practice, appropriate cost and efficiency considerations are specified to limit the iterations to a finite number. Such convergence considerations, in a more complex form, are also made for multilevel type systems.

More recently, also, the concept of 'fuzzy logic' has been applied to the design of some 'learning' control systems.

In summation, although, all approaches used in this branch of science have similar properties, the specific techniques used are very diverse and require varying amounts of prior information. This fairly new discipline has benefited from the rapid development of computer technology - which has enabled the quick processing of complex algorithms in various applications. The majority of adaptive and learning algorithms developed have involved stationary environments or stationary parameters and have focused on parameter selection; algorithms for non-stationary environments have also been devised based on 'nonlinear reinforcement' theories; or alternately, the non-stationary environments have been approximated by a finite number of stationary environments.

There is a remarkable degree of complexity involved within the particular adaptive techniques used for single-level domains of application, such as the control of various processes or plants, and numerous elaborate mathematical methods have been abstracted to deal with rigid 'learning' control problems. On the other hand, global axiomatic postulates and concepts which could be adequately applied to all adaptive control systems have not been developed to the same extent.

4.3 NEURAL-NETS APPROACH TO 'LEARNING SYSTEMS'

The concepts of 'neural-networks' were introduced in the previous chapter, both in the section dealing with the models of nervous system mechanisms and also as tools for the representation and investigation of cybernetic systems.

Neural-net approach to the modelling of learning processes has been one of the principal paradigms of 'machine learning' and 'learning systems' research. Historically, it is considered as one of the earliest significant formalizations of 'learning systems', mainly, involving systems with little or no initial structural or task-oriented knowledge. Because of the primitive nature of computer technology at the outset of these disciplines, the majority of early work was either theoretical or involved the construction of experimental hardware models. Learning in neural-nets view, generally, consists of changing the probabilities of certain functions of its elements (e.g., changing the probabilities of activation of neuron-like logical elements).

Neural-net approach is closely related to the automata-theory, cybernetic and self-organizing systems approaches (to be discussed later), and often the boundaries of their methodologies overlap. Adaptive and 'learning' control systems approach (section 4.2) was also developed in parallel to the neural-net and automata-theory approach in the more engineering oriented sciences. Later, the fields of 'pattern recognition' and 'decision theory' emerged from the extensive research on neural-networks and logical-networks, these subjects will also be further discussed later.

The present diversity of approaches in 'learning systems', is partly the reflection of a dichotomy which has been evident mainly from late 1950's. On the one hand, the so called 'mainstream' A.I. scientists have been involved in the modelling of the 'macroscopic' (organizational) aspects of human learning, without attempting to simulate any of the notions of nervous systems; their studies strongly depend on the use of computers and 'heuristic programming' notions; and the models they devise do not, generally, learn from direct experience, but, use conceptual interpretations of inputs to simulate the higher levels of human cognitive learning. These workers believe that the complexity of the task is such that they need to incorporate as much of the designers knowledge in the model as possible, this is achieved by using elaborate information processing mechanisms in their designs.

On the other hand, contrary to the high-level view of the learning process, workers in the field collectively referred to as 'neural cybernetics' (Feigenbaum and Feldman, 1963) have approached the subject from a 'cellular' (component) view point. They design models which are made up of large numbers of information processing units or are composed of rudimentary elements, these models normally start the learning from very little or no prior knowledge, and use specific simple criteria to improve their performance. Neural cybernetics, which includes diverse subjects such as 'pattern-recognition', 'self-organizing systems', 'neural-nets', and 'automata-theory', is closely related to the issues of neuronal plasticity in all animals; while 'mainstream' A.I. deals with the higher brain functions of humans.

Many of the pioneers of mainstream A.I. research (e.g., M. Minsky) were originally adherents of the neural-net approach. Continuous attempts have also been made to cross over the boundaries of this dichotomy, by explaining the functions at the microstructural level using the concepts and the obtained results of the higher cognitive level, and vice versa. It is evident that neural-net researchers are ever more trying to elaborate their concepts to higher and more complex levels, while, the A.I. workers try to simplify and find the basic elements of their notions; it is envisaged that ultimately these two levels should merge to yield a complete model of the learning process.

It is fair to say that due to the limited success of neural modelling techniques and the unfulfilled original high expectations of this discipline, the research using such methods has gradually been overshadowed by the more symbolic and knowledge based type A.I. work.

In the next sections, issues involving learning in neural-networks and logical-nets together with some examples will be discussed, followed by an analysis of systems which use the concepts of automata-theory.

4.3.1 MODELS OF NEURONS

The first logical step in producing any model is to look at its working example. In the case of 'learning systems', the nervous system was the obvious primary choice for copying the possible mechanisms involved. This endeavour has been persistently constrained by the limited knowledge of the workings of the nervous system. But, it is clear that the specific concepts of interest are not inherently apparent from the physical structures of nervous systems; since, similar physical structures can display entirely different behavioural characteristics, and conversely, different physical constructs can have identical behavioural patterns. Hence, it is futile to try to study models of the brain based on the exact copying of nerve-cell connections, rather, the real interest should lie in functional intricacies.

During the late 1930's, the popular analogy of 'brains as telephone switchboard centres' suggested that the nerve cell could be considered as a simple switching relay; based on this type of view points, the original postulates used to devise the early simple models of neurons were also dependent on the 'all-or-nothing' information content of a nerve cell. The neurons were seen as information coding devices, and the brain was thought of as a large matrix of these units of information which was reacting and communicating with its environment; an implication of this 'information' view being that the brain is considered as an 'information-receiving', 'information-transmitting', and 'storage' system.

A great deal of nerve-net research started off from the original work of McCulloch and Pitts (1943). Their proposals, discussed in the previous chapter, signaled the introduction of explicit forms of cybernetic modelling. They described mathematical principles for constructing a class of computing machines whose elements were modelled on neurons, and could be used to simulate behavioural and mental theories (the only provision was that such theories should be finite and causal). They also proposed hardware models of brain cells which incorporated some of the information theory and cybernetic concepts being developed during that period.

The neuro-physiological investigations of the brain had established various relationships between pulse frequency and levels of excitation in nerve cells, but, the basic criterion used in the early nerve models was the simple all-or-none aspect of neural pulse propagation (same as the binary logic of computers). The early neuronal models ('ideal neurons') functioned by activating their outputs when the summation of inhibitory and excitatory inputs exceeded some threshold value.

A vast amount of work has been done on artificial nerve cells, and many hardware based or mathematical models have been devised, simulating various neural properties of 'excitatory' and 'inhibitory' propagation and other aspects of neuronal information transmission (e.g., Walter, 1961; Harmon, 1961; Young, 1973; Deutsch, 1967; Kent, 1978).

Neural models are useful interpretations for low-level analysis and simulation of nerve action. But, even the protagonists of such simple models (McCulloch, 1959) acknowledged that to closely resemble the complexities of nerve cell behaviour, we need to define differential equations of several orders. The fact is that the basic biological neuron is an enormously powerful device, taking a variety of specialised forms for different applications, and making optimal connections to suit the requirements of a variety of functions. And although, in neural-net approach, neurons are considered as digital information processing tools, their overall behaviour is in fact composed of many analogue and digital components; whereby, spatial locations or time and frequency variations of inputs determine the output of a cell. The neuronal processes include many utility functions as well as the functions related with the information transmission or storage of data.

It has increasingly become apparent that the neuronal models based on a single pulse propagation are not an adequate representation of information processing aspects of neurons; and the significant notions which should seemingly be incorporated in the design of such models are the frequencies of firing and the summation of the frequencies of inputs (Powers, 1978). Hence, we can argue that the real usefulness of neural-net models have been to stimulate thought and to lead to other practicable applications such as their use in pattern recognition systems or the design of computers.

4.3.2 MODELS OF NEURAL AND LOGICAL NETWORKS

Models of single nerve cells have been investigated collectively as networks of interconnected elements. The primary objectives in developing 'neural-nets' were: to examine the possibility of their use as test beds for different postulates on nervous systems, or to directly simulate the nervous-system-activity. But, more generally, neural-nets can be considered as a methodology for representing, simulating, or synthesising various dynamic systems.

Neural-nets can be considered as 'finite automata', either in 'fixed' or 'growth' forms; the elements of these automata can, in turn, be seen as finite automata themselves. Numerous investigations of such networks have been undertaken, and many aspects of neuronal plasticity modelled and simulated; furthermore, attempts have been made to simulate certain higher cognitive functions of the brain. We should note that during the development of neural-net models of nervous systems a typical cycle of 'modelling-abstraction-simulation-verification-updating' has been in progress - starting from a simple model of a neuron, developing into an abstract form and finally applied to the original natural domain.

The neural-net approach to 'learning systems', firmly rooted in the biological studies of neural plasticity, originally resulted in the introduction of simple models of the workings of nerve cells and various associative networks and mechanisms. These early developments were hampered by the limited neuro-physiological information available at the time, and also by the lack of mathematical or computing tools for treating more complex models. Later, based on such studies, other formalizations and abstract analytical elements were devised, some with no direct reference to neuronal activity; networks of such mathematical (logical) elements could produce similar properties to the networks of biological nerve models.

Some rhythmic brain activities (e.g., alpha waves) have been simulated by hardware assemblies of nerve models. Psychological and neuro-physiological theories of learning such as Hebb's (1949) recirculating storage assemblies of nerve elements, Milner's, Lashly's, and many others have been simulated and studied using networks of neuronal models. Even, in the case of randomly or non-specifically organised nerve-nets, many workers (e.g., Beurle, 1962) have demonstrated some neural manifestations of simple learning. Later, nets were designed as models for specific sensory mechanisms or for higher cognitive central processes (e.g., concept-learning, problem-solving, thinking). The introduction of the discipline of 'pattern-recognition' is a consequence of the investigations of neural-nets by early researchers such as Selfridge, Rosenblatt and Widrow in late 1950's and early 1960's.

Logic is the basis of neural-nets, and the notion of logical-net was introduced as a more abstract formalize of such networks. The concepts of symbolic logic were applied to neural networks and used in the modelling of nervous system activities. The brain was seen as a network of logical elements assembled in a cellular or matrix form. It was argued that the same way which the nervous systems are seemingly composed of elementary neurons, some logical-nets could be devised from primary logic elements to mimic animal and human activities.

Theoretically it is conceivable that all logical functions and computations could be realised by appropriate logical or neural networks; however, it is the specific characteristics of such designs, and the similarity of their operation to the natural mechanisms which determines their usefulness. The severe restrictions involved in the methodology of logical-nets, such as the consideration of these networks only at disjointed instances of time, has posed many questions regarding the validity of such models as conceptual nervous systems.

In logical-nets, the basic building block is the 'threshold element', many variations of which have been adopted. In some threshold element models the inputs could have assigned relative values, as in the case of 'linear weighted models'; or, the functioning of the element could be based on specific criteria, such as the 'majority logic'.

The single threshold element is capable of implementing ordinary Boolean Logic functions by adopting different threshold values. Many properties of threshold elements and their networks, in particular, their ability to realise logical functions, have been extensively investigated. It is worth noting that conventional logic is a special case of general threshold logic.

The threshold element is a powerful computing tool, and can be used in the implementation of computing machinery; however, so far, its apparent advantages have not been fully realised in computer system designs. The more fruitful contribution of logical nets has been to the development of pattern-recognition systems.

Various noise and tolerance considerations are of prime importance in the design of logical-nets; and investigations involving these issues have led to the establishment of the 'reliability studies' of cellular type automata with stochastic elements (von-Neumann, 1956) – a computer is an obvious example of such automata. The notions of self-repair and self-reproduction have also been investigated in some computer-based models of neural and logical nets.

Some mathematicians like Turing (1950), von-Neumann (1956) and Beurle (1962) had proposed that randomly connected logical-nets starting from an unspecified initial state could, with experience, change into a specific and well organised goal state. Here, an important question is posed, whether neural and logical nets need to be 'fixed' or 'growth' type to adequately simulate various aspects of the brain's functions; the indications are that for the low level simulation the fixed type models can be sufficient, but as the complexity of description increases, it might become necessary to incorporate some growth or 'learning' capabilities into such models.

Although, the drawing of complicated neural or logical nets (and their hardware construction) is a cumbersome task, various forms of matrix representation have been developed which together with algebraic operations simplify this task greatly, and also facilitates their computer simulation. In the abstract matrix form, however, all anatomical similarities neural-nets and nervous systems have been sacrificed for functional considerations.

Memory, in its simplest form, has been represented by a 'unit delay organ'; but, other models have been devised which show some of the properties of human STM and LTM, this is achieved by changing the threshold values of the elements during their operation.

4.3.3 SOME EXAMPLES OF LEARNING IN NEURAL-NETS AND LOGICAL-NETS

In this section we will briefly look at few examples of neural and logical nets which were devised to display some adaptive or learning features. The more advanced cellular networks which model various perceptual aspects of cognition, in particular, the classification of percepts, will be discussed in the section dealing with the pattern-recognition approach to learning.

It is mathematically possible to describe any behavioural pattern in terms of neural-net configurations, provided the behaviour can be accurately described and translated to some analytical terminology. The implication of this observation for learning behaviour is that specific nerve-nets could be constructed to display apparently 'learned' behaviour. However, in the realization of such structures, the problems of complexity and size become overwhelming hindering factors.

Consequently, the main objective of workers in this field has become the construction of 'economical' structures that could exhibit complicated learning properties; such researchers, according to George (1973), wish to build automata that mimic learning by having a "build in capacity to learn and not the details of learning itself." An essential consideration here is the compromise which should be made between the generalization and the discrimination of data in order to faithfully produce the desired descriptive level of a learning process.

The principles founded by McCulloch and Pitts, and later elaborated and abstracted by workers such as Kleene, Rashevsky and von-Neumann, have been used to construct simple networks for various behavioural patterns described in terms of logical formulae. McCulloch and Pitts (1947) devised and constructed simple neural-nets with few elements which were able to recognise (learn) different musical intervals or various visual forms. Many other workers have used neural-net models to simulate visual and other neuro-physiological systems which can display simple forms of learning.

Stewart (1967) describes simple associative networks, originally suggested by von-Neumann (1956), which can be considered as being able to learn associations between inputs. Hardware models called "Flebus" have been built which could be programmed to simulate various cybernetic properties. An example of this type of basic associative net is outlined in the schematic and non-detailed form of FIG.4.3.

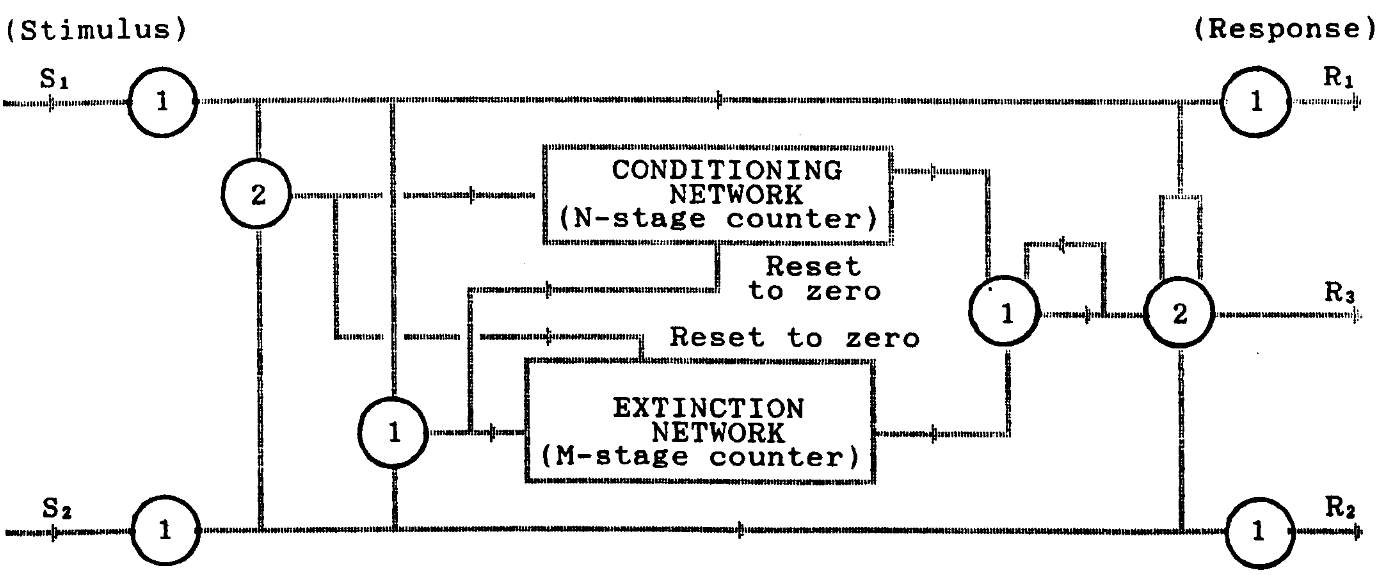


FIGURE 4.3. A simple neural-net capable of displaying response conditioning of one stimulus with a second stimulus. Initially, after firing, the stimulus element S_1 elicits response R_1+R_3 by itself alone, and S_2 elicits response R_2, after a unit instant of time. But after conditioning, S_2 should be able to elicit the response R_2+R_3 by itself.

The network of FIG.4.3 in its simplest form, with N=0 and M=0 (i.e., with no intermediate counting networks), is according to Stewart (1967): "nearly the simplest net that could exhibit 'learning' behaviour." More generally, we can say that S_2 may be conditioned to elicit the response R_3 by itself alone, provided it occurred simultaneously with S_1 for N consecutive occurrences - every non-occurrence of S_1 with occurrence of S_2 would reset the N-stage counter to zero. Similarly, the conditioned response will be extinguished if M consecutive occurrences of S_2 were unaccompanied by S_1.

It must be remembered that the use of psychological terminologies here is only a specific interpretation of a purely deterministic logical behaviour. Originally, in logical-net language, it was said that S_2 became 'a sign for' S_1 when the network established or remembered their prior incidental firings.

The above principles can be elaborated to much larger automata with memory capabilities. George (1973,1977) constructed units similar to the basic model of FIG.4.3, and assembled, from such units, automata which were able to realise some characteristics of simple learning.

George (1972) also illustrates a simple automata/environment configuration which uses neural-nets - the automata is a simple maze-running machine, and the environment is a maze. This abstract model is capable of displaying a simple 'intelligent' behaviour. The elements which are used as the basis of this system are also similar to the network of FIG.4.3, however, the conditioning and extinction networks are substituted for other appropriate networks which satisfy the specific criteria set by the design. The general class of this simple maze-running systems are also considered, such automata are deemed to have the potential to represent much more complex patterns of learning behaviour.

The above class of networks can be constructed based on a variety of other hypotheses for establishing associations between receptors and effectors; thus, various behavioural learning patterns can be described in a formal language, loosely emulating a neurological mechanism.

Stewart (1967) goes on to describe a method for the classification of input and output patterns of neural-nets. A layered net is described, which in its general form, with n input (or m output) elements can represent 2^n (or 2^m-1) different states or patterns on its receptors (or effectors). A schematic version of such classifying nets is illustrated in FIG.4.4.

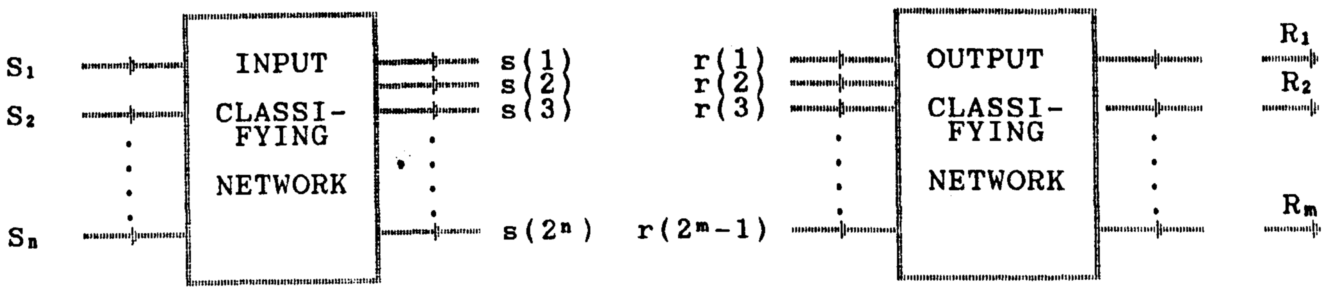


FIGURE 4.4. Schematic diagram for a possible input/output classifying method.

An automaton can be constructed by connecting the input and output classification networks together. There are $K=2^m x2^m x...x2^m$ (2^n-multiple) ways of realising such networks, including the open connections. Hence, it is conceivable to build K different types of automaton each having a distinct set of behavioural patterns.

The use of this general classification principle together with the introduction of simple memory networks, such as shown in FIG.4.5, will allow the design of automata which by employing serial or parallel information from their environments are capable of displaying quite complex behaviour for a range of situations - by selection of appropriate responses for various patterns of input, these automata may also exhibit some adaptive qualities.

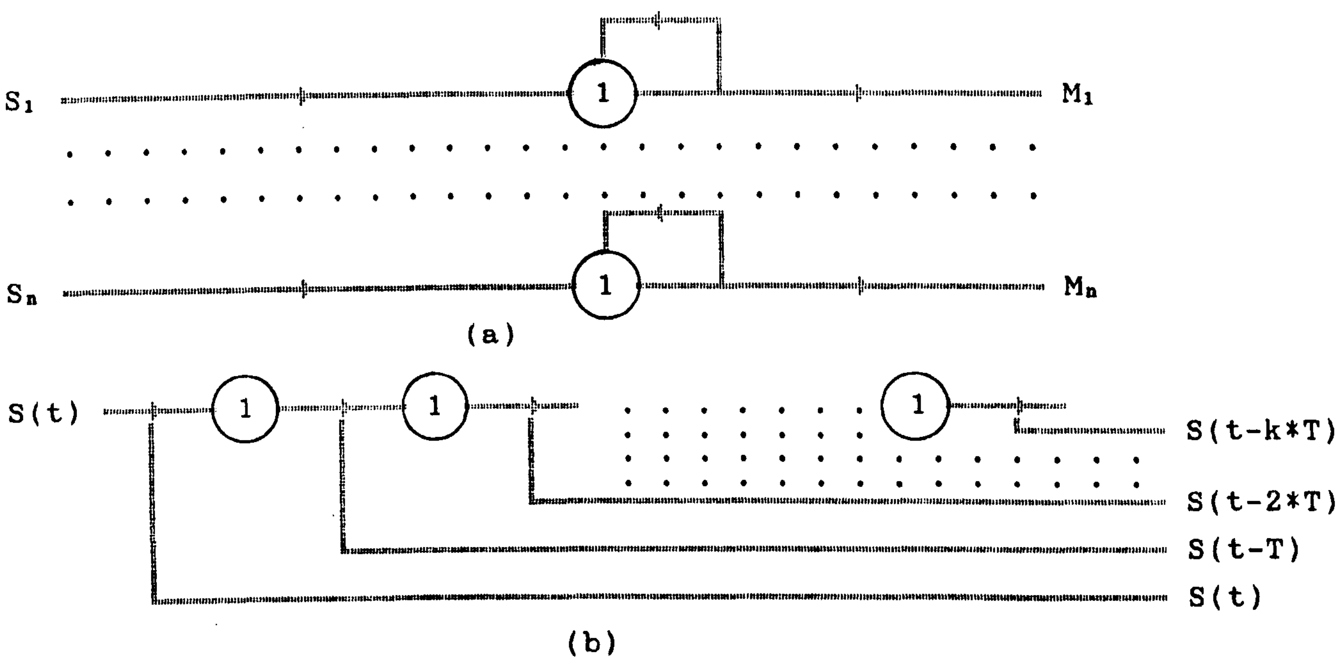


FIGURE 4.5. (a) - An n-stage simple parallel storage, once an element is activated it will keep firing; a reset function could also be incorporated.
(b) - A k-stage memory network which can be of arbitrary length, each stage of the chain will remember the preceding input that occurred a unit delay (T) time ago.

However, the feasibility of the more complex versions of such systems becomes highly questionable, since the number of the internal elements required for the realization of larger networks is extremely high; and a complete classification of inputs and outputs, and the establishment of all 'negative' and 'positive' associations becomes implausible.

A simple calculation will show that using this method of classification a typical human brain with 10^{10} nerve-cells will accommodate the sensory pattern permutations of only 33 sensory cells, therefore, clearly this type of classification cannot be a basis for natural neurological mechanisms - the patterns of activity must be the critical issue and not the individual cell firings. Hence, it can be concluded that neural-nets, to be able to learn efficiently from their environments, need to have a capacity for generalization (i.e., making inductive references).

The importance of the classification of input patterns, rather than the individual inputs, is acknowledged in this line of research. A possible hypothesis is suggested by Stewart for neural classification (and its possible use in modelling), whereby, inputs from the environment are received in the higher processing levels of nervous system in a coded form; this type of encoding can be done internally for the sensory type inputs, but for the symbolic type inputs, the encoding process can be thought to have been partly carried out in the external environment; after this encoding process, similar or associated patterns are assigned to various classes on the basis of some similar identifiable features.

It is obvious that this approach is diametrically opposite to the previous classification procedure of FIG.4.4 which involved an expansion of the number of input lines rather than their reduction. We can also deduce that only the adopting of this holistic view of the sensory percepts will enable us to devise possible efficient neural modelling methods which can explain the immensity of temporal information.

Stewart (1967) also describes a class of 'self-modifying' networks which can be devised by using the principles of 'reward and punishment'. Specific design features are incorporated which can modify stimulus-response connections of nets similar to the network of FIG.4.3. Sections of such automata could be regarded as 'motivational units', thus determining the effectiveness of associations. These motivational, rewarding or punishing, networks could be made to change according to the demands of their environments. The important observation is that as well as the association of stimuli on contiguity basis, by using some innate factors of nets, we can determine their actions by means of selective reinforcement. The principles involved could be extended indefinitely to encompass the more complex adaptive behaviours; a 'random' element could also be included to ensure that at some future time the correct responses shall be made. These models are also constructable in hardware and simple examples were given.

George (1973) describes a so called 'B-net' (Belief-net), similar to the neural network of FIG.4.3, which is the basis for his 'C-systems' (Cognitive-systems); B-nets are represented by the general form of B(m,n), n being the number of inputs, and m the number or the length of the storage elements. Generalised linear chained connections can be established between such B-nets and simple storage-nets (such as those in FIG.4.5) which are able to remember associations to any extent - as a function of: the order of occurrence, the frequency, or the number count of events. Various behavioural properties of such B-nets can also be described in terms of the probabilities of the occurrence of inputs.

George goes on to elaborate a 'cognitive' model of the central nervous system based on: 'B-nets', 'Control-nets', 'Cognitive-nets', 'Memory-nets' (long-term and short-term), 'Motivational-nets', and 'Emotional-nets'. A general form of this type of automaton described by George is illustrated in FIG.4.6; such systems will be able to classify, recognise and also reason, but not necessarily using any actual naturally occurring phenomena. The computer simulation of such complex automata are also thought to be possible, maybe, with the help of an automatic coding procedure.

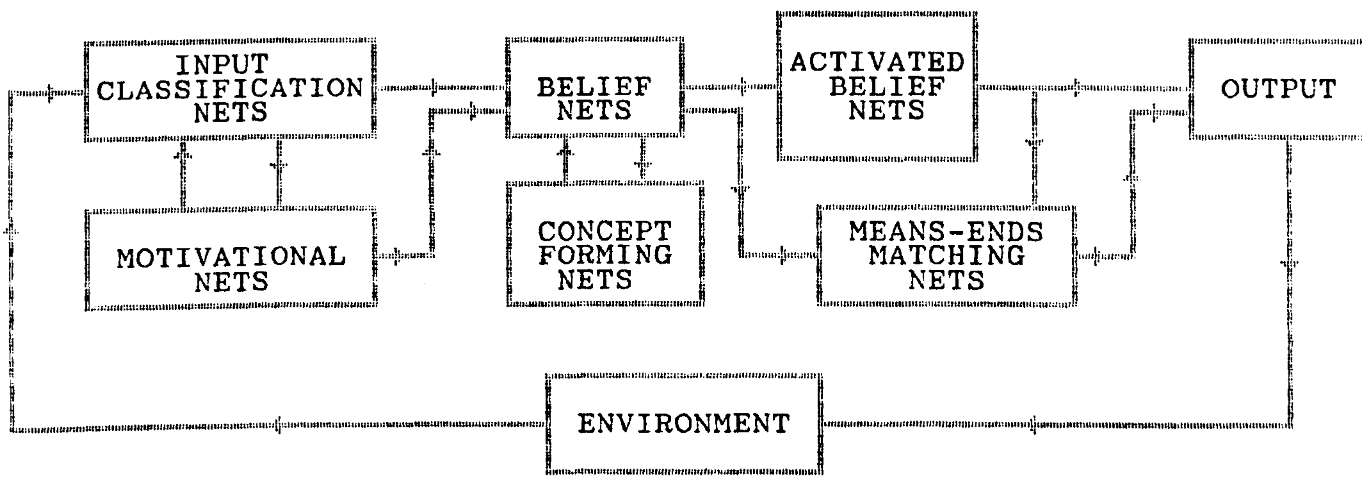


FIGURE 4.6. Block diagram of a general intelligent adaptive neural-net.

George (1973) discusses the question of perception within the theoretical context of neural-nets, proposes some molar principles for the neural-net realization of various perceiving adaptive automata, and highlights some basic issues such as: 'memory', 'motivation', 'attention', 'purposiveness', 'belief', etc. He also discusses the general problem of the simulation of behavioural and cognitive theories by neural and logical nets, with special attention given to the aspects of 'hierarchy' and 'growth' of such models.

Culbertson (1963) defines a theory of behaviour and consciousness based on the neural-net approach. 'Consciousness' is taken to mean 'mental' events or experiences (the 'mental' events are believed to 'consist' of physical events), and is deemed to be applicable to humans, animals, automata, and robots alike; here, a 'robot' refers to a finite automaton reproduced in hardware. He also abstracts various conscious automata which can display learning. It is argued that even 'memory-less' (unconscious) robots could exhibit some kind of 'intelligent' behaviour, and the 'intelligence' may be enhanced by adopting some probabilistic rather than deterministic model; and furthermore, by adding storage facility a fully conscious (intelligent) robot is conceptually realizable.

An important perceptual issue for a neural network is the 'meaning' it attaches to a stimulus, in other words, how can a neural-net automaton be made to react differently to apparently similar stimuli; the answer, generally, envisaged is that for the simple modes of behaviour the single-level logical design specifications of the net can in fact manifest such variations, but, for the more intelligent types of behaviour and learning, such as the formation of concepts, we need to incorporate a second meta-level into the design of the non-peripheral sections of the nerve-net. This feature will enable the automaton to 'evaluate' each situation and assign the proper meaning to an output. Even, a higher semantic level is also thought to become necessary for the understanding of the external descriptions of environments which are received in symbolic form.

Some neurological postulates of learning, such as Hebb's and Millner's cell assemblies, have also been represented by neural-nets in a branch of physiology sometimes referred to by 'neuro-cybernetics' or 'biological cybernetics'. George (1973) has shown how such learning cell assemblies could be constructed by neural-net configurations. Deutsch (1967) has proposed neural-net models based on neurological observations, and using an engineering oriented approach, has simulated auditory and visual neurological systems, and described models and hardware constructs for simple 'learning systems'. Amari (1977) presents primitive neural models of association and concept formation by a mathematical analysis of neural "pools"; the principle of neuronal 'self-organization', as proposed by Hebb's theories, is verified, and simple mechanisms for learning, storing and using of knowledge postulated. Other important mathematical treatments of neural-nets has been carried out by Griffith (1971).

More recently, the control engineering oriented researchers have applied some of their techniques to neural-net models of learning. Bobrowski (1984 & 1982) describes a set of formal mathematical learning algorithms and rules, which can be applied to both 'supervised' and 'unsupervised' learning neural-nets. In these models, neurons are considered as 'filters' which pass signals most frequently received; alternately, they are seen as 'detectors of rareness'. The stochastic methods used will approximate the optimal (adaptive) configuration of a network which, usually, has been laid down in advance - a decision rule at each instant of time varies certain weighting parameters until the learning is achieved. Kohonen (1984) describes a set of experiments for the formation of 'ordered neural-net maps', based on adaptive selection; these specific patterns are formed, in a self-organizing manner, as a result of sensory experiences. This process is believed to be occurring at the higher central processing levels of the brain; and simple two-dimensional neural-net arrays are devised which reproduce such maps using some 'primary laws of adaptation'.

4.3.4 AN OVERVIEW OF NEURAL NETS APPROACH

The above neural-nets have demonstrated possible ways by which both 'macro' (pattern, concept) and 'micro' (direct sensory impression) levels of inputs could be represented in a model, and also have suggested a methodology for constructing large scale experimental automata (both software and hardware). Although, mathematical techniques to sufficiently manipulate associations between such representations have not been developed, yet, to a great extent.

The introduction of neural-nets as a kind of Turing machine was considered as an important cybernetic development in the early 1940's. However, neural-nets have only resulted in the realization of trivial types of behaviour. Consequently, in the last two decades, the interest in this area has gradually waned, with some exceptions in specific subjects. On the other hand, some of the mathematical ideas and methodologies of neural-nets which spread to other disciplines, such as biology and neurology, seem to have retained their potency, and have been firmly established as a concrete tool for the analyses of natural systems.

Some conjectures as to the possible causes of this decline are: insufficiency of neuro-physiological findings; lack of early computing tools and machinery; contradictions with neurological observations; vagueness or imprecision of some terminologies to other disciplines; too many assumptions about the significances or interpretations of models; too much attention to the hardware realization; eagerness to simulate higher cognitive aspects, and define their neural-net correlates.

The comparing of the concepts of neural-nets and the actual processes of the brain has been one of the major issues in this field. McCulloch, in proposing his theories of neural-nets, assumed that humans started life with inborn fixed 'universals' consisting of various sensations, reflexes and appetites; and he thought of the brain as an interconnected mass of randomly organised nets, experiences would rearrange such connections and manifest learning, memory, prediction, and purpose. An important implication of this view was that 'mind', probably for the first time, could be described in scientific terms. However, von-Neumann (1956) recognised that although it was possible to model the processes of the brain in a digital binary form, the actual indications were that the brain itself was using a mixture of digital and analogue processes. Dreyfus (1972) in his analysis of the feasibility of computer realization of intelligence concludes that the analogy of digital computers and brains is a weak and outdated concept, and digital automata (neural-nets) cannot be used to produce a true 'artificial' model of intelligence. The problem of applying discrete criteria to the modelling of an essentially continuous system and environment, such as the brain, has also been argued by Andrew (1967,1983).

Furthermore, in view of the deficiencies of 'sequential' models in representing the processes of the brain such as learning, it is speculated that serial mechanisms are, in principle, not capable of imitating the brain which seemingly has non-serial characteristics. For this reason, some researchers have become interested in the design of parallel computing machinery and their possible applications to the neural-net type models of learning.

Digital computers have been developing alongside the more analytical computing automata of neural and logical nets, and in spite of possessing the same logical foundations, computers have come to simulate the human cognitive aspects using a very different approach. The new approach, A.I., has made the understanding at the cell level unnecessary; although, intuitively it is more appropriate to start the investigations from the point of trivial to the more complex, as promoted by the methodologies of neural-net. Hence, the 'information processing' rather than the 'information theory' aspects have dominated the field of artificial learning-systems, shifting the paradigm of 'machine intelligence' from energy and matter to information; and the computing machines have come to be considered as 'information processors' and 'symbol manipulators' rather than the crude copies of nervous systems.

Neural-nets, on the whole, have made an undeniable contribution to learning sciences. In the course of their development they have spanned a typical pattern of scientific progress, whereby, postulates are postulated upon and enormous bodies of knowledge amassed in a hierarchical form. Once the pattern of development is set, and a number of researchers commit themselves to a particular approach, generally, fundamental questions regarding the basis of a paradigm are not raised from within that particular paradigm, and it becomes quite a difficult task to break away from the mainstream views of a discipline. In addition, the discipline is diversified into complex and specialised branches which are heavily reliant upon the previously accepted postulates. Therefore, as in the case of neural-net methodology, if at some stage of development general problems are confronted and adequate solutions are not foreseen, then the whole paradigm suffers adversely rather than the specific path of the development.

Initially, the back-up necessary by some technological tools (e.g., computers) were not forthcoming; while, such supplements could have corroborated certain lines of development or proposals which were pushed aside and generally not pursued later on. For example, in the 1940's and early 1950's, it was inconceivable to compare the capabilities of the nervous systems and reasonable size computers, while, today, in many aspects (e.g., memory size and speed of calculation) computers have overtaken their neuronal counterparts.

Hence, it can be argued that some of the previously discarded ideas could be reinvestigated in the light of the new scientific and technological breakthroughs, and the questions which were thought to be impractical or vague might be seen to be pertinent now.

Various new lines of development could be envisaged if the elements of neural-networks were modelled much more closely on the actual actions of nerve-cells, or were of a more complex nature, specially in view of some recent neurological discoveries. Ultimately, the extent of future research in this direction will be governed by the interest shown in the simulation and the understanding of the structural aspects of behaviour.

4.4 AUTOMATA-THEORETICAL APPROACH TO 'LEARNING SYSTEMS'

Various concepts of 'automata-theory' and 'automata' were discussed in the previous chapter as a general mathematical tool for modelling. In this section we will survey the field of automata-theory as a conceptual framework for manifesting the process of learning in systems. A specialised class of automata, namely the combinational type cellular 'learning' automata, which normally feature the classification aspects of inputs, will be covered in the future section dealing with the pattern-recognition approach to 'learning systems'.

It is worth mentioning here that neural-nets are a special class of finite automata, and many of the previous section's discussions are also appropriate to the automata-theoretical concepts. The origins of automata-theory can be traced to the introduction of Turing-machines. However, during the early 1940's, the notions of neural-net and finite automaton were, usually, used synonymously; but, gradually, the concepts of automata-theory came to be discussed in a broader analytical sense of abstract computational machinery, while, neural-nets were seen more and more as application tools.

4.4.1 BASIC CONCEPTS OF AUTOMATA-THEORY

Automata, in analytical terms, are defined as tape-machines, or as a formal class of information processing machines. An automaton has three basic elements: 'inputs', 'outputs', and 'internal-states'; various definitions and categorizations of different types of automaton were outlined in chapter three. Of particular interest are the finite-state automata which can be described as: any system constructed from finite number of parts (e.g., cells, elements, etc.). The notion of 'finite-state-machine', in preference to Turing-machine, was introduced since it was more appropriate for abstracting fundamental models for a variety of natural and artificial information processing machines, especially, computers and switching circuits; a great deal of interest also stemmed from the rise of neural-nets.

An important distinction is made between automata, and that is the 'deterministic' against the 'non-deterministic'. A deterministic automaton in any particular state will give a fixed response to a particular input; and starting from an initial point, will behave in an identical manner for a given sequence of inputs. The 'stochastic' or 'probabilistic' automata, being a special sub-class of non-deterministic automata, on the other hand, can have a 'random' element or 'noise' incorporated, which may result in outputs not necessarily identical for every application of the same input.

Automata studies have two major components of: (a) - behavioural aspects, and (b) - structural (mechanism, anatomical) aspects. The main feature of the behavioural aspect of an automaton (provided it is not a random automaton) is that its output and state at any instant of time is dependent (deterministically or stochastically) upon its previous state and its previous input. Another point to mention here is that most learning related studies of automata involve the 'discrete' type, in which the changes of an automaton's characteristics are only considered at disjointed instances of time.

The state changes of an automaton are determined by its 'state-structure' which, in a sense, represents the dynamic organizational intricacies of a machine, yet, it does not directly refer to any actual physical hardware - the state-structure is closely related to the movement of information within an automaton. According to Aleksander (1978): "state-structure is a function of both the physical structure and the function of the elements of an automaton."

4.4.2 MODELLING OF SYSTEMS USING AUTOMATA-THEORY CONCEPTS

Automata-theory is concerned with the logical properties of dynamic and specially non-linear systems; and although, it has been developing separate from the subject of computing theory as a discipline in its own right, the concepts of automata-theory have contributed greatly to the development of computing machinery and computers. The 'reliability' studies of von-Neumann (1956) in the context of probabilistic automata is a prominent example of such contribution, whereby, stochastic automata have been used to model computers which have unreliable components; similarly, other automata-theory studies have been applied to the coding theory and the time-sharing of computers.

As well as abstracting the information processing aspects of machines, without actually referring to the hardware or the technology, the automata theorists have also been concerned with the actual input-output behaviour of various automata, and have developed methods for describing and analyzing the dynamic behaviour of mainly discrete systems. Some of these methods have been applied to computers, neural-nets, control systems, biological systems, and aspects of animal and human behaviour. In chapter three, we also outlined the various ways of representing an automaton - graphs, matrices, transition-functions, tables, etc.

Kleene (1956) demonstrated that automata could be built that for any given input situation choose an appropriate output by looking up in a kind of 'dictionary'. This means that once any pattern of behaviour has been translated to a mathematical language it can be represented by an automaton's isomorphic behaviour. The same line of argument was developed for the neural-nets of previous section, but automata-theory is concerned with the more abstract logical and mathematical consequences.

Automata-theory concepts can be applied to a broad spectrum of systems, starting from a simple clockwork mechanism to much more complex social systems. Animals and humans have also been looked at as a form of automaton, and their behaviour defined in such manner; a special advantage of this conceptualization is that it does not imply the equating of men and machines. A great deal of such analytical simulations of natural behaviour have been undertaken; similarly, many workers have been trying to translate the findings of automata-theory to the natural domain - by explaining the natural phenomena in the mathematical language of automata-theory. When psychological theories are used, the concepts of 'input' and 'output' are, normally, equated with the notions of 'stimulus' and 'response'; furthermore, the precise automata-theoretic concept of 'internal-state' is correlated with either the state of the excitation of the nervous system, or with the cognitive organization of the brain.

Automata-theory has also attempted to analyze the higher aspects of human's experiences, such as learning, thoughts, and other mental events, using a mechanistic view of these notions. Mental activities have been paralleled to the state changes of an automaton, hence enabling various quantitative analysis and evaluations based on some mathematical features of such mental activities.

Aleksander (1978,1983) argues that the state-structure of the brain is a kind of reflection of its environment, ignoring the unimportant details of the environment and making 'sense' of its experiences. He pursues the point that the automatic homeostatic/regulatory functions of the brain and also the concept of 'mind' are all attributable to the brain's state-structure. Furthermore, he describes various temporal concepts such as: 'learning', 'thoughts', 'emotions', and 'self-knowledge' in terms of the characteristics of this state-structure; conjectures are also made regarding the questions of 'free-will', 'awareness', 'attention', and 'perception', and it is argued that such features could be explained in the proposed automata-theoretic model -

by thinking of the state-structure as a hierarchical conglomerate of relatively autonomous sections which are able to control and coordinate each other. Some 'deeper' mental properties such as 'unconsciousness' and 'sleep' are considered, and speculations are made regarding the nature of various pathological problems - models proposed for a 'psychotic automaton' which could in a very basic sense display the symptoms of some psychological disorders.

Aleksander goes on to elaborate automata capable of modelling the genetic mechanisms of cell division, in particular, Kauffman's discovery of the existence of short stable cycles in the assemblies of randomly connected networks is modelled; these proposals exemplify the use of automata-theory methods in the studies of an autonomous biological activity. Finally, social interactions, such as the family relationships, are seen as another suitable domain for the application of automata-theory concepts.

An important and relatively recent landmark in the automata-theoretic research has been the introduction of Chomsky's linguistic theories, regarding 'syntactic structures'. His ideas based on the underlying structures of language has resulted in the introduction of 'artificial languages' for automata. The concept of an abstract language, having symbols and formulae rather than words and phrases, was defined for formal machines, the syntax of such a language could be expressed in terms of various 'production rules'; many developments have been achieved in this direction, some relevant examples will be covered in the future section on the A.I. approach to learning.

Before attempting to discuss learning in automata and outline some specific examples, it is relevant to make few general comments about the development of automata-theory within the past 40 years and its importance to machine-intelligence. Automata may be described as abstract machines that process information, but, it is true to say that the main influencing factor in the development of this science has been its relevance to the digital computers rather than the need for constructing abstract intelligent machinery.

As far as the modelling of the biological systems or the low-level aspects of human activities are concerned, there has been a gradual decline of interest following the initial enthusiasm; the reasons behind such a decline are more or less similar to those outlined and discussed for the case of neural-nets. However, automata-theory has been able to divert its attention and flourish in subjects relating to the formal representation of logic and language in computing machinery.

Automata theorists, on the whole, have a diametrically opposite view to workers in A.I.; and like neural-net researchers, favour the basic analytical approach to the investigations of intelligence, rather than the high-level cognitive outlook. And the computer, featured heavily in their techniques, is seen as a simple tool for simulation and not, as proposed by most A.I. workers, the sole means of attaining or simulating 'intelligence' in machines.

4.4.3 'LEARNING' AUTOMATA

Ordinary automata differ from 'learning' automata, since the majority of the non-learning automata are designed on the basis of specific state-structures which can only realise a particular information processing machine. The 'learning' automaton, on the other hand, can adapt to a specific task from a range of different tasks.

'Learning' automata interact with their environment, and modify their response on the basis of the responses they receive from their environment. Normally, the general concept of reinforcement, in the form of rewards and punishments from the environment, are used to update some aspects of the actions of the automaton for various given input situations. The automata-theoretic learning techniques have been used to model a variety of systems in biology as well as having applications in the engineering and control fields such as: optimization, adaptive control, statistical decision making, etc.

Aleksander (1983) argues that for automata (and robots) to be able to 'learn', it would be necessary to incorporate a facility for the design of new state-structures or self-programming within the machine. He also suggests a possible way for realising such automata - by devising the machine in a hierarchical form, with the higher sections being able to program (teach) some of the lower ones. In the mean time, it is pointed out that this approach could lead to an infinite regression problem.

If a deterministic 'autonomous' or an 'input-less and output-less' automaton (FIG.4.7a) with a finite number of states is considered, starting from an initial state will under the influences of its internal inputs either fall into a stable state or continue changing indefinitely (theoretically); it has also been shown that such continuous changes will invariably entail cycles.

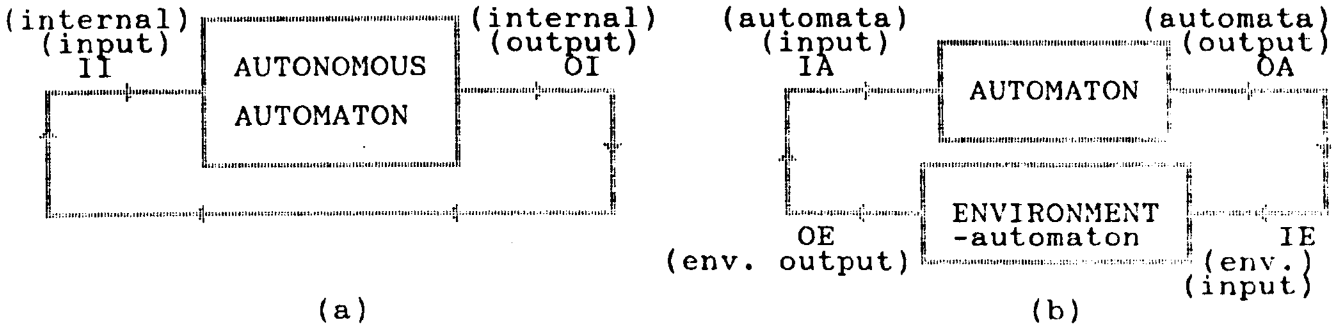


FIGURE 4.7. (a) - An 'autonomous' automaton
(b) - An automaton/environment configuration.

Now, if the interactions of an automaton with an environment is considered, as in FIG.4.7b, then much more complex systems may be envisaged; although, it must be remembered that such definitions of environment and automaton are quite arbitrary, since the environment can be looked at an automaton and the original automaton as its environment; or the whole of the automaton/environment could be seen as a single autonomous automaton.

The environment and/or the automaton may be of a probabilistic nature, whereby it reacts to the inputs according to some stochastic criterion - each output has a specific probability of occurrence for every value of the input.

An adaptive interaction of the automaton and the environment can also be considered, whereby, the environment is able to change the state-structure of the automaton. A general automaton/environment learning situation is one in which an automaton has a finite choice of actions including the option of 'no-action'. After a choice is made the environment may penalise or reward the automaton, and this, in turn, modifies the future choice of such actions in a manner that gradually a more 'favourable' response is attained; in other words, the automaton is said to be 'learning' from its environment. In most 'learning' automata, the environment is considered to be reacting instantaneously to the outputs, while the changes in the automaton occur at a unit interval after the application of input.

'Learning' automata can be categorised into two general distinct groups: 'deterministic', and 'stochastic'. In each case, a further distinction may be made, whether the automaton is 'fixed-structure' or 'variable-structure'. Deterministic 'learning' automata with fixed-structure are those whose state and output transition functions are deterministic, and although easy to implement they are rigid in response and their adaptive qualities do not improve with time. On the other hand, deterministic 'learning' automata with variable-structure could improve their learning abilities with exhaustive algorithmic strategies, however, their accurate implementation becomes extremely impractical and uneconomical.

The behaviour of a stochastic automaton can provide the type of variety needed in 'learning systems'; and the necessary modification may be achieved by changing the elements of the transition matrices of the automaton. Fixed-structure stochastic 'learning' automata are those which have time-invariant stochastic processes determining their transitions; this class of 'learning' automata, with no loss of generality, can be assumed to be deterministic in nature. The most widely used and the most suitable type of automata for the modelling of learning is the variable-structure stochastic type. These automata have probabilistic transition functions (matrices) which are updated as the process of learning evolves.

In the following, we will outline some examples of deterministic and stochastic 'learning' automata, and also briefly discuss the relatively recent concept of 'fuzzy logic' and the application of fuzzy-automata to the modelling of 'learning systems'.

(i) - SEQUENTIAL DETERMINISTIC 'LEARNING' AUTOMATA WITH SIMPLE MEMORY ELEMENTS

Aleksander (1976,1978) discusses the basis for a deterministic variable-structure finite 'learning' automaton. A general schematic diagram of such system is illustrated in FIG.4.8.

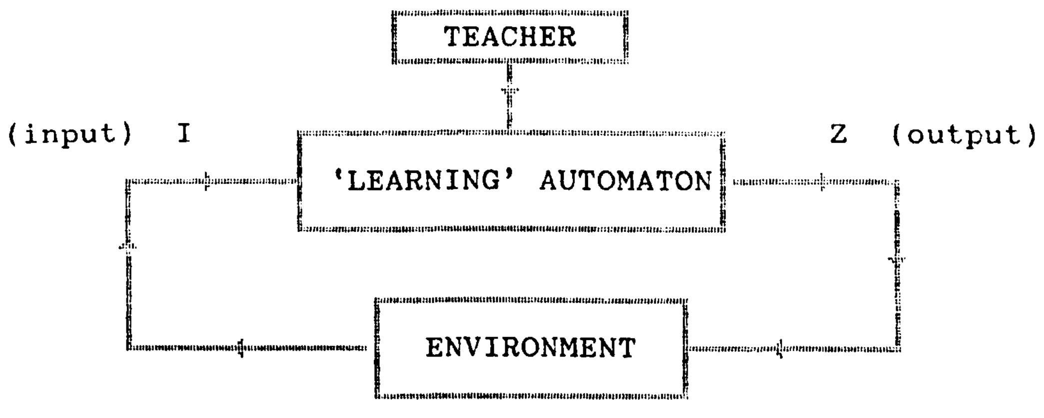


FIGURE 4.8. A trainable 'learning' automaton.

A distinction is made between 'combinational' and 'sequential' 'learning' automata, the combinational 'learning' machines are discussed and various types of decision making and classification schemes outlined; and some basic elements of such systems, in particular, Rosenblatt's 'perceptrons' are described - this whole area will be more fully covered in a future section.

The sequential 'learning' automata are devised using some feedback features from their outputs, they are able to compute and classify various properties of their input patterns, as in the case of the combinational type 'learning' automata. However, these machines are considered to be more suitable when the notion of 'training' needs to be incorporated in an automaton.

The 'learning' automata designed consist of a simple cellular network of memory elements (RAMs). Starting from an initial state, the automaton can be trained to retain a pattern of input which was applied to it. This trivial kind of 'learning' can be achieved by activating the "teach" terminals of the memory elements (RAMs), consequently, the automaton either moves into the trained state after a unit interval, or, will change into the trained state after a finite transitory steps, even though the actual training pattern has been removed. These automata are tentatively interpreted as being able to perceive, recognise, remember, learn, and switch their attention.

(ii) - AUTOMATA-THEORETICAL 'LEARNING' IN NEURAL-NETS

Veelenturf (1981) adopts an automata-theoretical approach in his investigations of learning in neural-nets. The described formal deterministic variable-structure models of growing neural-networks are based on the consideration of each element as an automaton, rather than a neuron-like unit. The 'learning' is achieved by choosing the correct behaviour from a finite set of examples, this 'learning' process includes the equivalent notions of memory and generalization or, in his words, the "abstracting of invariant structures underlying a finite set of examples."

Neural-nets are described as finite automata whose each element has both excitatory and inhibitory inputs, and can be defined by a neuronal equation which represents its firing at any time.

Using some neurological evidence, it is argued that the learning process brings about changes in the input-output behaviour of a net, and accordingly in the automata-theoretical descriptions - in the form of the state diagram and the state transition matrix of such learning and growing networks. The comparisons and the evaluations of the actual output of a net and the required output is carried out outside the system. The learning procedures used results in the gradual improvement of the automaton's response to stimuli in some desired way.

The neurological phenomenon used as the guiding principle is the observation that some cortical neurons can detect correlations between the activities of different sets of fibers, and change their synaptic effectiveness with correct response. The described learning strategies translate this observation to the analytical domain of automata-theory, by changing the weighing values of the links in the state diagram of the automaton.

Two other fundamental assumptions are made. Firstly, it is contended that to learn from the correct examples of input-output behaviours, we should only use the presently applied example to update and modify the automaton, rather than store all correct input-output examples and then simultaneously assess the information at the end of the learning period - it is thought that the biological learning process is unlikely to have this second type of data handling.

Secondly, by arguing against the two extreme learning methods of 'passive memorization' (i.e., forming a state diagram at each stage which includes all correct previous examples) and 'enumerative generalization' (i.e., finding a state diagram at each stage which is the least complicated model fully representing all correct previous examples), it is concluded that although with each method we will eventually attain the goal automaton, a more appropriate strategy would be one which adopts some essential features from both above extreme procedures; this compromised strategy is called the 'generalization memorization method'.

Furthermore, the indications are that neither of the extreme processes are involved in the neuronal level; since, with the first method, learning would be limited only to what examples had been confronted previously, and for the second method, it might become necessary to completely rearrange the network at each stage of learning, hence both strategies seem to be uneconomical and inefficient. Finally, it is suggested that this generalization memorization method, tentatively stemming from neurological observations, will be able to improve the speed of convergence of the learning process.

(iii) - STOCHASTIC AUTOMATA AS MODELS OF 'LEARNING SYSTEMS'

The ideas of stochastic automata have been applied in the implementation of various 'learning systems', in particular, to 'learning' control systems - Fu (1970), Glorioso (1975), Thathachar and Oommen (1983). Similarly, stochastic automata formulations can be applied to deterministic automata operating in stochastic environments. Some techniques have also been devised for the synthesis of stochastic automata involving their differentiation into a deterministic component plus a 'random noise generator'.

Now, without actually going too much into the intricacies of the mathematics involved, some of the basic concepts of stochastic 'learning' automata will be outlined here, particularly as we shall be using some of these formalizations in the later part of this work. This outlook is, mostly, from the control-system sciences point of view, but some psychological concepts of various learning theories, such as Bush and Mosteller's (1951) stochastic learning theories, have also been incorporated. In the following we will be, mainly, dealing with the variable-structure type stochastic 'learning' automata.

A stochastic automaton can be represented by a quintuple:-

$$Q = \{I,O,S,G,F\}.$$

I, O, and S being finite sets of inputs, outputs, and states, respectively, or:-

$$I = \{i_1,\ldots,i_r\}, \qquad O = \{o_1,\ldots,o_m\}, \qquad S = \{s_1,\ldots,s_q\}.$$

F and G are the next state and the next output functions of the form:-

$$F:\ SxI{-\!-\!-\!-}{>}S, \quad \text{and } G:\ S{-\!-\!-\!-}{>}O \text{ or } G:\ SxI{-\!-\!-\!-}{>}O,$$

and for every input i(n) and state s(n), considered at discrete instances of time, the next state and output are:-

$$s(n+1) = F[s(n),i(n)], \qquad o(n+1) = G[s(n+1)] = G[s(n),i(n)].$$

In general, the function F is stochastic and the function G may or may not be stochastic.

For each state-input pair, a probability p^k_{ij} of transfer from state s_i to state s_j with input k is defined, or more formally:-

$$p^k_{ij} = \text{Prob}\{\ s(n+1)=s_j \ \mid\ s(n)=s_i\ , \text{ and } i(n)=k\ \}, \quad \text{for all } i,j = 1,\ldots,q.$$

Also, for any present state si the total sum of the probabilities of transitions is equal to one, or:

$$\sum_{j=1}^{q} p^k_{ij} = 1.$$

The above is the basic implementation of a stochastic automaton. The problem now is to try to change the entries in the probability matrix of transition in such a way to bring about 'learning'. The general form of such a state transition matrix, M, with r inputs is as follows:-

$$M_1 = \begin{vmatrix} p^1{}_{11}, \ldots, p^1{}_{1q} \\ \vdots \\ p^1{}_{q1}, \ldots, p^1{}_{qq} \end{vmatrix}, \ldots\ldots\ldots\ldots, M_r = \begin{vmatrix} p^r{}_{11}, \ldots, p^r{}_{1q} \\ \vdots \\ p^r{}_{q1}, \ldots, p^r{}_{qq} \end{vmatrix}$$

Furthermore, the state transition matrix for a sequence of inputs $i_a, i_b, \ldots$ can be found by simple matrix multiplication of their state transition probability matrices $M_a, M_b, \ldots$ The output transition matrix can also be considered in a similar stochastic fashion, however, for simplicity reasons the output matrix is proposed to be deterministic for most 'learning' automata. For the special simple case of automata with only two value binary inputs of zero and one from their environment, "1" representing punishing reinforcement (or penalty) and "0" representing reward, the transition matrix M becomes:-

$$M_0 = \begin{vmatrix} p^0{}_{11}, \ldots, p^0{}_{1q} \\ \vdots \\ p^0{}_{q1}, \ldots, p^0{}_{qq} \end{vmatrix}, \qquad M_1 = \begin{vmatrix} p^1{}_{11}, \ldots, p^1{}_{1q} \\ \vdots \\ p^1{}_{q1}, \ldots, p^1{}_{qq} \end{vmatrix}$$

It can be easily shown that the finite deterministic automata are a special case of the general stochastic automata, with each $p^k{}_{ij}$ being either one or zero, and each row of M having all but one of the elements equal to zero.

The various stochastic automata/environment configurations are displayed in FIG.4.9. For random environments (FIG.4.9a), or deterministic or stochastic environments (FIG.4.9c and FIG.4.9e) which vary much too quickly in relation to an automaton's parameter modifications, only tentative and non-optimal strategies have been developed - since such environments do not have adequate features for making inferences.

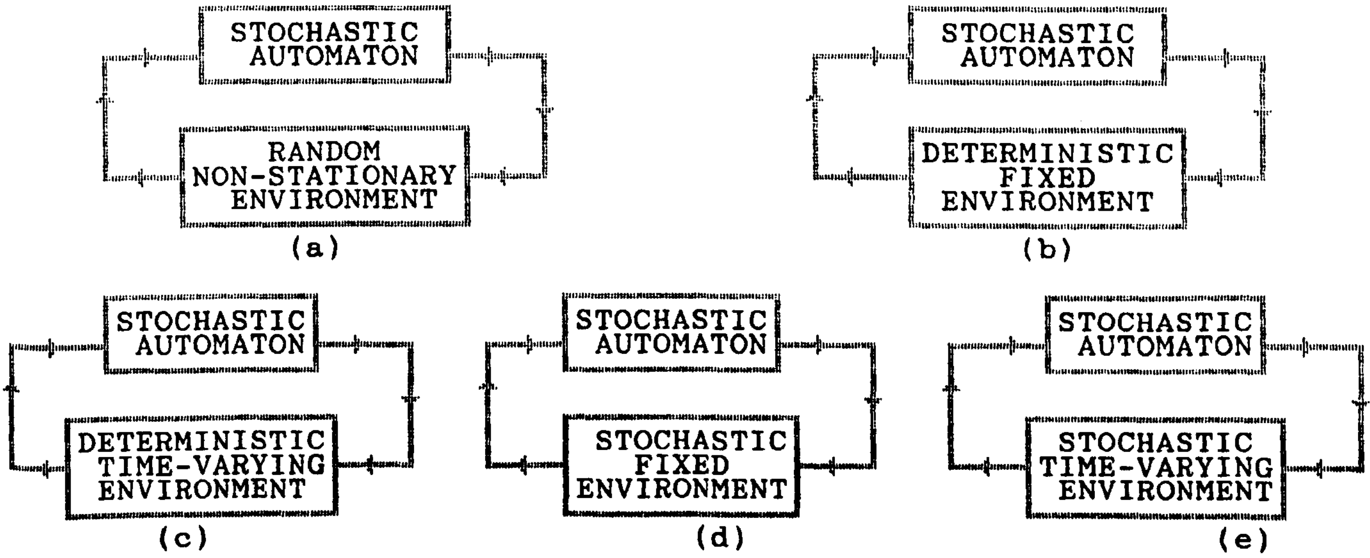


FIGURE 4.9. Five schematic configurations of stochastic automata and different types of environments.

For the fixed deterministic environments of FIG.4.9b, the input of the automaton is some function C of either its instantaneous output, or the preceding output (depending on how the environmental reactions are considered), in other words: $i(n) = C[o(n)]$, or $i(n) = C[o(n-1)]$.

The relatively slow time varying deterministic environments (FIG.4.9c) can be considered similar to the previous class of automata, and usually, the same techniques may be applied. In both cases, the function C is, generally, associated with a 'performance evaluation' element, as shown in FIG.4.10.

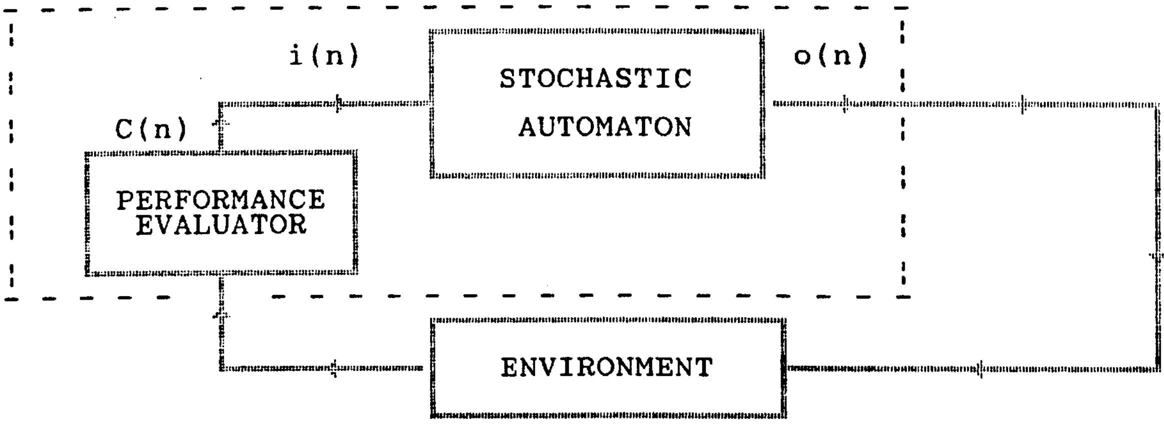


FIGURE 4.10. A generalised learning stochastic automata/environment diagram, showing the evaluation element intermediating between the automaton and its environment.

The overall measure of the performance of an automaton is given by the mathematical expectation of the negative reinforcements from the evaluation element (penalties, punishments), or T.

A measure of learning efficiency has also been introduced, called the 'expediency' which is related to the closeness of the automaton's behaviour to its optimal performance. Ideally, $T = T_{optimal}$, and the automaton is said to have optimal performance.

Now, if the assumption is made that every output of an automaton produces either a penalty (1) or a reward (0) from its environment; then, for an automaton with a fixed or slowly varying stochastic environment (FIG.4.9d and FIG.4.9e), the state s_i which deterministically leads to an output o_k (i.e., $o_k = G[s_i]$) is said to be able to produce 'punishment' with the probability:-

$$A_k = \text{Prob}\{ \text{punishment} \mid \text{given output } o_k \},$$

or in other words, $A_k = \text{Prob}\{ \text{input}=1 \mid \text{output}=o_k \}$.

Hence, the probability of transfer from state s_i to s_j can be expressed by:-

$$p_{ij} = A_k p^1_{ij} + (1-A_k)p^0_{ij}$$

For this special case, at any instant of time the value of T, or the expected penalty, becomes:-

$$T = \sum_{k=1}^{q} A_k \text{Prob}\{ \text{output}=o_k \}$$

Also, if initially the automaton attached equal probabilities to its outputs then:-

$$T_{initial} = 1/q \sum_{k=1}^{q} A_k$$

And, if $T<T_{initial}$ then the automaton is said to be 'expedient'.

Next, we consider how adaptation and learning may be manifested in such systems. If, the transition of a state s_i to s_j results in an output o_k which, in turn, produces a reward from the environment, then it is reasonable to increase the probability of such transition occurring in future, and decrease the probabilities of other transitions from s_i. Hence, p_{ij} should be increased and all other probabilities p_{ix} (where s_x is any state other than s_j) should be decreased.

The above can constitute a simple mechanism for 'learning' in a stochastic automaton - the new transition matrix reflecting the information received from the environment and the consequent self-improvement of the automaton. An alternate approach to this problem is to modify the state probabilities instead of the state transition probabilities - the probabilities for the automaton to be at each state; similarly, the modification of the input-output probabilities (i.e., Prob{ $o_k|ij$ }) could substitute the above criterion as the learning strategy.

In general, the set of parameters {m} of a matrix should be adjusted to achieve the 'adapted' or the 'optimal' system structure. If there is no a priori information about the nature of vector {m}, then, usually, equal probabilities are assigned to each state of the automaton.

If the performance of an automaton, as mentioned previously, is denoted by T(m), then we are trying to find $T(m_{optimal})$. This search for optimal performance can take several forms, depending on the complexity of the system, and also if the system is free from random fluctuations or not. There are two major classes of 'linear' and 'non-linear' reinforcement algorithms. However, for some complex problems, it becomes too difficult to devise appropriate analytic techniques, hence the so-called 'on-line' techniques have been formulated.

Probability updating algorithms and schemes, as well as the above classification, can also be distinguished into two general groups. The first group are the algorithms whose final probability vectors are dependent on the initial probabilities of the automaton; while, in the second group, the distribution of the limiting probabilities are independent of the initial probabilities - the former being the more desirable feature.

Possibly, the simplest learning scheme is the linear reward-penalty algorithm. The probabilities of actions (state transitions, input-output, state) are increased or decreased in a linearly proportional manner, depending on reward or penalty from the environment.

Here, a simple linear reward-penalty algorithm will be described, the more generalised non-linear form can also be developed in a similar fashion.

Suppose a stochastic automaton displays an output $o(n)$ at time instant n from a set of $\{o_1,o_2,\dots\dots,o_m\}$; and also there is a specific probability of p_j = Prob$\{\ o(n)=o_j\ \}$ associated with each output $o(n)$ at time n; also,

$$\sum_{j=1}^{m} p_j(n) = 1.$$

The interaction of this output with a random environment results in an input $i(n)$ of either zero (reward) or one (penalty) to the automaton. Hence, for each output a penalty probability is defined by:-

$p^1_j(n)$ = Prob$\{\ i(n)=1 \mid o(n)=o_j\ \}$, for all $j = 1,2,\dots,m$.

The environment is thus characterised by a set of penalty probabilities relating to each output; and hence the 'learning' can be manifested by updating the output probabilities of the automaton on the basis of its reinforcing inputs (either 0 or 1). In the beginning, the probability distribution of p^1_j is unknown. And for the 'learning' to be successful, the automaton should ultimately choose the actions which lead to the smaller values of p^1_j.

A measure of the performance of the automaton, as mentioned before, is:-

$$T(n) = \sum_{j=1}^{m} p_j(n) p^1_j \quad ,$$

Or the average penalty received at the instant n.

As before, the automaton is called 'expedient' if $T(n) < T(0)$; and is said to learn 'expediently', if as time tends towards infinity, the expected penalty is less than T(0).

Finally, a simple probability updating procedure with two parameters A and $B < 1$, and $\hat{o}_j$ denoting any output other than o_j, could be as follows:-

$$\begin{aligned} p_j(n+1) &= Ap_j(n), && \text{if} \quad o(n) = o_j \quad \text{and} \quad i(n) = 1 \\ &= Bp_j(n), && \text{if} \quad o(n) = \hat{o}_j \quad \text{and} \quad i(n) = 0 \\ &= (1-B) + Bp_j(n) && \text{if} \quad o(n) = \hat{o}_j \quad \text{and} \quad i(n) = 1 \\ &= (1-A) + Ap_j(n) && \text{if} \quad o(n) = o_j \quad \text{and} \quad i(n) = 0 \end{aligned}$$

The above scheme will ensure that for each penalty the associated action probabilities are reduced and other action probabilities are increased.

(iv) - FUZZY LOGIC AND FUZZY 'LEARNING' AUTOMATA

The relatively new concepts of 'fuzzy logic' and 'fuzzy sets' have been applied in the implementation of various automata-theoretic models of 'learning systems'. Some practical engineering applications of such techniques have also been envisaged.

The theory of fuzzy logic states that an assertion can have an infinite numbers of degrees of truth, represented by values between zero and one - telling us how sure we are about a proposition such as "he is tall". This is different from assigning ordinary probabilistic logical values to assertions, which tells us what chance there is of having an absolutely true value of an assertion - in fuzzy logic, there is no absolute certainty about a proposition being true (i.e., it is only true to some degree). The concepts of fuzzy logic have been particularly useful for relative notions such as temperature, height, size, etc. Fuzzy sets, loosely equated with the role of adjectives such as 'very', 'much' and 'below' of ordinary language, can be constructed as the basic quantifiers of fuzzy logic theory.

The formalization of fuzzy logic was introduced by Zadeh in the middle 1960's as an extension of Boolean logic to real-numbers; some equivalent concepts of 'AND', 'OR', and 'NOT' operators were also defined for fuzzy logic.

Zadeh (1973) describes the fuzzy approach as a substitute for the quantitative techniques of system analysis; and argues that in view of the inefficiency of computers in dealing with humanistic systems, possibly, this new approach could tackle the high degrees of complexity of such systems better than the conventional techniques. He defines the so called 'linguistic variables' in place of, or in addition to, numerical variables; and also characterises the simple relations between such variables by 'fuzzy conditional statements', and the more complex relations by 'fuzzy algorithms'. Furthermore, some 'compositional rules of inference' are devised which govern the execution of fuzzy algorithms. The principal use of fuzzy logic concepts is seen to be in the simulation of the behaviour of ill-defined or complex systems where precise descriptions are not readily available.

Fuzzy logic ideas have found some useful practical applications in engineering and A.I. But, it should be pointed out that various philosophical, theoretical, and practical objections have also been voiced. For example, it is argued that the 'context' of an assertion should also be taken into account, since it may influence and distort the objectivity of a fuzzy logic formulation.

Wee and Fu (1969) formulate a class of fuzzy automata based on Zadeh's fuzzy set concepts. Such fuzzy automata behave in a similar way to deterministic automata; yet, many of their properties are identical to stochastic automata. The inputs, outputs, and internal states of these fuzzy automata are defined as ordinary sets of finite numbers of points, while, the state and output transition functions are defined using fuzzy connotations. A non-supervised 'learning system' is proposed which is based on this formulation of fuzzy automata. The model is also applied to various engineering problems in pattern classification and control systems, and the computer simulation of a fuzzy 'learning' automaton undertaken.

Kitajima and Asai (1973), also, describe a fuzzy 'learning' automaton based on a similar formulation as above. They apply various learning algorithms and search techniques to the finding of the optimum performance of control systems with unknown characteristics. Such variable structure fuzzy automata are further investigated by means of computer simulations.

4.5 CYBERNETIC APPROACH TO 'LEARNING SYSTEMS'

Cybernetics originated from an engineering point of view, assimilating biological systems with electronic or mechanical devices; and at its outset, was, mainly, involved with tackling control problems in systems. Later, it was recognised as an all encompassing discipline, and was closely related to and overlapped a number of different subjects, including computing-theory, communications-theory, decision-theory, logic, biology, psychology, and numerous secondary areas such as linguistics, semantics, medicine, education, industry, management, economics, physiology, etc. Later, the interaction of cybernetics and computers also created many applied disciplines which have developed in their own right. Not all cybernetic workers call themselves "cyberneticians", yet, many of them acknowledge that they are working in the area of interest covered by the science of Cybernetics.

Cybernetics promised and proposed the idea of having a universal scientific language for explaining and exchanging various notions. However, as discussed in sections 3.3.5 and 3.6.1 dealing with the development and the methodologies of this subject, a common analytical language has not been forthcoming so far. The various mathematical developments discussed in chapter three (e.g., set-theory, logic-theory, probability-theory, neural-nets) which are used in the implementation of cybernetic models, in spite of being quite useful in specific domains, do not show the exactness and the generality needed in dealing with a wide range of problems.

An important observation about the science of Cybernetics is that for many purposes it does not distinguish between living and non-living systems, and it believes that various formal mathematical theories can be effectively applied to inanimate, biological, and social systems on equal basis. Ideally, cybernetics, in explaining the issues of control and communication, would need a mathematical language analogous to the Newtonian descriptions of the dynamics of interacting objects. Whereby, such a hypothetical language would be able to formally predict and explain the behaviour of all directive systems by using an axiomatic set of rules. These rules being equally applicable to the inanimate interactions in machines or systems, the biological processes, and the behavioural and mental processes of animals. So far, probably 'logic' has come closest to this notion of a cybernetic universal language.

As well as lacking in a unified explanatory language, cybernetics has also suffered from the extensiveness of its applications and the ubiquity of its ideas. It has been found that the definition of cybernetics can be extended to cover such a wide field that very few scientific disciplines are wholely excluded from it. The argument being that anything involving the concepts of 'feedback', 'control', and 'communication' qualifies to be included in cybernetics. Because of this diversity, many of the cybernetically oriented subjects have drifted away from the mainstream of cybernetics, and formed into distinct and independent research fields with no apparent reference to their originating paradigm or its underlying goals.

The definition of cybernetics has gradually become more imprecise (and sensationalised at times), and probably as a result, attracted workers from many diverse disciplines. It has also become a convenient label for use by many interdisciplinary researchers. But, in general, cyberneticians have strived towards the formalising of observations rather than giving qualitative explanations; and have been more interested in 'principles' (sometimes speculative) rather than elaborate working models. Furthermore, cybernetics has managed to unify the frame of reference of the descriptions of the living and the artificial processes.

The term 'Bionics' has also been introduced referring to the methodology of applying biological knowledge to engineering problems. Bionics, as a discipline, was established in the early 1960's (in the United States), but has not flourished to a great extent; it has been, mainly, involved in the studies of systems whose functions were based on or resembled to that of the living systems, in other words, the imitations of life functions. Although, the definition of cybernetics encompasses bionics, it has been suggested that cybernetics is more interested in the studies of living systems by the use of concepts of inanimate systems, while, bionics is more interested in the studies and the applications of biological concepts to other systems. However, such a distinction will not be pursued here and all such models will be discussed in the context of cybernetics.

According to George (1977): "...models, both in hardware and software, are a vital part of the driving force underlying cybernetics." There are, in general, three basic types of cybernetic models: 'hardware', 'mathematical', and 'descriptive'; in addition, cybernetic models are characterised by a certain 'efficiency' and 'precision' of application within a particular domain (i.e., they should be accurate and workable).

In the following section, we will outline some historical and developmental aspects of relevant cybernetic models, and discuss some examples in more detail, in particular, those involving the concepts of 'self-adaptation' and 'learning'. The cybernetic 'learning systems' included here are those which because of their particular terminologies, manifestations, or formal notations, have not been included in the other catagories of our classification of 'learning systems'. Although, many of the other approaches to 'learning systems' discussed could be considered as sub-divisions of the cybernetic approach.

4.5.1 CYBERNETICS AND THE MODELLING OF THE BRAIN, BEHAVIOUR AND INTELLIGENCE

The basic underlying assumption of cybernetic ideas is that humans and animals are essentially kinds of complex machines; and hence, in principle, it is possible to construct some artificial models which behave in a similar manner. This mechanistic behaviouristic view stems from the implicit beliefs of the sciences of biology and physiological psychology. The tendencies of cyberneticians, as behavioural scientists, seem to have been towards one of the following two general directions:-

(a) - Understanding and simulating human and animal behaviour and underlying mechanisms involved by devising or utilising precise analytical formalisations.

(b) - Developing and synthesising various systems and machines for specific tasks which can, usually, outperform humans in some aspects.

An attempt to model the brain in its entirety, irrespective of the triviality of the model components, will be an immensely difficult task. Hence, all work in this field has concentrated on specific areas of the brain, or on a specific principle involved in the nervous system. Examples are the models of the 'reticular formation', the 'cerebellum', 'eye-brain systems', 'perception', 'associative-memory', etc.

The essence of the science of Cybernetics is the concept of information and its implications in systems (i.e., coding, storage, noise, control, feedback, etc.). One of the original commitment of cybernetics was to model human behavioural and cognitive properties using some simple information processing models of neurons (i.e., neural-nets); this type of modelling did not entail a precise understanding of the physiology of nerve-cells, but a simple on-off analogue would be used. Later, the above conceptualization was also applied to the syntheses of some specific mechanisms of the brain (e.g., memory formation). George (1973) outlines the following as the basic types of cybernetic models involving the brain and the behaviour:-

(1) - Models of visual systems.

(2) - Classification models.

(3) - Conditional probability, counting and associative models.

(4) - Information processing models of particular biological systems.

(5) - Memory storage models.

(6) - Models of motivational systems.

(7) - Natural language programming.

(8) - Inference-making and theorem-proving programs.

(9) - Models using heuristic methods.

(10) - Various formal psychological and neuro-physiological theories as models for learning processes and mechanisms.

(11) - Analytical, abstract, or hardware models based on computational machinery.

At the outset of science of Cybernetics, Wiener, as well as introducing ideas such as 'feedback', 'entropy', 'information', and 'stable-state' to the study of organisms, had drawn parallels between the brains and computers. The resemblances seen between the computer and the cybernetic view of the brain are not points of much significance; however, it is assumed that, in principle, the brain can be synthesised and simulated on computers.

The various cybernetic analogues of the brain in the form of switching-networks, neural-nets, finite-automata, digital computers, etc., in spite of giving a very limited view of the mechanisms of the brain, have made useful contributions to the understanding of some aspects of the workings of the nervous-systems. In particular, the organizational aspects of the brain, such as 'classification', 'memory storage', 'generalization', and other autonomously defined functional subsystems of the brain, have been understood much better by the use of above analogies. But, in cybernetics, the emphasis has been, mainly, away from the higher 'semantic' and 'symbolic' representations of inputs and outputs, and more towards the raw analysis of information.

In the early 1950's, mathematicians such as von-Neumann and Kleene had proposed that mathematical models could be abstracted from the available neuro-physiological evidence. Von-Neumann saw the utility in the two-way interactions of concepts of 'natural' and 'artificial' organizations, and believed that the observations in one domain could be beneficial or applicable to the other. Cybernetics has been, generally, occupied with problems based on such endeavours; but, relatively, more emphasis is given to the applications of artificial criteria to the natural domain.

The behaviour of these models could be analyzed in their precise mathematical domain, however, the task of translating and equating of these analytical results to the actual neuronal processes was envisaged as the real challenge of this approach. Some McCulloch and Pitts type learning neural-nets were discussed previously, but many other cybernetic mathematically based models of learning mechanisms have also been devised.

Numerous workers have attempted to apply the concepts of 'feedback' and 'system-theory' to the motor control mechanisms of animals and humans. The spinal cords are considered as the primary centre for such processes in the body, but other regions of the nervous systems have a directing or modulating effect on the activities of the spinal cords. The motor control systems identified have, generally, a hierarchical nature of organization. Similarly, many researchers have been engaged in the identification of perceptual and cognitive sub-systems of the higher processes of the nervous systems; and have devised computer simulations or other descriptive representations of their models.

The cybernetic approach to the brain studies (as opposed to the A.I. approach), or according to Arbib (1972) the brain theory approach to the metaphor of "humans are machines", is heavily reliant on data obtained from biological and psychological observations. But he sees a definite place for the 'mathematico-deductive' methods of investigations of the brain, alongside the empirical observations of neurology and psychology. For Arbib the essence of perception and learning, in the brain or the machine, is the processing of information based on action-oriented computations taking place in 'somatotopically' (preserving information) organised networks.

However, some of the mathematical abstractions of the brain mechanisms seem to be very remote from their 'natural' counterparts. The gap between a cybernetic model and the subject it is trying to model is hardly surprising, since, normally, cybernetic tools are used to implement a conjecture which itself does not accurately represent a phenomenon - for example, in the neural-net representations of a neuro-physiological theory of learning mechanisms, such as Hebb's cell assembly theory.

Minsky (1968) acknowledges that the science of Cybernetics is the original forerunner of the majority of current trends in machine intelligence. In a brief account of the history of machine intelligence, he outlines the course of the development of the science of Cybernetics; and describes how the introduction of general-purpose computers resulted in the division of cybernetics into three main avenues.

Firstly, the main cybernetic search for simple basic principles; this approach leading into the investigations of the so called 'minimal' or 'self-organizing' systems. A paradigm of this approach was to devise collections of, generally, similar components that if arranged in specified structures and placed in an appropriate environment would eventually behave in an adaptive fashion, and hopefully display intelligent behaviour. The 'learning' models classified within this first approach included the various 'learning machines', 'adaptive' or 'self-organizing' networks, and 'automatic' or 'learning' control systems. However, in Minsky's view, the results of these undertakings were disappointing, and the explanations inconclusive, since, the systems developed worked quite well on simple specific problems but their performance deteriorated rapidly as the tasks assigned got harder, or the range of tasks were extended.

Minsky goes on to describe the second and the third directions of the development of the original cybernetic ideas as: 'cognitive simulation' and 'artificial intelligence', respectively. He is similarly critical of the cognitive simulation approach on the basis of its limitations in expressing the complexity and the diversity of human behaviour. Finally, as one of the principal proponents of the A.I. approach, he promotes the 'semantic information processing' outlook as the most promising route of enquiry; workers in this field attempt to build intelligent machines without actually having any prejudice towards biological, simple, or humanoid manifestations.

Most of the "mainstream" A.I. researchers do not consider themselves as cyberneticians. Their 'learning' models almost exoludes any direct 'learning' from raw experience, and, mainly, use an external teacher (perhaps with some exceptions; e.g., Samuel's or Andraea's work).

Many philosophical issues can be raised when discussing the nature of 'intelligence'; including the paradox of "how the brain can study itself". But, assuming that the notion of 'intelligence' can be defined by certain non-biologically dependent criteria, based on the manipulations of the information received by an organism, then the original definition of 'Cybernetics' by Wiener (1948) would imply that the concept of 'intelligence' has an important relevance to the science of Cybernetics. Hence, the question "can intelligent machines and systems be made" can be considered as one of the central issues in cybernetics.

The higher temporal capabilities of humans have been the principal subject of interest for the cybernetically oriented model-builders in disciplines such as 'computer-sciences', 'psychology', 'A.I.', 'cognitive psychology', 'pattern-recognition', etc. Although, some dedicated hardware machines have been built to simulate various cognitive aspects, in general, these class of cybernetic models have been based on the electronic digital computers. 'Problem-solving', 'theorem-proving', 'classification', 'discovery', 'decision-making', and 'emotional' properties are some of the major themes in this area.

At the higher semantic and symbolic processing levels of information, models have been devised which display capabilities such as: 'question-answering', 'language-translating', 'speech-recognition', 'text-reading', 'music-composing', 'story-composing', etc. This class of models will be discussed in the A.I. section.

It must be pointed out, that although the 'intelligence' shown by some of these machines is, at times, quite astounding; and even, it can be envisaged that some such machines will in future pass Turing's criteria for intelligence, nevertheless, there might not be any actual 'learning' involved in their processes.

On the other hand, as seen from our analysis of learning so far, the concept of learning is a very relative notion, depending on context and performance; and even the simple registering of a pattern on a photographic plate can be interpreted as a form of trivial 'learning' of a visual input which can be reproduced later. Hence, in a stricter sense, 'learning' machines should be characterised by features that bear more resemblance to our common conception of the definition of learning.

An analogy often made in literature when discussing the possible methods of devising 'intelligent' cybernetic models is the way man has managed to solve the problem of artificial flight - whereby, the problem was solved by means other than mimicking the natural wing-flapping behaviour of birds and insects. Hence, it is thought that the precise copying of the actual mechanisms of the brain is a futile task, and only some general principles which are appropriate to the specific class of models should be included in the design.

Historically, it has been the case that once a machine is devised to achieve a specific human-based task in a proficient and expert manner, and the ambiguities of a problem is solved, such as the playing of expert games of 'checkers' or 'backgammon', then the behaviour of the machine is no longer considered 'intelligent'.

If this argument is pursued for the future expert machines which will be able to display many of the intellectual faculties of humans, then we can foresee that even the most human-like machines will never be considered as 'intelligent'. Hence, in that sense, the true 'artificial intelligence' may be unattainable.

4.5.2 CYBERNETIC 'LEARNING' MODELS

The cybernetic approach to 'learning systems' has been striving towards the 'universalization' rather than the 'specialization' of the principles involved in the learning process. This alternative approach can be used as a way of analyzing the learning process which supplements disciplines such as psychology, ethology, and physiology.

Pask (1963) surveys the field of modelling of learning, and distinguishes the various types of cybernetic approaches to this problem. He points out that cybernetic model-builders have been involved with both psychological (behaviour oriented) and physiological (mechanism oriented) domains of interest.

The challenge of making true 'learning' machines and systems, which could ultimately become more intelligent than man, has constantly fascinated and occupied the minds of cyberneticians. A variety of cybernetic models have already been outlined in this chapter; but, because of their particular characterization of problems or the use of a specific formal language, such models were discussed in the independent context of their methodology.

The definition of cybernetics, however, covers a much wider range of models; here, we will attempt to list the typical characteristics of cybernetic 'learning' models in very general terms, without actually referring to specific classes of 'learning systems', and without implying that all cybernetic 'learning' models should necessarily have such features.

(1) - Cybernetic techniques are applied to the simpler forms of the learning process based on the lower levels of the hierarchy of learning in animals.

(2) - 'Learning' is, normally, manifested without an external 'teacher' or 'supervisor', and the source of learning is the past experiences.

(3) - Most cybernetic 'learning' models involve the property of 'feedback'; in other word, they can evaluate their own outputs and recognise their own mistakes.

(4) - The principal questions raised about the elements of a system are 'what' and 'why' they achieve something, rather than 'how' they achieve it or 'what' they represent.

(5) - Although, natural systems are not exactly duplicated as far as their mechanisms are concerned, the biological processes often provide the guiding principles.

(6) - The comparisons of the behaviour and the performance of cybernetic 'learning' models against biological systems, and the use of 'natural' terminologies, are quite prevalent.

(7) - A cybernetic description, normally, involves a formal non-semantic mathematical language; generally, entailing a quantification of the concept of 'information'.

(8) - Hardware models are, normally, constructable from the abstractions of a 'learning' model, and are often used for purposes of demonstration and experimentation.

(9) - While, cybernetic 'learning' models are not, generally, based on computers, there is a widespread use of computers as tools for the simulation of cybernetic ideas.

(10) - It is believed that the use of simple cybernetic principles of trivial 'learning' models will lead to the abstraction of different and more complex classes of systems; hence, these principles are thought to be applicable to many different levels of enquiry.

Another important point which should be emphasised here is that, so far, there has been no indication that any of the hardware or software cybernetic 'learning' models have had direct influences on the furthering of our understanding of the natural learning processes or mechanisms; unless, of course, the 'pure' simulation of these biological phenomena is the objective of modelling.

The researchers who set out by, initially, incorporating some 'living' phenomena in their work, usually, get engulfed in the intricacies of their formalisations or their hardware constructs; and, later, only relatively insignificant (and sometimes superficial) parallels are made between the subsequent behaviours of their models and the processes in the natural domain. Although, undeniably, this kind of research has contributed a great deal to some practical disciplines; yet, the cumbersome use of 'living' criteria for such ends is still difficult to justify.

4.5.3 CYBERNETIC MACHINES AND HARDWARE MODELS

The history of simple counting or calculating machines goes back many centuries. The earliest machines were, probably, the ancient Chinese 'abacus' type devices; later elaborated by early calculating machines, such as those of Pascal and Leibniz in the 17th century. In the late 18th century, some simple programmable mechanical machines were also introduced which used a series of perforated cards as their programs.

Babbage's (1792–1871) calculating machine (Analytical Engine), which was to be the first truly 'learning machine', was never realised, but his ideas signified the start of the modern era of computing and logical machinery, and the search for 'intelligent' machines.

Cybernetic machines which could realise the mathematical inferential nature of a logical formula in some hardware form date back to at least the early 19th century. Various examples of such machines are described by Nemes (1969), mostly capable of solving a specific logical problem; the early versions were, generally, mechanically operated, but later on, electrical or electronic components were also incorporated. These machines can be considered as forerunners of today's electronic computers which can easily be programmed to solve logical or propositional calculus problems. Another class of logical machines evolved were mainly concerned with the simplification or the minimising of logical expressions, or the economical design of switching circuits. Some of these machines, in a trivial sense, 'learned' to 'recognise' classes of inputs, and 'concluded' causal relations from logical truth tables.

The trend in the development of hardware models has been from the basic fully predictable models which simulate simple theories to the more elaborate constructions that are not fully deterministic a priori. In the latter stages of their evolution, the electronic or computer based models have almost totally dominated this area of science. The precise behavioural details of complex hardware models are not, normally, known to their constructors in advance. Although, even in the case of a simple hardware model, certain interactions of the machine with its environments might result in some unforeseen patterns of behaviour.

Various theoretical issues and the possible constructability of different classes of abstract machines have been investigated by many automata theorists and other mathematicians. Another issue of importance which must be recognised here is that all hardware models could, in principle, be programmed on a digital computer, with a complete description of the machine and its environment realised in software by an equivalent computer program. But, unless some sort of abbreviations of the physical descriptions are used, great difficulties with the speed and the memory size of the simulation will be encountered - if the variety and the sheer amount of environmental and machine information is to be represented accurately.

Machines and robots have been developed that either under the direct control of a human operator, or in a toy-like imitative fashion, will try to duplicate some motor functions of living beings. But, of cybernetic importance are those machines which, almost in an autonomous way, show some simple levels of adaptation and variability, and also bear some relevance to the natural systems. Hardware models, although interesting phenomena in their own rights, need to have a high degree of complexity to be of real interest from the point of view of a psychologist or cognitive scientist.

Many cybernetic hardware machines have also been built to copy various specific subsystems or properties of living organisms (plants, animals, and humans). The principle of 'self-reproduction' and the 'phylogenetic' developments of living entities have been simulated by some physical, as well as abstract, models; examples are the self-reproducing cellular type automata, introduced by von-Neumann. Other 'purposive', 'goal-directed', or 'motivational' behavioural simulations, generally, involving the concepts of 'pain-aversion' and/or 'pleasure-seeking', have also been carried out. Furthermore, the science of 'robotics' has been established as a direct consequence of the desire to build machines which mimic functions of life.

When assessing a cybernetic hardware model, it is, probably, as important to enquire the motivations of its designer, and the goals the model is trying to achieve, as it is to evaluate the complexity and the intricacy of the model's behaviour. A researcher may simply want to incorporate the 'state-of-art' technology in his model, or solve a specific engineering problem; on the other hand, some workers try to simulate a particular principle, or display a specific pattern of natural behaviour.

As the attention of cybernetic researchers has been increasingly focusing on the more complex subjects, there has been a gradual move away from the construction of hardware models; and software models and computer programs have been more favoured because of their relative ease of design and economy. Yet, the important instructive and demonstrative qualities of physical models cannot be denied.

4.5.4 CYBERNETIC 'LEARNING' HARDWARE MACHINES

Cyberneticians believe the conjecture that every aspect of learning, or other features of intelligence, can, in principle, be precisely described, and consequently lead to the construction of machines which simulate these descriptions.

As to what are the most important principal characteristics which should be included in a simple learning-machine. Firstly, it should have a behaviour which is not totally predicted or determined by its designer. Secondly, the learning-machine should be able to make a 'generalization' or 'inference' on the basis of the 'frequency', the 'contiguity', or the 'value' of its experiences - this feature implying a storage of information (or memory) capability, and also a 'judgement' or 'motivational' facility. Thirdly, the learning-machine should, according to its incoming and stored information, take appropriate actions which will reflect an 'improvement of performance'.

However, the above characterization does not necessarily imply that a pre-wired machine with a fixed program will not be able to display the main features of the learning process; a 'contingency', 'conditionality', or 'randomness' could be incorporated in the design, which will allow the machine to behave in the desired 'learned' manner.

In practice, the 'learning-machines' that have been designed so far do not just reiterate the abstract theories imbedded in their blue-prints, but genuinely surprise their designers, by the unforeseen capabilities and novelties of their behaviour - due to their interactions with environment, or other interferences. In some cases, the interest is directed towards discovering the full potentialities of the constructed machine, be it to simply prove all possible theories based on the logical consequences of a set of assumptions and a set of rules of inference (e.g., 'perceptrons')

Some cybernetic machines, in loosely arranged structures, can in fact behave in a learning or adaptive fashion, if placed in an appropriate environment. But, although such machines have not resulted in a general theory of intelligence, nevertheless they have made considerable contributions to some fields of engineering, control, and the design of feedback regulatory systems.

The hardware 'learning' models that will be discussed here are predominantly fixed in their actual physical structure, and the modifiability of natural neuronal mechanisms of the learning process is not, in general, depicted in such models - although, if the atomic level of changes are considered, then it can be argued that there is, in fact, a structural modification of the machine or the computer circuit.

Indeed, in future, it may be discovered that the feature of growth and change are essential characteristics of the structure of any true 'learning' entity. However, presently, the 'fixed structure' pre-wired models are thought to be capable of adequately representing all behaviour displayed by the 'modifiable structure' or 'growth' models.

(i) - HARDWARE MODELS OF SIMPLE INNATE BEHAVIOUR AND CONDITIONING PROCESS

Nemes (1969) describes some simple models which were used for the investigation of the 'process of choice' and 'instinctive behaviour' in animals. In one example, Luxe's (1920's) simple model of instinctive feeding behaviour of a protozoan, an electro-mechanical model is outlined which is capable of showing some properties of learning and remembering according to the Pavlovian conditioning criteria.

Other mechanisms, based on the concept of 'negative feedback control', have also been devised which by controlling an error function are able to keep the course of their behaviour within some defined pathways. An example of this type of machine was the "Philip dog" (1920's) which, on the principle of negative feedback from its two photocells, could operate two motors in such a way as to keep the mechanical dog on track towards a light source.

A later version of these primitive machines was Walter's (1950) "Machina Speculatrix", this machine was capable of demonstrating the interesting property of 'photo taxis' of animals, in a simple electronic device with few components; however, it lacked the capability to 'learn'. This tortoise-shaped electro-mechanical cybernetic machine could randomly move around a room and if a light source was detected by its photocell, depending on the intensity of the light, would steer towards or away from the light - simulating the innate reflexive behaviours of an animal which explores its surroundings and seeks out 'favourable' conditions. The shell of the "tortoise" was connected to a touch sensing switch which upon hitting an obstacle would send the machine into a 'search' mode.

The interesting behavioural patterns of these machines were in fact only a consequence of their 'clever' design features, and no real information processing, in the accepted 'contextual' sense of the word, was carried out.

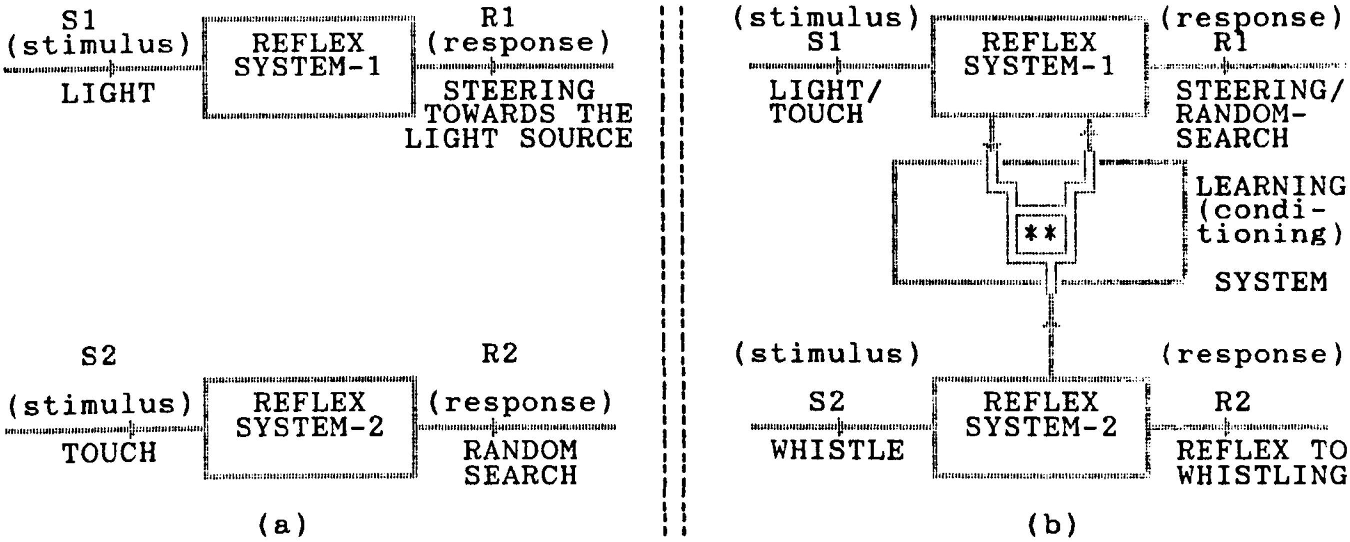


FIGURE 4.11. (a) – A functional block diagram of Machina-Speculatrix's behaviour; two reflex systems determine all the activities of the machine.
(b) – A block diagram of the functions of Machina-Docilis showing how the light or the touch reflex could be conditioned to the sound of the whistle – the '**' gate of the 'learning system' fully establishes the connection S2--->R1 as the number of the coincidences of the two stimuli exceed a certain limit over a specific period.

A simple block diagram of the reflex reactions of Machina-Speculatrix is drawn in FIG.4.11(a). The two stimuli could have priority ordering, whereby, if both occurred simultaneously, one could override the second stimulus; or alternately, their responses could oscillate. In Machina-Speculatrix, the touch input had priority over the light input.

Although, the behaviour of the tortoises were totally deterministic, many novel and unpredictable manifestations of such machines within different environments were observed; and various 'natural' interpretations of their behaviour in terms of 'feeding', 'searching', 'avoiding', 'hunger', etc. were elaborated. In some experiments, even the 'social' interactions of two or more of these devices, each having a mounted light source, were also investigated.

Walter's (1953) 'Cora' (COndition Reflex Analogue) machine was a simple response conditioning device. He used the 'black box' techniques of transmission engineers to design this machine. Initially, it would only respond to a light stimulus by flashing a neon tube, however, after the conditioning was complete, the sound of a whistle (the conditioned stimulus) alone could elicit the flash of the neon tube. Walter devised his Cora machine on the basis of the evidence provided from physiological psychology; but as pointed out by him, at times, many speculations were made regarding the actual mechanisms of the living brains in terms of the "explicit clarities" of the model.

Although, no physiological evidence could support such inferences. He also considered that one of the most important issues arising from the design of the Cora machine was the realization that three types of analogue 'memory' were seemingly necessary: (a) - a very short 'prolongation' of the effect of a neutral stimulus (iconic memory); (b) - the 'summation' of the combined effects of several neutral and specific stimuli (STM); (c) - the 'activation and preservation' over a long period of the conditioned reflex (LTM), this type of memory is also reinforced by further coincidences of stimuli.

The basic 'strengthening' and 'forgetting' characteristics of a conditioned reflex were also incorporated, by having a kind of statistical relation between the number of occurrences of the stimuli S1 and S2, and the establishment of the association between them. Hence, the design of the circuits stipulated that only significant coincidences of these stimuli would enable the stimulus S2 to evoke the response R1 by itself.

Walter differentiated between the various processes involved and defined seven distinct operations for the above process. The electronic circuits built on such basis consisted of a summing element which after a certain number of coincident inputs would trigger a positive feedback oscillator; this oscillator, representing the establishment of the conditioned reflex, had a damping component, and its response would eventually die out - hence, representing the extinguishing (or forgetting) of the conditioned reflex, unless additional reinforcement was applied by further training.

The seven operations involved in the manifestation of the conditioning (associative learning) process in Cora, or according to Walter "the seven steps from chance to meaning," are shown in FIG.4.12.

At the next logical step, Walter developed another device called "Machina Docilis" (teachable), similar to Machina Speculatrix but with the added circuitry of Cora which enabled it to manifest a simple Pavlovian (classical) reflex conditioning (or associative learning). The conditioning of the sound of a whistle with the ON state of the light or touch detecting sensors could result in the sound of the whistle alone evoking 'direction seeking' or 'search' movements in the machine. The behaviour of the machine is outlined by the block diagram of FIG.4.11(b).

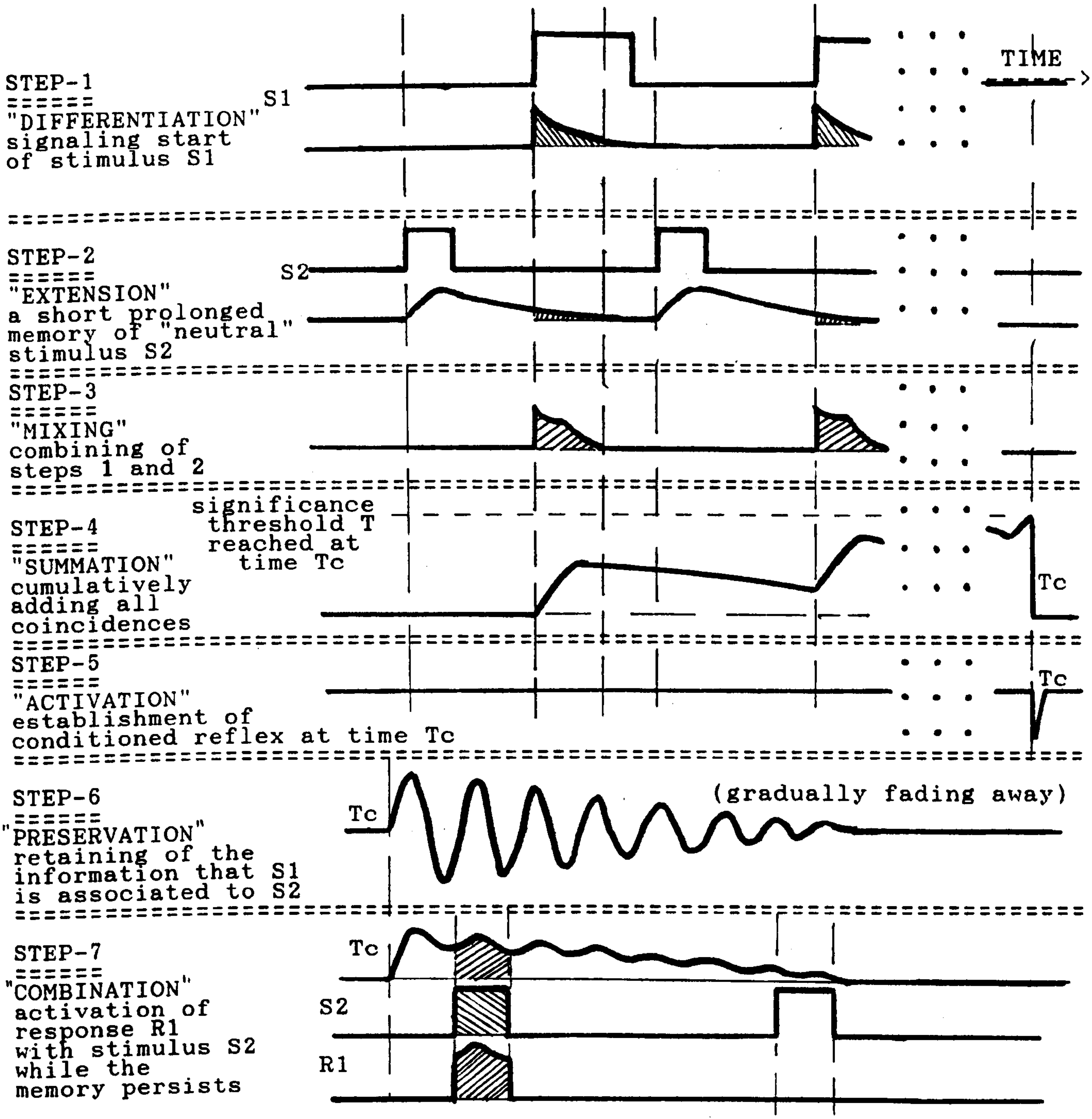


FIGURE 4.12. The seven stages involved in the establishment of a conditioned reflex in a Cora machine.

But, as emphasised by their designer, these models were only a low-level 'first approximation' of a single conditioned reflex behaviour, and many of the 'natural' features of conditioning process, such as the 'inhibition' of a response, were not included in the design. Hence, any comparisons between these simple models and living creatures would be purely speculative.

Walter (1953) also discusses the feasibility of constructing artificial nervous systems, and concludes that the building of identical number of neuronal mechanisms with appropriate interconnections is an implausible undertaking. And also suggests that the elaborations of cerebral functions are, possibly, derived from the "richness of their interconnections," rather than the sheer number of units involved. His experimental hardware models of a hypothetical simple animal with a brain having only two cells, yet capable of showing seven different modes of existence, were a demonstration of this conjecture. But, it was foreseen that constructing the more complex models on such simple principles might lead to instability problems.

Angyan (1959) describes an analogue hardware model, 'Machina Reproductrix', based on Walter's models, to demonstrate some aspects of neuronal adaptation. This mobile three-wheeled electro-mechanical device could also simulate the conditioned reflex behaviour of animals, using three types of inputs. But, the characteristics of such natural processes were depicted more accurately in these machines. The psychological notions of 'habituation', 'generalization', 'spontaneous recovery', and 'neurotic behaviour', omitted in Walter's Cora machine, were incorporated (in a trivial sense) in Machina Reproductrix. However, Angyan saw the principle contribution of such simple analogues in their usefulness in clarifying various biological and psychological terminologies.

Another class of conditioning devices have also been devised, primarily for the purposes of demonstration and teaching in psychological sciences. An example was Hoffman's (1962) 'analogue lab.' - a simple electronic device made up of switches, capacitors, resistors and batteries. It could be used as a basic model of simple reflex conditioning mechanisms (e.g., salivation response of a dog). The device was found to be able to illustrate or duplicate many aspects of results obtained from experiments on living animals.

In the 1950's and early 1960's, numerous other researchers endeavoured to build tortoise-type electro-mechanical, non-computer based, versions of these machines to simulate instinctive and simple conditioned behaviour of animals machines. The more recent computer based examples of such devices (the 'mobile robots') will be discussed in the next chapter in the context of 'learning robots'.

Young (1973) by taking into account various engineering considerations, describes a 'learning' machine which has a sort of associative memory. His basic criteria of the design of such 'learning' machines are: (1) - detecting and recording of the coincidences between the occurrences of a number of stimuli; (2) - making some use (deductions and inferences) of these stored information; (3) - having a probabilistic basis; (4) - having 'reinforcing' and 'forgetting' characteristics; and (5) - being capable of extension to higher number input machines. The frequencies of the occurrences of inputs are calculated on the 'joint probability' basis, rather than 'conditional probability' basis of 'learning' machines such as Uttly's - the justification given is the 'economy' in the necessary storage devices.

Based on the above criteria, hardware 'learning' machines called "Astra" were constructed; these devices were able to show the reflex conditioning of inputs (photocells) in a similar manner to Walter's Cora machine; however, they had some additional 'forgetting' and 'inhibition' features. The Astra 'learning' machines were also simulated on computers, and various industrial and engineering applications were envisaged, especially in the field of automation and robotics.

(ii) - CYBERNETIC MAZE-SOLVING AND TRIAL & ERROR 'LEARNING' MACHINES

Attempts to create machines which could imitate other simple aspects of animals' learning behaviour have also been made. Ross's (1933) machine and Wallace's (1952) (computer operated) machine, were both based on the simple trial and error learning of tram-like creatures, running in a system of tracks; these machines could find their goals by 'exploring' and 'discovering' the correct path through 'choice points'. But, probably Shannon's (1951) maze-running machine is the most prominent example of this type of machines.

The procedure for finding a maze can be easily formalised by a mathematical algorithm, and after this stage, it is a simple enough task to construct a machine which can 'learn' to go through a 5x5 maze by remembering its previous moves. These machines, in fact, demonstrate the apparent simplicity of some of the lower types of learning processes; in the mean time, they do not really lead us to generalizations which are applicable to the more complex learning behaviours.

Shannon's electro-mechanical "mouse" could find its way round an arbitrary maze, by seemingly learning through trial and error. This simple digital device had obstacle detecting sensors and action-programming and memory relays which could register the sequence of correct solutions on the basis of a systematic search. It searched a grid (maze), and 'learned' or 'remembered' the correct 'pathway' through the maze towards a goal. Once the solution was found ('learned'), the 'mouse' could transverse the maze in a few seconds without any hesitation. It could also show 'forgetting', if the situation was changed and the solution was no longer applicable.

The algorithm used for search did not contain any random element, every time the sensory finger hit a partition of the maze it retracted and chose a predetermined alternate direction. However, this algorithm was not fool-proof, and certain mazes could simply not be solved. To overcome this problem, if the goal was not found after a certain number of moves (24), then the previous solutions were thought to be invalid, and the machine would restart the maze-learning process - this was because the machine had probably got into a sort of 'neurotic' cyclic loop.

Shannon's machine also had the interesting property which if the connections to its memory relays were changed, or the sign of the feedback from the sensory finger was changed, it still managed to operate correctly - analogous to the plasticity of some animal neuronal tissues.

Later, Minsky also constructed similar 'learning machines' (based on electronic tubes) which could simulate some maze learning capabilities of rats. Similarly, an electronic maze-learning "mouse" is described by Jacker (1964) which was operated as a component of a so called 'bionic computer' - a conglomerate of artificial neurons. This device was able to find its way round a maze and also associate various features of the maze with its correct responses.

(iii) - CYBERNETIC GAME-PLAYING MACHINES

Game-theory, established as an independent discipline by von-Neumann, has been instrumental in the introduction of some hardware models which were capable of playing various games.

Even, as early as 1900 simple electro-mechanical devices were built which could play the particular endings of chess games - by following an algorithmic strategy. Later, much more complex fixed-wire electrical or electronic machines were introduced which showed a greater diversity and aptitude in playing complete games of chess, or solving specific class of chess-puzzles.

As well as chess, many other games such as 'checkers', 'Go', 'Nim', 'naughts and crosses', and various board or card games have also been implemented in hardware models. In some cases (e.g., naughts and crosses), a precise rigorous mathematical solution has been found for optimising moves, but other game-playing machines rely on heuristic (and sometimes 'learning') criteria.

Although, the majority of cybernetic game-playing machines do not strictly 'learn' from their experiences, their proficiency has been continuously improving. Thus, today, the descendants of the early crude mechanical devices are able to expertly play games, and outwardly show a high degree of 'intelligence'.

The development of game-theory concepts and, in particular, various search and evaluation techniques, together with the advent of digital computers, has resulted in the realization of many specialised game-playing programs (this area will be covered more fully in later sections).

(iv) - HARDWARE 'LEARNING' MODELS BASED ON ABSTRACT OR PHYSICAL PRINCIPLE

Ashby (1952) outlined the process of 'homeostasis' as a simple foundation for the entire working of the brain. He also constructed a hardware model to imitate this biological phenomenon. The term homeostasis was introduced by physiologist W.B. Cannon in the early 1930's, meaning the self-regulation of body functions, or the maintaining of the equilibrium of internal states.

Ashby, in his book "design for a brain", introduced the analytical notion of 'ultra-stability', based on the mathematical concept of stability, which was said to mathematically represent various homeostatic physical and psychological phenomena. This book was more concerned with adaptation than learning, however, he pointed out that his proposed models could display a rudimentary learning capability.

Ashby's main goal was the copying of the functions of the living brain by using alternate mathematical criteria. And his underlying contention was that most human behaviour could be explained in mechanistic terms. Hence, the idea of ultra-stability was extended to design a 'mechanical brain' in 'objective' terms; and based on such criteria, he also proposed a means of simulating the brain's adaptive and learning qualities. Ashby also described a hardware machine which was able to respond to inputs, and in a self-organizing manner change its behaviour and structure in order to achieve stability. This device, the realization of the principle of ultra-stability, was called the 'Homeostat'. The adaptive behaviour of the Homeostat was deemed to be brain-like, and was of special interest to behavioural psychologists.

The Homeostat consisted of four identical units, constructed from magnets, coils, switches, potentiometers, and other electro-mechanical components. Each unit in isolation could be regarded as a simple analogue regulator (or an analogue computer) with manual variable settings. The four units, coupled in an interacting manner, could also be considered as a multiple-feedback complex regulator which would either come to rest at an equilibrium, after a few oscillations, or would continue oscillating indefinitely. However, these oscillations could also come to rest if a new appropriate setting was found manually.

The interesting features of the Homeostat were only displayed when the control of the settings were performed by the machine itself. A process analogous to the learning and adaptive behaviour of animals could be simulated, if a 'random search' of the settings of the Homeostat was carried out - the pattern of behaviour depending on the initial settings of certain parameters. If by virtue of this search an equilibrium was attained, then the machine would stop the search, and only embark on a similar search if some of its variables were changed.

Of course, this randomness of search is an inefficient way of attaining stability, since, no information from the previous results of the settings are utilised. On the other hand, it simulates the 'inquisitive' nature of a learning animal's behaviour when it experiments with alternate options - temporary deviations from the course of a goal-directed behaviour occur, provided such deviations are in search of a more favourable outcome.

The immensity of number of moves required to find the 'stable-state' settings of a Homeostat was also recognised by Ashby (1952), and the concept of 'multi-stable' system introduced to overcome this difficulty. In FIG.4.13, a simple diagram of the 'Homeostat' is shown, each unit having inputs from the other three units.

Ashby (1956) discusses the possibility of constructing systems that are more intelligent than their designers, and argues that, in principle, it is possible to devise a so called "intelligence amplifier" which could have many applications in various socio-economic systems. The concept of ultra-stability and homeostasis is again suggested as a possible mechanism for designing such amplifiers; thus, they will be able to attain or maintain their essential variables within specified goals. The theoretical boundaries of this hypothesis is also investigated.

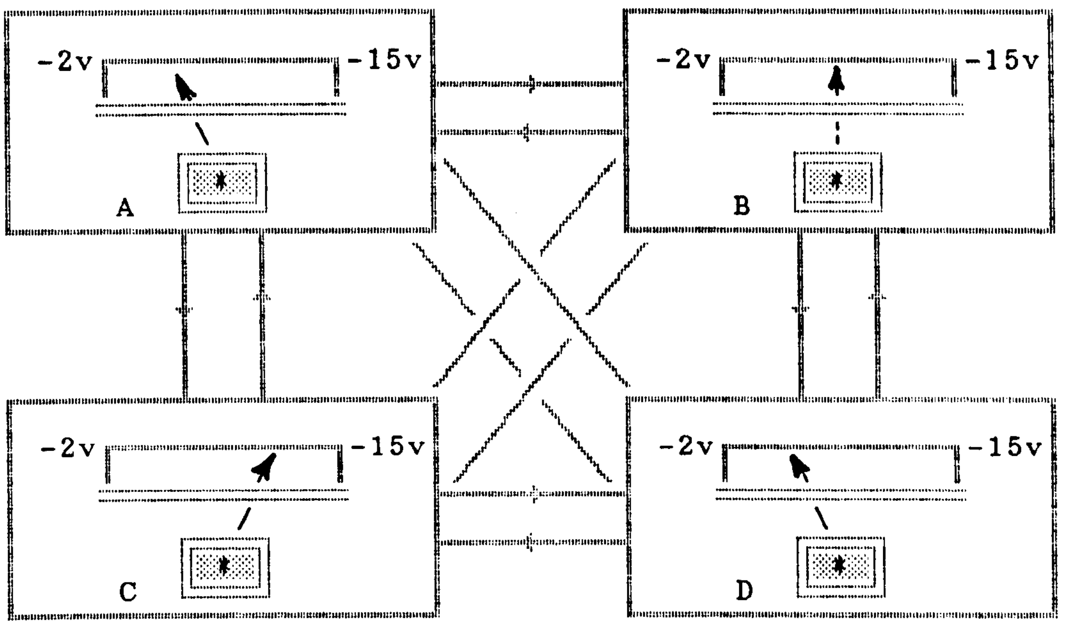


FIGURE 4.13. A four unit Homeostat; A, B, C, and D are the manual settings of the coil currents of the units. The actual physical processes involved were the changing of currents into the units, the changing of the polarity of the interconnections, or open circuiting them; but in all cases, the Homeostat would make equilibrium-seeking adjustments.

Later on, some Homeostat-type devices were designed for specific practical applications, such as the controllers for generators or autopilot systems; some devices which utilised the random selective features of the Homeostat were also described for the synthesis of certain sociological or ecological systems.

Capehart and Terry (1968) proposed a more generalised version of the Homeostat for the application to or the modelling of systems that are characterised by a random set of parameters, and where the evaluation of an optimal performance is desired. The modification suggested to the original design of the Homeostat include the addition of memory and learning capabilities, so that the exponential growth of the time required for the achievement of stable states would be avoided. The memory could store the past successful adaptations, and the learning could utilise such adaptations. Furthermore, based on a first-ordered differential equation which describes the behaviour of the Homeostat at any instant, firstly, they simulated the behaviour of a Homeostat by a computer program; and secondly, using a simple dual-type (i.e., short/long term divisions) memory and a reinforcement learning criterion, they devised a more elaborate model and its computer simulation. The results of these simulation runs were compared, and the conclusion made that a significant improvement is achieved when the additional properties of memory and/or learning are incorporated.

Uttly (1956) used the statistical notions of conditional probability, and the mathematical relations of class inclusion, to devise his theoretical 'learning' machines. He, also, constructed two non-mobile hardware models to show some basic cybernetic principles. The 'classification' model which depicted simple aspects of perception will be discussed in the pattern recognition section of this chapter. His 'conditional probability' machine was a device capable of showing simple 'learning' behaviour, it embodied basic principles which could be used to develop control mechanisms for industrial purposes.

The conditional probability machine would use the latest received information and the previously stored information to form a prediction for the future patterns of behaviour. The nature of this behaviour was not known initially, and was organised depending on the frequency and the recency of the occurrences of events - the more frequently occurring events, and also the more recently occurring events, were remembered better.

For example, if the event X followed the event Y on every occasion, then the Prob{ X ¦ Y } = 1, which meant that, based on the past information, X will always follow Y. Furthermore, Uttly weighed such probabilities in terms of their recency, so that events happening in the distant past did not have too much bearing on the behaviour. The 'patterns' of the occurrences of the inputs of Uttly's machine could be regarded either in parallel (spatially), or in series (temporally), and the machine can be thought of as simply carrying out some computations on these patterns.

Another example of an actual hardware conditional probability computer has been described by Andrew (1959); it was a machine consisting of five inputs, but machines with arbitrarily large number of inputs could also be constructed on the same principles. This machine had 31 similar counting units which could count the number of incidental occurrences of 2, 3, 4, or 5 inputs. If the value of a conditional probability exceeded a predetermined threshold, then an 'inference' was made about the activity related to that particular conditional probability. The counts were indicated by the amount of charge on leaking capacitors, hence, their value was gradually decreased.

This type of conditional probability models bear a close kinship to the notion of 'expectancy' in psychology, and we can also intuitively see similar processes at work in humans and animals. But, as admitted by Uttly, his machines were a simulation of 'idealised' human behaviour; and it has been realised that to accurately model a pattern of human behaviour in terms of conditional probabilities a great deal of elaborations, constraints, and other new concepts (e.g., hierarchical ordering) should be introduced to the basic conditional probability model. Consequently, it becomes very difficult to devise complex models of behaviour on such basis, and also the original principles are all but lost in the featured complexities.

George (1977) sees the principal shortcoming of Uttly's conditional probability computers in their inability to distinguish between the 'linguistic' considerations of stimuli and the, so called, 'factual' considerations. In other words, the inherent differences which exist between physical symbols and what they represent.

(v) - OTHER HARDWARE CYBERNETIC MODELS INVOLVING LEARNING OR ADAPTATION

Pask (1969) constructed a cybernetic hardware model which, to some extent, depicted the process of growth of nervous tissues. An electrochemical system, consisting of wire electrodes suspended in an acid solution, could simulate a chain of increasing and diminishing growth patterns, in the form of electro-deposition. Some non-reversible permanent changes could also occur. This type of model is, however, only an 'analogy', and not a hardware realization of some mathematical postulate.

Young (1973) adopts a 'cybernetic engineering' approach in the construction of 'learning' models. He constructs electrical analogue circuits to simulate various properties of nerve cells and their assemblies in a simple form. These circuits were able to, in a very trivial sense, by using the transmission of electrical pulses, depict neuronal propagation of information, generate rhythmic activities, and simulate closed-loop recirculating type memory storage systems (which could also be effected by various noise considerations). Therefore, simulate a simple neurological basis of learning and memory.

Mathematical formulations of these rigid engineering models were also undertaken, based on some transfer functions. The "Astra" 'learning' machines were an extension of this line of research. Many other researchers have also attempted to simulate the neuronal activities in this electrical-mathematical form.

We must not forget to mention that the hardware neural-net models outlined, and discussed previously, form an important class of cybernetic machines. Neural-net and logical-net blue prints could be realised in hardware (e.g., George, 1961; Stewart, 1959), directly from their logical descriptions; and some characteristics of simple 'learning', 'memory', 'classification', 'motivation', 'reinforcement', and 'conditioning' could be displayed; in such endeavours, the hardware systems were found to be much easier to construct than to describe the complete behaviour of its equivalent in logical form.

4.5.5 CYBERNETIC 'LEARNING' MODELS AND COMPUTERS

Digital and analogue computing machines with some sort of memory storage are very important tools for the cybernetic model-builders. A digital computer can easily be programmed to display a typical process of learning or response conditioning. Similarly, various learning theories can be simulated and tested by means of computer programming. The major difficulty in programming computers to actually 'learn' from their experience is the problem of the 'generalization' of data.

The cybernetic designers of 'learning' machines are not, generally, interested in the step-by-step imitations of the natural learning process, specially, in view of the impracticality of storing the large number of possible variations of behaviour - again, emphasising the importance of the concept of 'generalization' to the learning process. Now, there are four basic ways that 'intelligent' patterns of behaviour (e.g., learning) could be implemented on computers:-

(a) - They can be programmed to follow a set of instructions step-by-step, and display the appropriate behaviour, in an open-loop fashion.

(b) - Some form of working (previously conceived) 'intelligence', relating to the task in hand, can be incorporated within their programs.

(c) - Their programs could be governed by some tentative 'heuristic' instructions which do not guarantee a solution.

(d) - They could be programmed on the basis of some general and elementary principles, not, primarily, devised for a particular class of problems, and whose final outcomes are not normally predictable.

'Feedback' could be an important but not an essential feature of the last three catagories. These four basic methods, illustrated schematically in FIG.4.14, could also be associated with 'simulated', 'algorithmically determined', 'heuristically determined', and 'minimally determined' behaviours, respectively.

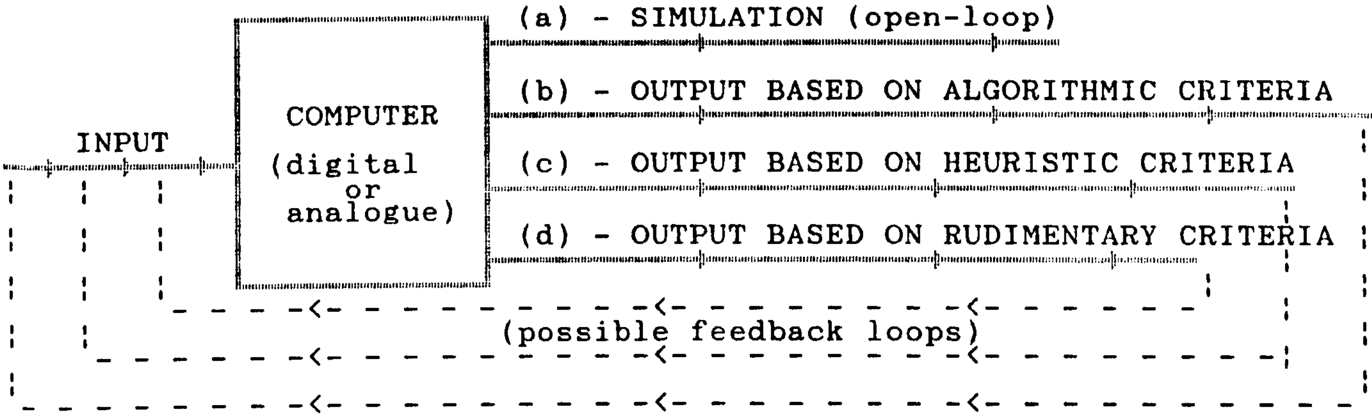


FIGURE 4.14. A schematic diagram of the four principal methods of programming a computer to display interesting "intelligent" patterns of behaviour.

Computer programs, in general, can be described as a special class of cybernetic models. However, because of our particular taxonomy, 'learning' or 'adaptive' computer programs will be discussed in later A.I. sections.

4.5.6 ABSTRACT CYBERNETIC MODELS INVOLVING LEARNING AND ADAPTATION

In this section we will attempt to outline some analytical software cybernetic models which have tried to embody the notions of learning and adaptation. It is clear that simply collecting empirical data and restating them in a different form, by itself, is not a very fruitful scientific pursuit, and some general interpretations and integrations of results are necessary. Descriptive or mathematical models provide a vehicle using which various deductions can be made and theories can be constructed. A principal advantage of the mathematical representations is that a much higher level of objectivity of terminology can be attained, and logical imprecisions can be detected much more easily.

The mathematical formalisations used for the modelling of learning have been quite diverse. The actual choice of the methodology depending on the specific application, and the individual worker's background. The trend in using mathematical techniques for devising learning or adaptive systems has spanned from the classical modes of mathematics, such as linear algebraic equations, to the more symbolic mathematics of modern (abstract) algebra, such as logic or set-theory.

An often neglected issue in this area has been the translation and the interpretation of results in one formal descriptive languages in terms of other terminologies. Similarly, the relative merits of different techniques in tackling a particular problem is another important consideration in comparing mathematical models of the learning process.

Within the past few decades, the behavioural sciences have become increasingly more reliant on the use of mathematics for understanding various phenomena. The two most important conceptual tools, by far, are the statistical probability theory and the digital computers.

Uttly (1956) translated some characteristics of animal behaviour, such as the evoking of the same response by different stimuli, into the mathematical domains of set-theory and probability-theory. From such formal basis he described a machine whose behaviour was similar to those of animals' in a number of ways; also, the structure of the machine could be identified with some elements of the nervous systems. Again, the simple all-or-none representation of inputs were adopted, in addition, various other restrictions were imposed on this abstract model.

The analytical machines he described could simulate a hypothetical experiment in conditioning a reflex. The basic underlying assumption was that the incoming data are classified according to resemblances between sets of input data, rather than the concepts or the phenomena which give rise to such data. In support of this premiss, Uttly states "the nervous system is limited similarly to assessing resemblances between signals in sets of fibers, not between sets of physical quantities external to this system, and from which those fibre signals were derived - between internal representations, not between external 'configurations'."

Uttly's models also typify the trend in the early cybernetic models which, based on some simple mathematical criterion, promised and predicted a great deal, but in reality their rigid levels of abstraction did not allow any significant developments in devising practical models of interesting qualities.

Mackay (1956) discusses the issues involved in designing intelligent automata, especially, the representation and the limitations of the "universe of discourse" of an automaton (the field of its activities) is scrutinised. He distinguishes two different approaches to constructing such automata, vaguely reflecting the dichotomy which has developed between the cybernetic and the A.I. approach in designing 'learning' machines - 'trial and error' versus 'fully informed'. He also proposes a statistical mechanism which could display 'mind-like' behaviours such as learning. Furthermore, he contends that using such principles any level of abstraction, including 'meta-linguistic' concepts, could be developed - by simply adaptively considering the regularities of sensory inputs. He concludes that some similarities between his abstract formalisations and the actual brain processes could be seen, but they are inadmissible as evidence for the workings of the brain.

Uttly's 'conditional probability' model and Mackay's model were examples of stochastic models of behaviour, and could be regarded in the same category as the mathematical descriptions of behaviour and learning in psychology. Indeed, in a sense, all mathematical learning theories could be regarded as cybernetic.

The essence of mathematical analysis of the learning process in psychology is the probability theory; in addition, many related mathematical techniques, such as Markov chain theory, have also been applied to such investigations. It is generally thought that Hull's (1943) learning theories signified the introduction of precise mathematical descriptions of the learning behaviour, on par with other formalisations used in different physical sciences. The principal trend has been, mainly, away from finding a universal mathematical theory of learning, applicable to all learning situations, and more towards devising specific theories and techniques for individual learning situations and paradigms.

But, on the whole, it must be emphasised that most mathematical learning theories are only the mathematical expression of a verbalised descriptive theory of learning formulated prior to the construction of the mathematical abstraction. However, in some cases, a mathematical model is devised to explain or fit to data obtained from a set of experiments. An example of this second type of model was Bush and Mosteller's (1951) 'learning' models, they used probability of outcomes as a measure of behaviour, and attempted to relate their analytical results to experimental observations on learning, in particular, to reinforcement (acquisition and extinction) aspects involved in simple learning experiments.

Another cybernetic theory of behaviour is also proposed by Deutsch (1960). Deutsch's models are an attempt to put to order the immense amount of accumulated psychological and neuro-physiological evidence on the subject of learning. His theories are 'structural' explanations of behaviour in terms of possible, previously postulated, neural mechanisms. He postulates a process by which learning might occur in organisms, and defines five separate units, as the principal components of any 'learning system': (a) – an analyzer (input subsystems), (b) – a link (channel), (c) – a motor system, (d) –an internal medium, (e) – an environment. The elements of the system could have three types of relationships: 'activation', 'inhibition', and 'causing change'. The basic elements of such models are shown in FIG.4.15 which illustrates the components of a postulated mechanism of 'need'.

The basic process of 'learning' is defined as the sequential firing of two analyzers, leading to the formation of links between each other - which signify the establishment of associations between analyzers. The manifestation of learning also results in the detachment of some system elements from analyzers. Similarly, 'extinction' and 'forgetting' are defined in equivalent terms.

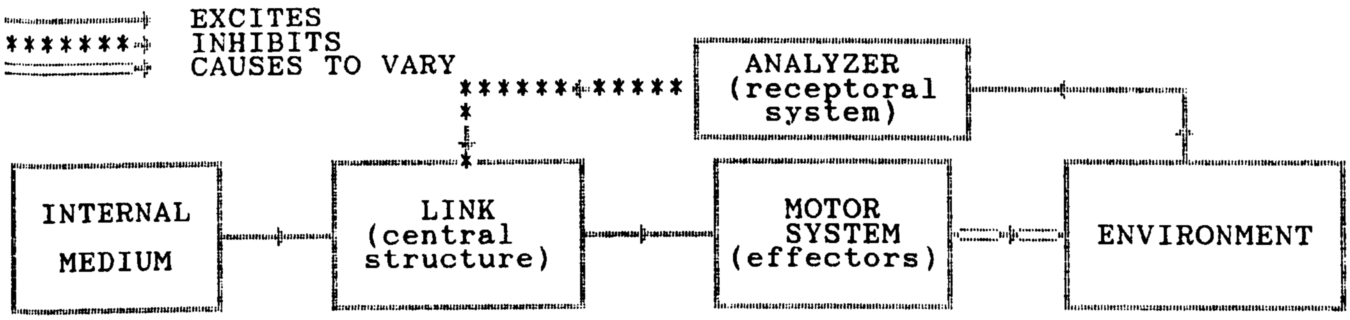


FIGURE 4.15. A simple model for 'need' mechanism, based on Deutsch's elementary definitions of components and processes involved.

A mechanical device was also constructed to demonstrate a simplified version of Deutsch's theories. The machine consisted of a turtle-type trolley running in a maze, having touch sensors, and was capable of being remotely controlled from a central part made up of relays and uni-selectors. Besides the instructions received from the control centre, the machine had a set of in built fixed stimulus-response actions. The main distinction from the earlier similar devices, described previously, was that the photo-sensitive sensors placed around the maze, which told the central part the position of the machine in the maze, were in fact regarded as part of the system, rather than elements of the environment. The machine could learn simple mazes, by running through the correct path in a 'training' stage. Some properties such as 'generalization', 'reasoning', 'insight' were also shown in a very limited sense. The mechanical trolley was able to demonstrate an interesting variety and generality of learning behaviour within the context of its simple domain. Although, Deutsch emphasised that the properties observed are only a consequence of arrangement of relays, and not those of relays themselves; and envisaged that such systems could be expanded, without a loss of efficiency, to demonstrate more complex psychological properties.

Ashby (1967) also develops his ideas on homeostasis within the more rigorous but abstract framework of set-theory. He postulates a basis for representing all types of adaptive mechanisms, natural or otherwise, by using a formal "simplification" which is homoeomorphic with our observations - not based on intuitive rules of thumbs. The theories and hypothetical mechanisms proposed, according to Ashby, could embody and demonstrate various notions, such as Shannon's concept of information or Sommerhoff's 'directive correlation'. However, because of their high degree of abstraction (in search of a high degree of precision), they have not been used for many practical modelling purposes.

As outlined in chapter three, many other mathematical developments have also been used as such cybernetic tools. In the following, some other important (historically or otherwise) examples of cybernetic 'learning systems' which utilise a specific analytical methodology will be discussed.

(i) - GENERAL SYSTEM THEORY APPROACH TO CYBERNETIC 'LEARNING' MODELS

Systems that can autonomously modify and adapt to their environments (changing or stationary) have been described in a variety of forms, 'mathematical', 'logical', 'descriptive', etc. Cybernetics, in general, is interested in the behavioural and structural studies of such 'self-adapting' systems, particularly, in those systems related to some aspects of living organisms.

The general system theory promotes the idea that a body of tools, predominantly mathematical, can be applied to a variety of apparently dissimilar fields. Thus, it is an appropriate methodology for studying the learning process which has applications in diverse subjects. However, the alternative view of studying the 'differences' between various systems is an equally valid undertaking.

In general system terms, 'adaptation' and 'learning' represent the processes of changing control policies through interaction with the environment. Other, properties such as 'goal-directedness' and 'purposiveness' have also been attributed to these processes of systems.

'Feedback control systems' or 'servo-systems' are a special class of the general adaptive systems, involving the notion of closed-loop control. 'Governors', 'thermostats', and other stability-seeking systems which use feedback have always been considered as important cybernetic models. Some physiological and behavioural sub-systems of living organisms which show simple goal-seeking or homeostatic characteristics can be included within this category. Hence, control theory formalisations have been applied to these biological and even to some sociological systems.

For example, cardio-vascular regulations or particular aspects of economic systems could be modelled quite adequately using servo-system type descriptions, and mathematical techniques such as differential equations. But, the methodology of control system theory, which involves the finding of a mathematical definition of a system, or a search for an appropriate 'transfer function', is not very useful for applying to the more complex biological, human-based or organizational systems; neither it can faithfully represent elementary cybernetic principles.

Similarly, the 'learning and adaptive control systems', discussed at the start of this chapter, are the more complex (and recent) manifestations of the processes of learning and adaptation in general systems.

Many 'natural' scientists have applied the mathematical concepts of general systems theory to the domain of their investigations, and, using a cybernetic outlook, have noticed that many seemingly different systems have various features in common. The central problem for the cybernetician, in such enquiries, is the finding of well-defined behavioural variables and well-defined relations between them.

Ashby, by introducing the concept of 'ultra-stability', described adaptive analogue systems which could change and modify according to different environmental circumstances. Adaptation (deterministically or by trial and error), in organisms and machines, was defined as the controlling of 'essential variables' by forcing them to remain within specified limits - through the manipulation of the environment by outputs. Ultra-stable systems were, thereby, defined as systems that were able to reorganise (adapt) themselves to abrupt changes in their variables and parameters. The abstract models which he devised, and also the hardware model the 'Homeostat', could depict the equilibrium-seeking (also called 'purposive' or 'goal-seeking') behaviour of animals - once the stable-state was achieved, the searching behaviour would stop, the particular path to stability being a function of the state of the system and also the degree of the initial displacements. An organism's nervous system was also looked at as a 'multi-stable' system (a system made up of many ultra-stable systems) which interacted with a complicated multi-variable environment.

The underlying criterion used by Ashby, in general system terms, was that behaviour was considered 'state-determined', or in other words, at any instant of time, the present state of a system could exactly determine the next step.

However, clearly many of our observations in nature are based on probabilistic rather than deterministic criteria. In any case, within the abstract formal framework of systems, Ashby managed to postulate simple axiomatic principles that could lead to more complex and interesting subsequent developments.

Gause (1971,1983) proposes a taxonomy of general systems which considers 'adaptation' and 'intelligence' as two separate independent system attributes. Briefly, a system is said to be adaptive if it is capable of improving with experience; and is considered intelligent if it is capable of performing a given task in human-like manner. The performance of a system is taken to be describable in terms of a real, single-valued function of system states, time, and the task in hand. The attributes of adaptability and intelligence are only considered within the context of specific tasks, and in non-human terms - in an effort to reduce the ambiguity of the definition of such terms. Using these two attributes, a classification of different types of systems is made as follows:-

(a) - 'Innane-Systems': non-adaptive, unintelligent. Having a non varying performance characteristics (e.g., a simple machine or computer programming system).

(b) - 'Meretricious-Systems': non-adaptive, intelligent. With a constant level of performance for a given task (e.g., most chess playing programs).

(c) - 'Homeostatic-Systems': adaptive, intelligent. Performance oscillating within certain bounds (e.g., adaptive process control systems).

(d) - 'Learning-Systems': adaptive, intelligent. Long range improvement in performance as the system interacts with its environment.

Gause further proposes a general blueprint for a cybernetic 'learning system' with the following two principal components: (1) - 'evaluation mechanisms' which perceive the system's performance through goal-directedness or reinforcement; and (2) - 'control mechanisms' which should be able to reduce entropy, maintain variety, and display selective forgetting.

Pask (1967) discusses a cybernetic model of human cognition by viewing man as a special kind of control system. He postulates a set of criteria which characterise such a control system; in particular, the human learning process is analyzed, and the notions developed in terms of general systems theory are applied to experimental learning and problem solving situations.

Reuver (1978,1980) introduces a conceptual tool-kit for the realization of the process of learning in the context of general system theory. The formalisations described is seen appropriate for use in the modelling of real learning situations in psychology, engineering and sociology. A general description of a 'learning system', according to Reuver is:-

> "A learning system is a system purposefully adjusting its behaviour to constant or changing environmental conditions, availing itself of past experiences. Past history of the behaviour in the environmental situation is processed in order to find an optimal adaptation. Information processing is an essential characteristic of the learning process."

The notation used is a more elaborate version of the representations of 'learning systems' outlined in the previous sections dealing with 'learning' control theory and automata theory. The basic elements are:-

T :	$\{..,t_{-1},t_0,t_1,..\}$	- linearly ordered discrete parameter set (time)
X :	$\{x_1,x_2,x_3,...,x_m\}$	- input space (set of inputs)
Y :	$\{y_1,y_2,y_3,...,y_n\}$	- output space (set of outputs)
S :	$\{s_1,s_2,s_3,...,s_j\}$	- state space
U :	$\{u_1,u_2,u_3,...,u_k\}$	- decision space
Z :	$\{z_1,z_2,z_3,...,z_l\}$	- information space
P :	$\{p_1,p_2,p_3,...,p_g\}$	- parameter space
V :	$\{v_1,v_2,v_3,...,v_h\}$	- linearly ordered set of variables, mostly real numbers

Furthermore, a 'learning system' is described in terms of four simpler components of: 'real system' (an unknown or partially known black box); 'system cell' (a mathematical model of the 'real system'); 'decision cell' (a mathematical model of a decision maker); and 'learning cell' (a mathematical model of the learning process) - any 'learning system' should have at least two of the above four distinct components. The goal functions and learning algorithms are also defined in terms of such formalisms.

Finally, various catagories of 'learning systems' are defined according to the different configurations of the four above basic sub-systems; and some specific examples, namely those for 'goal learning', 'learning inventory systems' and 'social learning', are investigated using this conceptual framework.

Reuver (1980) also extends the notions of his formalization to the studies of psychological learning theories of Gal'perin - a Soviet psychologist who had based his theories of behaviour on three distinct level of human action: (1) - material, (2) - verbal, (3) - mental. On such foundations, 'action' is characterised by parameters which specify the 'extent', the 'detail', and the 'efficiency' of behaviour; and any learning is proposed to be directed on a so called 'orientation basis' (the motivation behind the process).

On the whole, this type of 'learning' models, by virtue of their universality, have to contend with an extremely high level of abstraction which itself alienates them from the scopes of the actual applications they are ultimately intended for - by making them incomprehensible or irrelevant.

(ii) - INFORMATION THEORETICAL APPROACH TO CYBERNETIC 'LEARNING' MODELS

Information Theory, which can be considered as a subsection of Cybernetics, has been developing parallel to the latter. These two disciplines are sometimes grouped together, however, information theory is not, generally, involved with control. This discipline, mostly developed in the past three decades, is mainly concerned with various aspects of coding, decoding, and transmission of signals (and information) within communication channels. The basic ideas behind the quantification of the amount of information were founded by Shannon in the late 1940's. But, Shannon's pure quantitative approach to 'information' did not take into account the concept of 'meaning' of acquired information. Today, information theory has more and more come to be concerned with the wider qualitative aspects of data as well as the measurement of various signal capacities.

When the concept of information is used, as a generalised explanatory language, to describe the dynamic behavioural observations of the world, then the distinction between the quantity and the meaning of a message (also the ambiguities of such terminologies) is noted much more clearly. Elstob (1980), in an analysis of concepts of 'information', 'meaning', and 'knowledge', stresses the underlying paradoxes and confusions which exist in the usage and comprehension of such cybernetic ideas. He proposes a distinction between 'physical-information' and 'semantic-information' based on the differences of types of behaviour they are associated with - involving some energy transfer considerations and specific consequential properties of input-outputs. Using this distinction, he also discusses the concept of meaning within the contexts of 'goal-directed' behaviour and 'purposeful systems'. It is argued that to convey the meaning of a message we need to involve the semantic component of information, and also take the behavioural intentionality considerations of a system into account.

Some simple regulatory processes which use error-control or anticipatory cause control are intimately related to the quantitative aspects of information. Ashby (1956) by introducing his law of 'requisite variety' established an explicit basis for these relationships.

Many of the 'information' related findings of the learning process have been accomplished by the engineering oriented cyberneticians who considered the disparities between the capabilities of man and machine. Information entering (or leaving) an adaptive or 'learning system' (natural or artificial) is primarily of the physical type, and hence the quantitative aspects of information theory could be applied at this rudimentary level. At the next level, the manipulations of this raw data in the form of experimentation, constructing of various hypothesis, and a constant feedback from the previously accumulated knowledge and information takes place. Hence, it becomes extremely difficult to apply the quantitative formalisations of information with any degree of cohesion.

Information theoretical concepts developed by Shannon, such as 'information content', 'channel capacity', 'redundancy', and 'entropy' have all been applied to some specific aspects of neuronal functioning and brain mechanisms. For example, in the language of information theory, 'messages' in the nervous-system are thought to be 'encoded' and 'transmitted' between its different elements. Similarly, if a nerve-cell is considered as a source emitting information, then its output is the 'action potential' moving along its axon and regenerated by the subsequent firing of other neurons - hence, information is carried from one cell to other. Although, the use of information theory is sometimes criticised on the grounds that the actual neuro-physiological mechanisms are not themselves clearly understood yet, nevertheless, this methodology has been one of the fundamental approaches in the investigation of memory related aspects of learning.

The usual holistic cybernetic view of the brain is that of the brain as an information processing system, having large memory storage, and operating on coded information derived from its environment and from within itself. This, of course, implies a level of 'analysis' of the brain's physiology which falls well short of detailed neurological scrutinies, but, is more specific than the investigations of electrical wave patterns of the brain (EEGs).

Similarly, in cognitive psychology, a human could be regarded as a channel of communication, if viewed from the information theoretical angle - perceiving and reacting to stimuli. The novelty of this outlook is that it considers all of the set of relevant stimuli at any time, rather than considering individual stimuli for every occasion.

Various aspects of human perceptual capabilities could be studied by using notions of information theory. Broadbent (1965) discusses some areas of such applications. For example, the critical level of distinguishing simple perceptual stimuli is found to be approximately 2.5 bits of information, or 7+2 different signals. Similarly, the understanding of perceptual stimuli, as well as being dependent on the physical noise factors present, are very much context dependent; and probabilities could be attached to the order of occurrence of events, such as the probability of consequential ordering of different physical events and sounds, or the chance of certain words following other words. Estimates of an average human's memory capacity have also been calculated using these notions, but its value ranges diversely between 10^{8}–10^{12} bits of information.

Other quantitative characteristics and capacities of human pattern perception, signal perception, and attention systems have also been investigated. These findings suggest that, in fact, there are some limitations to the human brain's perceptual mechanisms governed by external informational aspects of inputs, and not internal sensory factors.

As far as the output (or the response) side of such a communications system is concerned, similar 'information' based analyses of the brain have been carried out. Some temporal aspects of information processing, such as the rate and the capacity of the brain in reacting to various stimuli, have been experimentally measured. These quantitative results, being more exact and consistent than the ones obtained in the case of perceptual inputs, point to the fact that human beings are also limited, in some fashion, as far as their reactions to stimuli is concerned - this limitation being independent of the information they convey.

The implication of such work on the studies of the learning process is that the various information related quantitative changes which occur in the mappings of stimuli onto responses could be determined as learning progresses. Yet, the causes or the nature of these changes cannot be explained by observations based solely on this paradigm.

(iii) - DYNAMIC PROGRAMMING APPROACH TO CYBERNETIC 'LEARNING SYSTEMS'

The classical mathematical techniques used in describing the behaviour of systems, deterministically or stochastically, most involved the use of differential equations; and optimising problems, generally, involved the minimising of specific functions. The basic assumptions were that the number and the value of variables, the cause-effect relationships, or their probability distributions were known a-priori. However, as the complexity of problems increased, and not all information about a system and its behaviour was known in advance, then the use of adaptive and learning criteria in the solving of problems, such as those encountered in engineering and control, saw the introduction of various mathematical abstractions. Dynamic programming techniques are one of the principal methodologies in such analyses, they are used as an alternate approach to the classical methods in dealing with the modelling of adaptive and learning processes. This subject is closely related to the fields of adaptive and 'learning' control, but has also found applications in psychology, economics, and biology where similar processes are seen to be at work. Dynamic programming, basically, involves the finding of the optimal procedures in a series of observations, working backwards from a prescribed final stage to the first stage.

From a cybernetic engineering point of view, if the various features of a system's characteristics are unknown, then estimations of some parameters could gradually lead to the compilation of a complete knowledge base about the system - by hierarchically improving the precision of system's description. Dynamic programming techniques can be used to formulate an optimal policy for a series of decision choice-stages; the decision could be based on choice of numbers, expected values, or probability distributions.

A highly abstract mathematical formalization of 'learning systems' is proposed by Doberkat (1978) who utilises the view of dynamic programming. He uses a formal framework to show the existence of an optimal strategy in achieving goals. The notions of group and set theory are used to describe a 'learning system' and its processes; and, based on such abstractions, mathematical models of specific examples of learning, such as classical conditioning are simulated.

4.5.7 CYBERNETIC LEARNING MODELS BASED ON THE MECHANISMS OF THE BRAIN

The notion of direct use of the brain's mechanisms as the basis for constructing artificial 'learning systems' was adopted by many researchers prior to the introduction of the science of Cybernetics. However, it was the founding of this formal discipline which allowed the widespread hypothesising of such endeavours, and opened up opportunities for future developments.

The language a cyberentician uses to construct his abstract or physical nervous system only differs in degree from the so called actual observations of a neuro-biologist - since, they both are different types of conceptual descriptions. However, it is hoped that cybernetic models of the functioning of the brain will, also, eventually lead to a better understanding of the neuro-biology and neuro-psychology of the brain.

The neuro-physiological evidence for cybernetic model-builders are very extensive. Animals' purposive activities have been related to various neuronal substrate. Many fixed neuronal mechanisms are thought to govern the biological functions or the reflexive innate modes of behaviour (some purposive behaviours are also controlled by the electrochemical mechanisms of the body such as the hormones).

The brain, seemingly, in a constant state of activity, has constituent neurons which rhythmically discharge, and induce patterns of impulse propagation. The summation of excitatory and inhibitory impulses of neurons when exceed a certain threshold value evoke further patterns within cell assemblies. The progress of the excitation not being limited from the 'sensory' to the 'central' onto the 'motor' outputs, but in a kind of bi-directional reverberations. The generally accepted postulate for the neuronal basis of learning experience states that the flow of messages along particular pathways of neurons leaves some kind of facilitation for the future passage of similar activity along the same neurons. But, as a whole, the knowledge is fragmented, imprecise, and mostly speculative.

In addition, while our knowledge of the cognitive processes of the brain has been constantly improving, the sheer size of the nervous system has meant that the conceptual models (e.g., finite-automata, neural-nets, control-systems, etc.) cannot, in practice or theory, represent the brain's higher functions accurately.

Hence, the current 'learning' models based on some aspects of the brain should all be looked at with a certain degree of skepticism. This point is strongly emphasised by Aleksander (1983) in discussing the difficulties of devising artificial analogues of the brain:-

> "Knowledge of the human brain is currently confined to two fragile bridgeheads bordering on a central gulf of ignorance. On the one side, neuro-physiology can give a superficial picture of the structure of the brain and offer some educated guesses about its operations; on the other, psychology provides a variety of hypothesis about its output. How, exactly, the one give rise to the other remains mysterious."

Griffith (1971) gives an elaborate mathematical account of the functioning of the elements of the brain, and analyses various propagatory aspects of signal transmission within networks of neurons. Furthermore, various randomly connected neuronal models are investigated and mathematical deductions made. Similar type of research, defining a mathematical form of neuronal activity, has been carried out by Caianiello (1967); this work being mainly directed towards devising pattern-recognition 'learning' devices. The implication of such work to the modelling of cybernetic 'learning systems' is that it can give some clues as to the possible ways artificial associative memory systems involving the contiguity of neuronal activity could be devised.

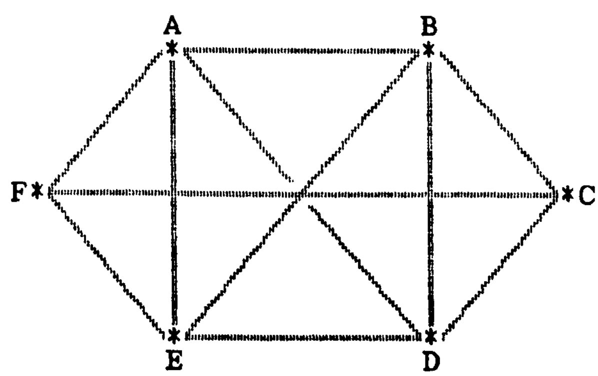
Other hardware (electrical, mechanical, hydraulic, chemical) simulations of the actions of neurons and neuronal systems have been carried out by many researchers. Similarly, computer programs and other software mathematical models have demonstrated some properties of nervous systems. Some servo-system type engineering based models of the brain mechanisms (e.g., Deutsch's, 1967) have also been designed, involving the use of differential equations and transfer functions. The neural and logical nets discussed previously were also examples of cybernetic models of the brain mechanisms.

A simple method of representing nervous systems in an abstract form is the use of matrix notation. Inputs could be depicted by the rows of a matrix and outputs by its columns, as in FIG.4.16(a). An inborn reflex action could be represented by a permanent connection at the junction of the relevant input and output. Other properties such as 'learning', 'conditioning', 'inhibition', 'forgetting', etc. can all be represented by the formation or the modification of the strengths of connections between inputs and outputs. A great advantage of this type of model is that it can be easily realised by various hardware devices (e.g., electronic, magnetic).

A well known early example was Steinbach's (1963) "Lernmatrix" which was based on arrays of magnetic memory cores (similar to those found in early computer memories). Other abstract models (e.g., Griffith's, 1971) have used structural matrices to represent the input-output transformations of the nervous-system.

\OUT- IN-\PUT PUT \	O_1	O_2 , . . .	O_n
I_1	a_{11}	. . . .	a_{1n}
I_2 . . .	. . . .	. . . .	. . . .
I_m	a_{m1}	. . . .	a_{mn}

(a)



(b)

FIGURE 4.16. (a) – A matrix representation of associations between m-inputs and n-outputs; a_{ij} (fixed or changeable) is the associative factor between input I_i and output O_j
(b) – A graphic representation of single associations between pairs of independent inputs.

The associations between any two independent (orthogonal) inputs can also be represented by a matrix notation. Alternately, as shown in FIG.4.16(b), inputs can be graphically represented by points (A,B,...) in a 2-dimensional space with interconnecting lines which depict associations between inputs; furthermore, the strength of association can be shown by the thickness of the lines or their number. Graph-theory has also made some useful contributions to this type of analysis.

An example of cybernetic learning model based on graphs is described by Palvolgyi (1982). He outlines a later version of neural-net type systems which can simulate the processes of adaptation and learning. Such networks are made up of elements that are either in 'active' or 'passive' state, and 'activity' can be transferred to some neighbouring points of the graph according to certain algorithmic rule. The designer contends that this kind of abstract cybernetic approach is a useful tool for studying various basic perceptual processes, memory, learning, and thinking.

Although, there have been many objections and criticisms of these simplistic views of the brain, yet, by in large, most cybernetic brain models are based on simple binary-state elements, as depictions of neurons. The ultimate exactness of these type of information processing (digital or analogue) cellular models of the nervous-system will depend on the clarifications provided by discoveries in the neuro-physiological sciences.

Another cybernetic approach to the modelling of the brain mechanisms of learning, along the same lines as neural-net models of the learning process, is described by Harth (1966). An attempt is made to devise a formalism to depict 'thought processes'. First, the epistemological problems are highlighted, and then a tentative model, heavily reliant on the available neuro-physiological evidence, is proposed. The basis of the model is an artificial abstract neuron, similar to McCulloch and Pitts type element, but, with an additional randomness incorporated. It also differs from Rosenblatt's 'perceptron' (will be discussed later), especially since it is not designed for particular cognitive purposes, but, for the simulation of more general aspects of brain functions. Matrices similar to that of FIG.4.16(a) are used to depict the neuronal connections between the elements of the system; and learning is defined in terms of the changes in the coupling coefficients of such matrices. Finally, a computer simulation of this mathematical model of the brain mechanism is carried out for systems comprising of few elements. Various comparisons and parallels of the results obtained is made with the established physiological data.

The nature of memory is one of the most intriguing aspects of brain studies, and relatively little is known about its actual mechanisms. The establishment of associations between memory traces is one of the brain processes which has attracted the attention of many cybernetic model-builders. Some examples of the modelling of the memory mechanisms were discussed in the previous sections dealing with neural and logical networks - simple models could depict the short and long term characteristics of the memory.

Computer based models of memory mechanisms have been devised which simulate some basic types of associative learning. Generally, inputs are first stored in a short-term memory, and, later, either lost or transferred to a long-term storage; some cognitive aspects of 'retrieval' or 'forgetting' is also, normally, characterised in most of these models.

Drozen (1970) proposes a layered mathematical model of associative memory. He sees the principle contribution of the neuronal network models in the possibilities they provide for filling the gap between neuro-physiological and psychological studies of various aspects of living organisms. Yet, he envisages possible applications to the problems in A.I. His model, in its simplest form, consists of three layers of input, associative, and output elements. These elements can change autonomously, and can produce a 'diffused' (distributed) memory. Again the matrix notation is utilised to characterise this model. The algorithm used for the formation of memory relies on the synaptic weighs of the inputs and outputs - which are considered at discrete time intervals. If some inputs are stimulated simultaneously, a kind of correlation is established between the associated inputs; by adding a kind of short-term (dynamic) memory, a process analogous to the conditioning of reflexes can be simulated, in the form of a simple mathematical realization of associative relations.

Booth (1970) describes a simple model for the organization of memory which is based on the optimal stacking problem of a collection of items (e.g., words). This cybernetic model, also realised in hardware, is an attempt to suggest a possible way of optimising the access time in memory retrieval process, but the existence of equivalent psychological or neuro-physiological correlates have not been conclusively established.

The main developments in the area of associative memory modelling have been directed towards the formation, association, and recognition of 'patterns' in cellular models of neuronal networks. This type of work will be discussed more fully in future sections dealing with self-organization and pattern recognition.

On the whole, cybernetic models of the brain have managed to demonstrate that the so called 'mentallistic' activities can be displayed by machines and other artificial systems. The goal-directed, purposeful, or teleological behaviours which the early behavioural scientists believed to be too subjective, and hence not precisely definable, have, in fact, been simulated artificially. Although, these demonstrations have not actually disproved the mentallistic theories of teleological behaviour, nevertheless, they have managed to broaden the definitions of such properties from the exclusive realms of 'living'.

4.5.8 AN OVERVIEW OF CYBERNETIC APPROACH TO 'LEARNING SYSTEMS'

The science of Cybernetics has had an undeniable impact on the learning process in most areas of research. The immense accumulation of empirical knowledge in many learning related disciplines has necessitated the introduction of cybernetic models, in particular computer based models, for the mathematical analysis and ordering of such knowledge. Cybernetics has drawn special attention to the concepts of 'organization', 'information', 'control', 'feedback', 'stability', 'goal-directedness' within natural systems, and emphasised that no system can be solely specified by its physical descriptions alone.

The interdisciplinary nature of cybernetics has also brought a new vitality into the studies of the learning process; and enabled the scientists from various disciplines to communicate with each other. Hence, increasing the possibility of establishing a unified view of this phenomenon.

During the past 40 years, cyberneticians have been constantly probing the foundations of scientific enquiries and methodologies, and trying to expand the scope of various disciplines by introducing concepts not intrinsically obvious from the empiric of observations. However, cybernetics has evolved into a subject that primarily gives a particular perspective to problems rather than providing specific class of solutions.

Wiener's original goals of devising human-like machines, based on the concepts of information theory and control-theory, and also his suggestion of building analogous mechanisms to the brain which can self-adapt or learn have been, largely, superseded by the exploits of digital computers. On the whole, presently, none of these objectives seem to be apt to the problem of devising intelligent-machinery, or to the furthering of the understanding of human-intelligence. However, with a better insight into the workings of the nervous system, these outlooks might again be favoured at some point in future.

Today, the science of Cybernetics is regarded with different degrees of appreciation and extent of usefulness around the world. In the Eastern Block countries, cybernetics is featured much more extensively in research subjects (mainly computer related and control engineering applications); but cybernetics is, probably, defined in a wider sense than the Western understanding of the science. In the Western European countries, various aspects of this science (philosophical, practical, theoretical, etc.) are still seriously investigated; normally, referring to abstract generalised systems approach to a subject. On the other hand, in the United States, the rise of subjects such as A.I., pattern-recognition, robotics, system-theory, and other more specialised subjects have all but replaced the generalist view of cybernetics in scientific research - the primary feature of these new subjects being their functionalism.

Before moving on to the more recent trends and approaches in the design of artificial 'learning systems', the 'self-organizing systems' (SOS) approach will be discussed next. The term self-organization systems is often used in conjunction with cybernetics or neural-nets, and sometimes these terms are used interchangeably. The SOS approach, however, has managed to arouse enough interest which justifies its appreciation as a distinct discipline, it has also lead to the establishment of a specific viewpoint and terminology. But gradually, as in the case of cybernetics, the enthusiasm towards this subject has also declined. Particularly, in view of the apparent deadlocks reached following some concentrated research in this field, and lack of spectacular results.

4.6 'SELF-ORGANIZING SYSTEMS' APPROACH TO 'LEARNING SYSTEMS'

The notion of self-organization was discussed previously; firstly, as a concept closely related to adaptation and learning; and secondly, in relation to specific views of 'learning systems' and machines, such as the control engineering view.

According to Feigenbaum and Feldman (1963) the 'self-organizing systems' (SOS) approach has an intrinsic fascination, and the basic objective of its workers is the design of intelligent machines which have simple information processing elements, arranged in a random or organised network, and certain processes for facilitating or inhibiting their activity.

Andrew (1972), in a survey of the field of SOS, points to the initial enthusiasm during the 1950's about this subject, and the possibility of it throwing a good deal of light on the workings of the nervous system.

The peak of popularity of the discipline of SOS was perhaps during the early 1960's, when many symposia and conferences were held on this topic, and various researchers presented their work in the context of this newly formed and promising approach.

The main impetus behind this subject has been the desire of learning more about the nature of the nervous system, which clearly has some self-organizing properties; and hence, by direct implication, the possibility of constructing similar artificial SOS.

On the one hand, psychologists, embryologists, neuro-physiologists, sociologists, and workers of other natural sciences such as evolutionary sciences have been trying to understand the self-organizing principles of living and human based systems. On the other hand, mathematicians, engineers, cyberneticians, computer-scientists, and physical scientists have attempted to design systems which show some self-organizing properties.

4.6.1 DEFINITIONS OF SELF-ORGANIZING SYSTEMS

The very definition of a self-organizing system has been met with some controversies. Von-Forester (1959) and also Ashby (1962) have argued that a system, in a strict logical sense, cannot really be self-organizing by itself. The basis of their argument, outlined more formally by Ashby's 'law of requisite variety', is that only order in the environment could result in order or organization in a system. Hence, a system can only become self-organized if it is defined with respect to some external source of order; and therefore the use of this term is only seen to be justified if the interactive component of the accepted system-environment distinction is regarded as an implicit part of the system.

The term "Self-Organizing Systems" was, according to Yovits (1962), first used by Farley and Clark in 1954, and it was defined as: "a system which changes its basic structure as a function of its experience and environment."

But, even in the early 1960's the rise of the more modern trends in 'machine learning' had already raised some questions regarding the feasibility or benefits of using the self-organizing approach in the design of efficient interesting (intelligent) systems. The principal issues involved are argued by Selfridge (1962), and it is contended that a great deal of initial organization (in the form of automated routines or in-built knowledge) is needed to achieve a truly intelligent machine.

Andrew (1972) proposes an additional proviso to the definition of SOS by Yovits, which is to incorporate the property of purposiveness within such a definition. The main difficulties he envisages in devising a rigorous definition include temporal considerations, storage considerations, and the amount of initial organization inherent in the system.

Ashby (1962), also, in a fundamental discussion of the principles of SOS, scrutinises the basic issues in machines and systems, such as: 'organization', 'whole and parts', 'conditionality', 'reducibility', 'communication', and 'good' or 'bad' organization. He gives a definition for two catagories of SOS. First referring to systems which change from 'parts separated' to 'parts joined', with no considerations of system utility. Second definition refers to the change from 'bad' organization to a 'good' one. However, he goes on to qualify this second definition by saying that a system (or machine) can only be self-organizing if it can be considered as coupled to another system (or machine), and in view of this contradiction, he suggests that: "the phrase (self-organization) better allowed to die out."

Another distinction which can be made within the class of SOS is whether the goals of a system are determined externally, such as in self-organizing process-controllers, or if they are evolved from within a system following behavioural interactions with its environment, as in the case of biological evolution.

Andrew (1978) describes SOS in more general, and presently accepted, terms: "If the changes which occur in a system, constituting its adaptation, are sufficiently fundamental, the system is classed as self-organizing."

4.6.2 LEARNING AND SELF-ORGANIZING SYSTEMS

The concept of self-organization, although not treated in isolation, was implicitly advocated and had been present in the adaptive and learning models of many early cyberneticians.

Andrew (1967) discusses two possible general ways by which the property of self-organization could be manifested in a 'learning system'. First method involves the adjusting or adding of parameters to some polynomial control functions. The second method involves placing certain constraints on control actions, by seeking a set of key-points in phase-space and relating the incoming information to these key-points, and trying to maintain the performance within specific bounds.

The learning capabilities displayed by SOS are primarily the result of their interactions with their environments; and the intention of their designers is to emulate, in a sense, the self-organizing features of the nervous system, while choosing alternate simplified abstract or physical means. Hence, the studies of neural plasticity become highly relevant to this approach. In addition, other psychological theories of learning and cognition can also be looked at from the SOS point of view (e.g., Estes, 1959; Rosenblatt, 1959; Farley, 1959).

4.6.3 PRINCIPLES OF SELF-ORGANIZATION

The concept of 'organization' itself is also subject to some ambiguity. But, within the context of self-organization it, generally, refers to the structural organization in relation with the potentialities for action, and not the orderly positioning of the physical constituents of a system, such as in growing crystals. In other words, it is functional considerations that tell us if a system is self-organizing, and not physical considerations - this issue can be traced to the classical distinction between 'form' and 'function'.

Organization of a system, in general, can be separated into the formal organization that is given a-priori to the system, and the organization that is acquired adaptively or otherwise following some changes to its initial form.

Andrew (1970) makes a tentative distinction between 'self-organizing' and 'self-optimising' systems. The former referring to the systems that can change their own internal connections, and the latter being only able to change the value of some system parameters. However, such a distinction is not, according to Andrew, a firm and conclusive one. Hence, a more precise definition for this intuitive principle is deemed to be essential.

The primary assumption about a SOS is that it should start from a fairly low level of organization - but some information about the 'motivational' or 'goal-achievement' aspects could be included in this organization a-priori. A further consideration is whether the system utilises its past inputs in the evaluation of its outputs at any instant of time - and if so, to what extent. If the complete history of inputs is taken into account, then the problem becomes one of classification and discrimination of similar situations. Now, the exhaustive classification of most non-trivial system becomes an enormously difficult task. Hence, as seen in the other previously covered approaches, the focus of problem is switched to designing an efficient 'generalization' technique for the system, so that it is able to make an 'inductive inference'.

Andrew sees a possible way of achieving such generalization is to incorporate the inherent 'continuity' aspects of inputs (for systems that have some or all input-outputs continuous) in their discriminative processes; in other words, to classify and compare patterns of inputs as continuous functions, rather than as points in Cartesian space. Hence, he proposes a procedure for computing output signals in terms of a polynomial function of inputs; and the self-organization of such a system should involve the adjustments of the coefficients of this polynomial by the system itself. His approach to this problem is based on the techniques developed by some workers in self-optimising control systems, who had devised 'learning filters' that operated on a similar basis. The scheme works by evaluating coefficients using a hedony measure, normally in the form of 'error-information'.

The specific algorithm used by Andrew for the adjustment of parameters utilises the well established statistical techniques of 'regression analysis' - which can determine the correlation measures of a set of data points. However, in this case, the estimates are evaluated continuously rather than in discrete steps. Computer simulations of this system are devised and results compared with other similar techniques. Further discussions are also undertaken into the extension of such ideas into the topic of pattern-recognition; and possible methods of analyzing SOS which have no error-information feedback, or are of a more complex multi-layered nature.

Dalenoort (1982) attempts to outline few principles which are seen essential to processes of self-organization. Such definitions are devised to distinguish SOS from systems based on the current computer paradigm (i.e., A.I., cognitive-psychology). Firstly, 'self-organization' is distinguished from 'construction' on the basis of their design particulars - a constructing system, normally, needing a higher degree of incorporated intelligence. Self-organization is said to come about in one of two ways: (a) - by adaptive evolutionary method, (b) - by training a system to internally change its structure or knowledge for future better performance. Other essential considerations are: the trainer should not know the exact nature of the results of his actions; the system should have an internal capability to change its structure - the existence of Ashby's requisite variety is implicitly assumed.

Secondly, to efficiently extract information from an environment it is seen necessary to establish correlations between different variables of the environment (and system). This principle forms one of the basis of the subject of pattern-recognition.

Thirdly, the distinction between 'data' and 'procedure', which currently applies to most computer based systems, should not apply to SOS. The particular development of digital computers has necessitated such a differentiation. However, the rich complexity and the high efficiency of storage and recall of information in the brain can provide many possibilities to manifest 'structural information storage' in SOS. Whereby, incoming information is organised according to the structure of present pathways, and is, in turn, able to interact with and change these structures.

4.6.4 SELF-ORGANIZATION IN CELLULAR NETWORKS OF IDENTICAL ELEMENTS

One of the principal frameworks for the investigation of the property of self-organization has been networks of identical elements based on neural-net type models, similar to those discussed in the earlier parts of this chapter. In the following some distinct aspects of self-organizing networks will be discussed in more detail.

(i) - SELF-ORGANIZATION IN RANDOMLY CONNECTED NETWORKS

As mentioned previously, some workers such as Beurle (1962) and Kauffman (1969) had investigated the formation and the propagation of some kind of organization within randomly connected networks. One of the first attempts in designing self-organizing learning networks was made by Farley and Clark (1954). Of course, numerous McCulloch and Pitt's type mathematical neural models were devised prior to them, but mainly involved the modelling of perceptual mechanisms, and were of a fairly organised pre-fixed structure.

Farley and Clark's (1954) model consisted of a network of non-linear elements, a segment of which is shown in FIG.4.17(a). The elements were connected in a random fashion and each had a threshold value of T_i. When an element 'fired' its threshold rose to infinity and, after a refractory period, exponentially fell back to its normal value. Furthermore, a 'firing' element could transmit excitation to all connected elements for a short period after its excitation. The model, therefore, could show both 'spatial' and 'temporal' summation. The pathways themselves had variable attached 'weight' measures which could determine the efficiency of transmission of excitations between two elements. These networks, unlike logical or neural nets, operated on analogue basis. Another point to note is the lack of inhibitory inputs, which is probably a reflection of the state of neurological knowledge at the time.

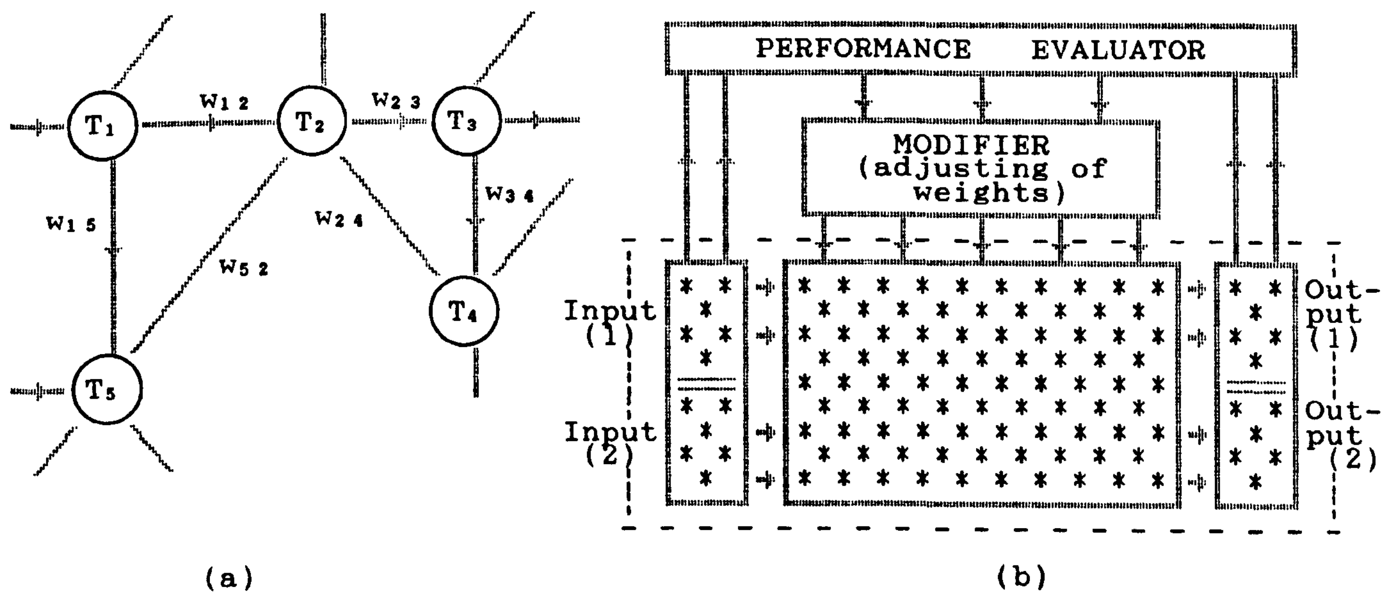


FIGURE 4.17. (a) - A segment of Farley and Clark's self-organizing network shown in more detail. An element fires when the summation of the weighs of its excited inputs exceeds its threshold value.
(b) - A simple two-input two-output learning self-organizing network, showing the arbitrary differentiation of systems elements, denoted by *'s.

A simple process of 'learning' could be displayed by the following exercise - which was also simulated on a digital computer. Arbitrary sets of elements were defined as the two inputs and the two outputs of the model, as illustrated in FIG.4.17(b). The state of the, so called "push-pull", output was determined by the relative number of firing elements in the two output sets. Now, if it was desired for the inputs to drive the outputs in a specific predetermined manner, then an algorithm for the modification of weighting parameters could gradually result in the self-organization of the system into a state which would show the appropriate input-output sequences. The procedure involved increasing the preceding weights (of active pathways) if the change was a favourable one, and decreasing them otherwise.

Beurle (1962) also examines some properties of functional organization in random networks - in an attempt to model some neuro-physiological processes. His devised hypothetical randomly distributed elements are able to make random connections with other cells, propagate excitations and inhibitions, and change their sensitivity during a refractory period. He goes on to discuss the results obtained (various graphic patterns) from simulations of such networks, in relation to some basic neural observations, in particular, the notions of long-term and short-term memory.

(ii) - SOME MATHEMATICAL ASPECTS OF SELF-ORGANIZING NETWORKS

Andrew (1972) describes two general principles which could guide the process of self-organization in multi-layer parameter adjustment-type networks - being distinct and more universal than other earlier self-organizing networks, such as Rosenblatt's (1962) single-layer simple "perceptrons", or Selfridge's (1959) "pandemonium" which could operate in two different layers. The first principle involves the breaking down of the overall goals into 'locally-evaluated' sub-goals, this should be done with particular attention given to the reduction of signal redundancy in the system.

The second suggestion is the incorporation of special pathways in the network which can facilitate a, so called, 'significance feedback' to the elements. This is because, in the more complex networks, the effect of weight adjustments of particular pathways on the output are not known exactly. Hence, a measure of the sensitivity of the output to activity at each point is required to determine parameter adjustments at that point. A simple mathematical algorithm is suggested, and based on such, computer simulations of nets of both neuron-like and continuous elements are devised and their results discussed. Finally, the significance of such notions are considered in relation to the actual neuronal observations, or within the context of socio-economic systems.

Uttly's (1956) work on systems based on 'classification' and 'conditional probability' concepts (discussed previously) has a great deal of relevance to SOS, and has influenced many later developments in this field. Chapman's (1959) self-organizing classification system was an early example of this type of models. It had both of Uttly's classification and conditional probability properties incorporated within its design. However, there were two basic differences. Firstly, the counting of significant inputs was done before classification was achieved, hence the frequency of occurrence determined if an input was classified. Secondly, an arbitrary number was imposed on the sets of combinations of inputs that could be classified. Abstract cell-networks were also devised on such criteria, arranged in configurations which displayed the property of self-organization in an adaptive manner; and were said to 'learn' to recognise patterns in a very trivial sense, using a simple set of rules - the operations of cells involved analogue processes. These models had obvious connotations to the plasticity and growth aspects of the nervous system.

An electronic hardware device was also constructed whose storage pathways consisted of moist cotton fibers, that reduced in conductivity as the moisture evaporated. This simple hardware was found to exactly duplicate the predicted logical workings of the abstract system. However Chapman saw the real potential of these artifacts in the much larger complex models whose behaviour exceeds the range of accurate predictions.

Mathematically these models could be represented by matrices of the form described in the previous section. Links between outputs and inputs could be denoted by the elements of a matrix, and after each input-output sequence the whole of the matrix structure can be modified to take the new reinforcements into account.

MacKay (1962) discusses the problems involved in designing SOS in terms of adjustments of effective topologies for neural-net type systems. He outlines some ways which pathways and connections between elements be modified, on the basis of reinforcement of the 'weight' of causal links, or a change in their 'conditional probabilities'. These processes could be continuous or discrete, and could involve parallel ,as well as serial, operations. He goes on to argue that a much greater efficiency could be achieved if topological organizations of SOS in the time domain could take up important functional roles - unlike the structure of most computing machinery where no particular functional significance is attached to the topological relationships of, for example, its memory cells. The various aspects of storage of information in temporal structures of SOS are also discussed, in terms of hypothetical elements called "coincidence detectors", and compared to some analogous aspects of neural organization.

Justice and Gervinski (1968), in an investigation of possible functional equivalence of the process of biological evolution and self-organization, describe a so called 'SOBLN' (Self-Organizing Binary Logic Network) device, whose functional diagram is shown in FIG.4.18. A network of elements, called 'statistical switches', organises itself on the basis of probability modifications occurring as a result of interaction with environmental variables and a 'goal circuit'.

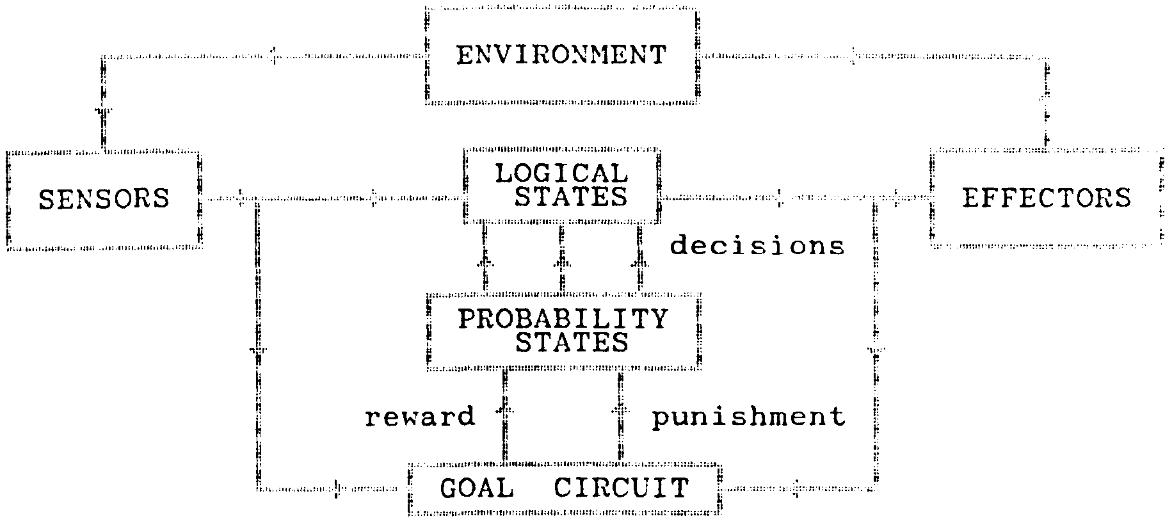


FIGURE 4.18. A schematic diagram of a 'SOBLN' machine.

Th design of this type of self-organizing networks, based on elements and procedures defined in some mathematical or logical terms, has been one of the principal trends in SOS. However, the consistency and the strength of these formalisms, and their theoretical analyses, have not managed to overshadow the weaknesses of their basic criteria and the deficiencies of their underlying premises.

(iii) - SELF-ORGANIZING NETWORKS BASED ON SOME ASPECTS OF THE NERVOUS SYSTEM

A belief in the parsimony of nature will imply that the problem of the organization of stimuli (and information) in the nervous system has been solved in the most efficient way, minimising all forms of redundancy.

Information in the nervous system is compressed, generalised, classified and organised according to processes which are very much in the dark. Uttly's classification system was a simple hypothesis about such processes, and he suggested the possibility of finding similar mechanisms in the brain.

Some principal developments of the topic of pattern recognition, to be discussed in the next chapter, are a direct consequence of the work carried out on neuron-like self-organizing networks.

The major thought provoking landmarks in this area were the work of Hebb and, later on, of Milner on the organization of neural assemblies. Their neurologically oriented theories were also complemented by the work of cognitive psychologists, mainly from the gestalt school, on the various aspects of perception.

Later manifestation of the property of self-organization within networks of neuron-like units have been undertaken by some engineering oriented workers. The type of cells suggested are, normally, the threshold digital elements; and the tasks involve some form of learning of external patterns or sequences of external patterns. This class of investigations are closely related to 'associative memory', 'recall', and 'concept formation' studies.

Threshold elements, as non-linear models of nerve-cells, have been the subject of many modelling endeavours which deploy their information-processing capabilities. But, their non-linear feature has meant that they are difficult to analyze in very general terms. Amari (1972) discusses self-organizing networks of threshold elements in relation to learning and recalling of patterns, and sequences of patterns - again, based on Hebb's hypothesis on the neural organization during a learning process. Amari uses the criterion of stability to define equilibrium states or state-transitions for the network, and the system is said to learn from its repeated encounters with stimuli, hence self-organize itself into stable states. Once a pattern is remembered as an equilibrium state, then it can be recalled and reproduced by an associated stimulus. Similarly, if a sequence of patterns is remembered, a segment of this pattern can act as a 'cue' stimulus and recall the rest of the sequence. Furthermore, it is shown that a kind of generalization of patterns can be achieved, by forming a representative pattern for a given set (class) of stimuli.

The analysis are all carried out theoretically. The formalisms devised involve defining a mathematical network of elements (having weighting and threshold values), and also investigating the stability of various states, state-transitions, and state-transition sequences. The effect of 'noisy' disturbances on the learning of patterns is also analyzed. Briefly, the basic algorithms used for modification entail the increasing of weighting value of connections of inputs and outputs (or other inputs) that coincide, and the decreasing of weights if they differ.

Fukushima and Miyake (1978) also describe a self-organizing neural network model of the functioning of human associative memory which, according to its designers, show some of the self-organizing properties of neural plasticity. It is, principally, based on the hypothesis that reinforcements of the pathways between two cells occur when the high states of their activities coincide. Their previously conceived model, the "Cognitron", was a multi-layered network, it was also simulated on a computer. The cognitrons were able to selectively respond to frequently occurring patterns, after a number of stimulus presentations. This later development of the cognitron model, the 'feedback-type cognitron', included the additional feedback feature - by making connections between the last-layer cells and some of the front-layer cells. A mathematical formal model was described for this new network, and its computer simulation undertaken. Because of the feedback loop the stimulus, after their presentation to the system, continued to recirculate within the network (even though the inputs were removed). Hence, after a period of training with different patterns, the connections are self-organized into desired states, depending on the characteristics of the externally presented stimulus patterns. The results of simulations showed that these self-organized associations could be recalled even if fragmented, noisy, random, or difficult stimulus patterns were presented to the system - on the basis of best correspondence with memorised patterns.

4.6.5 INFORMATION-THEORETICAL APPROACH TO SELF-ORGANIZING SYSTEMS

The concepts of information theory have been used to describe the basic criteria involved in SOS. For example, it has been established (by Ashby and Von-Forester) that the information theoretic concept of 'uncertainty' (or entropy) of information-content, or the variety, of a self-organizing system should be greater or equal to the uncertainty of the environment it is trying to control (or the uncertainty of disturbances to its goals) - as a necessary condition of achieving organization.

Similarly, the 'certainty' of a self-organizing system should increase as the system becomes more organised and the probability of response to a particular input increase. However, as certainty increases the system becomes more redundant, and, hence, the 'redundancy' of the self-organizing system should increase with time. Therefore, a specific condition can be imposed on the rate of change of the 'entropy' functions of SOS.

Von-Forester (1959) proposed, in a strict logical sense, the thesis that: "There are no such things as self-organizing systems;" and went on to justify this postulate on the basis of thermodynamic considerations of entropy within finite universes that contained the SOS. Other philosophical issues in relation to this subject were also scrutinised, in particular, the conceptions of order and reality.

He, further, proposed a particular measure for 'order' involving Shannon's 'redundancy' notions of information. Redundancy is defined by:-

$$R = 1 - H/H_m$$

where H/H_m is the ratio (or the relative entropy) of the entropy (in information theoretic sense) to the maximum entropy (or the maximum disorder) of a system.

Hence, if $H = H_m$, or in other words the system is in its highest disorder then $R = 0$. But, if the elements of the system are arranged such that given one element the position of the rest are precisely determined, in other words, there is no uncertainty, then H becomes zero, and $R = 1$, implying perfect order.

Now, the criterion for self-organization becomes:-

$$\frac{dR}{dt} > 0 ,$$

meaning that the rate of change of redundancy should be positive; and, it can be deduced that:-

$$H\left(\frac{dH_m}{dt}\right) > H_m\left(\frac{dH}{dt}\right) \qquad (1)$$

The two special cases of equation (1) are:-

Firstly, when the maximum entropy (H_m) is constant, then:-

$$\frac{dH}{dt} < 0 ,$$

meaning that the value of entropy should be reducing with time.

Secondly, if the entropy (H) is constant, then:-

$\frac{dH_m}{dt} > 0$,meaning that the maximum entropy should increase with time;

hence, implying that additional elements should be added to the system to achieve self-organization.

Von-Forester also suggests two general mechanisms which lie at the core of understanding of SOS; namely, the principles of "order from order" and "order from noise".

The general information theoretic aspects of self-organization can be applied to fairly simple combinatorial type systems. An example is the models described by Rapoport (1962). He discusses two sets of experiments in three-person learning groups, on the basis of their self-organizing properties. The, so called, "parameters of self-organization" are singled out as a measure of the performance of the system - indicating the degree to which the system had managed to organise itself. The learning tasks are comprised of finding a sequence of associations, or optimising a reward. The empirical observations are supplemented by some information theoretic analysis of a mathematical model, and the analytical predictions are compared with the actual results. Some indices and parameters are extrapolated which can quantitatively determine the learning efficiency of the group. Once such parameters are established, then their relation with respect to the composition of various groups, environmental conditions, experimental parameters, collusion or communication between the members of groups, and other controllable or observable variables could be studied.

Pask (1967) applies Von-Forester's formalised information-theoretic notions on self-organization to his elaborate cognitive model of human learning, and discusses the consequences of such analysis with the help of computer simulations. His ideas are extended to the domain of teaching and education. Man, in a general systems context, is also regarded as a self-organizing system, interacting and learning from its environment. Even a more intricate and general model of 'inter-nation' politics and relations has been studied and simulated by Guetzkow (1962), using the context of SOS.

4.6.6 GENERAL SYSTEMS THEORY AND SELF-ORGANIZING SYSTEMS

There are many different, and sometimes conflicting, conceptions of SOS which mainly refer to a specific class of system. Mesarovic (1962) analyses the SOS from a broad outlook of general systems theory. The process of self-organization is seen as an orderly change in a system's structure. In addition, some high level examples of SOS in nature (e.g., the human brain) are only considered as a special type of the general class of all SOS. The interdisciplinary feature is seen as a very important aspect of such analysis, but it is emphasised that widespread careless generalizations should be avoided, and: "SOS defined in terms of the activities or behaviour of the general system and not in terms of the specifics of the system under consideration."

Mathematical optimization techniques, on the whole, cannot be used on the general class of system structures that have no orderly relationships, and mainly apply to functional problems. Hence, Mesarovic outlines an abstract basis for the description of systems that can accommodate the global view of the property of self-organization - enabling the identification of a behaviour of a system as self-organizing, distinct from other properties such as adaptation and learning.

He also distinguishes two general methods of describing the behaviour of SOS: (1) - 'causal descriptions', and (2) - 'teleological descriptions'. The causal SOS change their structure according to a preprogrammed set of input-output relationships, but the teleological SOS change their structure in an attempt to achieve (or pursue) certain goals. The latter systems are the more flexible, and hence can function in a wider range of situations. The essential question is, therefore, seen as the design of elaborate and complex decision-making procedures and strategies to attain a goal more effectively, which might involve the changing of the teleological structure itself. Mathematical description of such systems are abstracted, and their behaviour further investigated by studying the computer simulation of analytic models.

4.6.7 AN OVERVIEW OF SELF-ORGANIZING SYSTEMS

Andrew (1967) sees the way forward for the future generations of much more complex computers lying in the incorporation of self-organizing properties within the fabric of computers; and he anticipates that basic units of these computers to be of a more complex nature than McCulloch and Pitt's models of neurons or the threshold logic elements. Furthermore, in this vain, he discusses a set of features that all SOS are likely to have, and proposes some hypothetical elements, namely "wondering correlators", which either dynamically or statically search for various significant levels of correlation in the system and then perform particular functions. The shortcomings of available experimental models are also discussed, in particular, the lack of a hardware device for adequate representation of analogue storage.

However, these cybernetically biased views, although intuitive, have not produced any concrete results in practice, thus far. Similarly, abstractions emerging from the topic of SOS have not lead to any fruitful discoveries about the nervous system. Probably, because of the complex way the inherited (genetic) and the environmental sources of order interact and influence the organization of the nervous system. Nevertheless, some workers feel that this approach is still potentially fruitful.

Furthermore, no precise global theories have been devised for the accurate functional analysis of this criterion. The formalisms proposed either deal with this phenomenon in a very abstract and general sense, or are only applicable to a specific type or a limited class of systems.

More recently, Andrew (1980) (also Dalenoort, 1982) reviews the development of the subject of SOS, and its current status, indicating the recent apparent abandoning of such an approach is, mainly, due to lack of disappointing results (unlike the spectacular achievements of A.I.). The SOS are also discussed in the context of 'autopoiesis' (referred to in the previous chapter), and the autopoietic view is regarded as a possible means of reviving the interest in SOS studies.

CHAPTER 5

APPROACHES TO MODELLING OF LEARNING – PART-II

5.0 INTRODUCTION

This chapter is in fact a continuation of the preceding chapter; a fairly arbitrary division is made here to distinguish the more recent trends in the modelling of the learning process from the earlier trends labelled by a variety of terms such as 'cybernetic', 'self-organising systems', 'neural-net', 'automata-theory', 'control systems', etc. (those outlined in chapter four).

However, the taxonomy pursued in these two chapters does not follow a precise historical ordering, neither does it have clear cut boundaries. Therefore, many of the subjects discussed in this classification span the entire course of the evolution of the field of artificial 'learning' systems, in one form or other. The overlap of domains of research for many of these approaches has also lead to a complex inter-linking of interests of workers from different fields.

Nevertheless, it is fair to say that, generally, most of the topics covered in Chapter 4 reached the peak of their popularity at some point during the past four decades, and, on the whole, have had a historical precedence over those that will be covered in this chapter (which are still extensively and progressively researched). Two major approaches to the modelling of learning will be discussed here: 'Pattern-Recognition' and 'Artificial-Intelligence' (specifically some of its important subsections: 'problem-solving', 'concept-learning', 'game-playing' and 'robotics'). In addition, some other specialised approaches to learning such as the 'evolutionary-programming', and the 'educational' approaches will also be briefly considered.

5.1 PATTERN-RECOGNITION APPROACH TO 'EARNING' SYSTEMS

The subject of "Pattern-Recognition" or PR has been referred to previously in numerous occasions. But, here, we will attempt to outline a more detailed and broad account of this topic, and issues involved in relevance to the learning process. Many of the topics discussed in the previous sections on SOS and neural/logical nets have had a great deal of bearing on PR models; and some of the models described could indeed be categorised as PR models.

Firstly, it must be emphasised that PR does not solely refer to visual aspects, and more fundamental explicit qualities of percepts and their relationships are involved. The process of PR is also so intrinsically involved with the process of learning that it is hard to imagine any learning situation where the detection, recognition, or classification of input variances are not prime concerns.

The discipline of PR came to life as a result of the scientific desire to find physical counterparts of physiological and mental events. Yet, the diversification of its enquiries has even extended to philosophical domains, such as discussions about the nature and validity of various epistemological assumptions on perception - usually involving aspects of mind-body problem.

The science of PR has two principal facets; on the one hand, models are devised for the 'simulation' of relevant psychological properties; on the other hand, models are devised as 'artificial' counterparts for such properties. Such a dichotomy was also implicit in our previous approaches, however, it is only in the case of these more recent trends that the division has become explicit.

The above, apparently arbitrary distinction is hinged on the underlying intentions and backgrounds of scientific workers interested in the subject. If their objective is the understanding of the 'natural' aspects of cognition by a step-by-step imitation of perceptual observations, then they are said to be engaged in simulation. But, if they are occupied with the design of models or machines that try to achieve the task of PR by, principally, non-biological means (although some 'natural' considerations may be incorporated), or with the intention of out-performing humans in particular task-domains, then they are said to be involved in the artificial aspects of PR. However, so far, the majority of PR models designed have involved practical tasks of discriminating patterns by "artificial" means, and not simulating the way an animal or human performs the same task.

A survey of literature on PR reveals the variety of abstract and practical techniques that have been adopted in solving these problems. The engineering-oriented system theorists have been engaged in the design of efficient input-filters for improving a specific performance criterion of a system. The workers interested in machine and computer 'intelligence' have been trying to devise models that can display some human perceptual capabilities, but not necessarily following the same processes. While, scientists that have an interest in either simulation or synthesis of animal and human PR mechanisms have endeavoured to construct a variety of models which emulate some such properties.

The science of PR has, also, been distinguished from the more recent topic of artificial vision (or scene-analysis) by some workers. Artificial vision is said to be concerned with structural properties and relationships of idealised object patterns, while, PR is concerned with identification from raw data. Yet, ultimately, we could argue that both processes should be incorporated hand in hand in the true intelligent machines of future.

Similarly, various researchers have distinguished PR problems from 'pattern-classification' and 'pattern-formation' problems. It is contended that in PR a set of known classes of patterns normally leads to a classification rule. While, pattern-classification is only concerned with applying a previously established classifying rule to some unknown inputs; and pattern-formation refers to problems dealing with the identification and definition of appropriate class sets.

However, here, neither of the above two distinctions will be pursued, and pattern-recognition and its related processes will be discussed in a broader context. Although it must be remembered that problems posed in PR are quite different from the engineering type problems involved in the design of automatic character or symbol 'reading' machines, where the principal issues of concern are the coding of a set of standardised patterns.

The development of the science of PR has been land marked by specific solutions that, although spectacular in well defined small domains, are mostly dependant on ad-hoc practical methods. On the other hand, the more global solutions to the problem are characterised by mathematical formalisms which, seemingly, preoccupy their designers with the intricacies of their abstractions without paying too much attention to practical aspects. Furthermore, experience has shown that when the first type, the ad-hoc, techniques are extended to the more generalised situations, they are disappointingly inadequate.

A classification for the different types of intelligent activities modeled by PR researchers was proposed by Dreyfus (1972), which, in his view, distinguished, although not sharply, the four general areas most PR workers direct their efforts towards. These fields are: (1) - 'pattern matching' rigid responses to fixed templates; (2) - 'algorithmic rule based recognition' of simple rigid patterns, involving search of feature classes; (3) - 'heuristic recognition' of more complex, or noisy, patterns which involve search for regularities; (4) - 'recognition' of varied and distorted patterns, based on generalising stimuli, and by having some insight into the meaning or the relevance of patterns - which usually involves learning from experience or examples.

Within the previous three decades, numerous PR systems have been developed, tackling varied tasks such as:-

(a) - Reading of codes, symbols or characters
(b) - Recognition of various objects
(c) - Speech-typewriting
(d) - Voice-recognition, speaker-identification
(e) - Identification of radar, sonar, or EEG signals
(f) - Weather-map analysis
(g) - Finger-print identification
(h) - Chemical analysis of stained tissue slides, blood-cell classification
(i) - Medical diagnosis
(j) - Game-Playing
(k) - Automatic translation and programming
(l) - Industrial process and quality control

The degree of proficiency of a system depending on the number and the complexity of patterns in its recognition range, and also on the ease of detecting and defining features of such patterns.

Although, most research in PR has been directed towards the recognition of visual patterns, a significant amount of work has also been progressing in the area of 'speech-recognition'. Early works on speech recognition approached the problem by trying to first recognise words on the basis of their phonetic patterns, and then, by using some rules of grammar, recognise whole sentences, and finally the embedded meanings. However, it has become clear that much more complex interactive processes are involved, whereby, recognition is influenced by a variety of semantic and syntactic factors, and also environmental cues and clues in a holistic manner. Currently it is regarded that the progression of methods developed for simpler type lower level sub-problems (phonetic recognition techniques) is not a viable route for finding a solution to the problem of artificial speech recognition.

There are also certain types of PR processes which do not depend on a specific sensory modality (vision, speech, etc.), such as the recognition of symptom patterns in a medical diagnosis system, or some geophysical patterns in a weather forecasting system. However, the fundamental problems of speech recognition (or other types of recognition), which are of interest to this thesis, will be covered within the general discussions of the topic of PR in this section.

5.1.1 SIGNIFICANCE OF PATTERN

Some early workers in cybernetics had reached the conclusion that recognition of patterns was the critical operation in any process of learning. For example, Walter (1953) embarks on an interesting discussion of 'pattern', and the important part it plays in every facet of life and development of scientific enquiry. Pattern, as the 'raw material of order', is defined as: "any sequence of events in time, or any set of objects in space, distinguishable from or comparable with another sequence or set." Its significance as an intrinsic quality of living is highlighted; and all sciences, in a broad sense, are said to stem from 'pattern-seeking'. The developed capabilities of the nervous system are also seen as the solution provided by evolution in dealing with the problem of perceiving complex patterns.

A more engineering-oriented definition for 'pattern', given by Glorioso (1975) as distinct from 'signal', is: "If one has no convenient mathematical description of the information structures of signals, then they are generally referred to as patterns. Also, patterns often take on a multidimensional nature and are expressed as n-dimensional vectors."

The significance of 'pattern' has also been evident in investigations of the workings of the brain at the neural level - it is the pattern of inputs to a neuron which determines its 'firing', and not the excitation/inhibition of individual inputs.

Hence, the questions of how pattern is formed, classified, preserved, retrieved, changed, and what it signifies, become central to the investigations and modelling of the learning process. In Walter's case, he approached these questions by trying to analyze and model the patterns of holistic activities of the brain - in the form of EEG recordings of the brain wave patterns.

5.1.2 PATTERN-RECOGNITION AND PHYSIOLOGICAL PSYCHOLOGY

Evidence from psycho-physiological observations show that animals and humans can recognise shapes independent of their size, angle, mirror-image, distance, movement, brightness, colour, incompleteness, and their relative retinal location. The less developed organisms possess inferior perceptual capabilities, due to the more primitive central processing features of their nervous systems, and not the deficiencies of their visual systems. In humans, visual pattern-recognition and image-processing is carried out by very fast parallel sensors (estimated at 10^6) that transmit information from the eye to the brain - this process is in contrast to the slower and more limited sequential auditory signal processing achieved by the ear. The visual signals are further manipulated in the brain by utilising its huge data base of information.

Some workers have attempted to devise artificial PR models which manifest a variety of psycho-physiological observations. As in the case of earlier neural-net and SOS trends, the theories of learning introduced by Hebb (1949) have featured prominently in many such endeavours.

Psychologists have, in general, divided the process of recognition into: 'primary-recognition', referring to the process of recording physical attributes of an object; and 'secondary-recognition', referring to the interpretation rules imposed or the classifications of sensory data. Furthermore, other higher aspects of perception such as 'expectations' and 'beliefs' are assumed to arise from these two lower levels. It is also recognised that various motivational factors influence the process of PR, and that motivational centres are intricately connected with PR mechanisms - although, most artificial PR models do not encompass motivational aspects.

The difficulties encountered in devising artificial PR systems are better understood when we carefully consider the natural PR systems. Clearly, the biological mechanisms present in the nervous system cater for this facility of animals, and in most cases no conscious or direct effort is made to classify or recognise an object as a member of a class of similar objects. However, the complex interplay between experience, learning and PR (and many other contextual and linguistic factors) has hampered the discovery of exact postulates about the nature of this process, or the mechanisms involved. Furthermore, due to the particular nature of most psychological experiments carried out on PR, which are generally based on logico-linguistic criteria, no real insight can be obtained into the actual way physical inputs are processed or utilised within the brain.

As far as a direct connection between the processes of PR and learning in the psycho-physiological sense is concerned, perhaps only 'habituation' can be described in terms of a crude process of PR. Whereby, an internal representation or a model of a stimulus pattern (or its general class) is formed and stored within the memory, and subsequent input patterns compared with it. Hence, in this sense, we can say that PR is only akin to the primitive levels of the evolutionary hierarchy of the learning process. However, PR is today being used as basis for a wide range of 'learning' models, many of which involve the more complex and higher strata of this hierarchy - 'learning' of meanings or concepts.

5.1.3 BRAIN, NEURO-PHYSIOLOGY, AND PATTERN-RECOGNITION

The lack of precise knowledge about the workings of the brain, and the complex nature of problems associated with its analysis, has prompted many scientists from various disciplines to apply the conjectures of their own paradigms to the modelling and explaining of the functions of the brain. Pattern-Recognition is one such discipline; an initial biological bias culminating in models which are more or less non-physiologically oriented has been the feature of most endeavours of PR workers. There has been a striving to find engineering or mathematical analogues of natural PR processes, and also a desire to explain these natural processes in terms of some engineering or abstract criteria.

Some engineering-oriented workers, having this latter objective, consider the whole brain as a kind of modular pattern recognizer. The brain in their view functions by extracting meaningful patterns of information from the mass of incoming data, compares these patterns with previously stored patterns, and based on the past and present patterns makes intricate judgments on which internal or external outputs to activate. Consequently, this simple explanation of biological reaction mechanism is extended to the entire hierarchy of intellectual capabilities. Such scientists see PR to be of prime importance to all biological organisms, and contend that it is only the size or the complexity of the memory-storage, comparing, or decision mechanisms of the PR systems in species which determine the phylogenetic ordering of intelligence.

At the neuro-physiological level, also, many workers such as McCulloch and Pitts (1959) have tried to uncover the mechanisms of PR and perception. They have directed most of their research towards the analysis of neural information processing, in particular, in the visual systems of humans and animals (e.g., frogs, cats, monkeys). These studies have been a valuable source of information for PR model-builders. For example, it has been discovered that certain retinal detectors of these animals are dedicated to preprocessing and responding to special types of visual features (e.g., edges, contrasts, convexities), more or less functioning on a sort of 'template matching' criteria.

Rosenblatt's (1959) 'Perception' machine, constructed with photocells and lights (inputs and outputs), was also able to display some basic qualities of human vision, and was even able to make simple classifications of inputs.

Many other anatomical studies of simpler visual systems have managed to reveal some of the mechanisms that extract more basic features involved in transforming visual information from the eye to the higher processing levels (as well as showing some general principles of neuronal plasticity and specificity). Subsequently, most of these concepts, in a more simplistic form, have been appearing in various PR schemes.

In addition, numerous engineering oriented scientists have attempted to apply various hypothetical neuronal models to the problems associated with PR. Using analytical neuronal models they have designed systems to emulate the 'feature extraction' or the 'stimulus discrimination' mechanisms of the brain of some animals.

Examples are Deutsch's (1967) models of the visual and auditory PR mechanisms of animals, which use established engineering techniques such as transformation matrices or functions, information theory, Laplace Transform, frequency analysis, etc. Deutsch's line of investigation firstly involved the understanding of biological processes at work, followed by the application of known engineering methods in analyzing these processes, and finally the introduction of a simple hypothetical model, based on arrays of elements represented by matrices. These models were capable of recognising simple patterns, such as the decimal numbers - the recognition was done by comparing and correlating features of inputs with sets of test patterns.

Arbib (1972) also approaches the problem of PR from the brain sciences point of view. He embarks on his line of argument by first considering the various hypothesis and discoveries made in the physiological investigations of visual perception and visual recognition in animals, in particular that of cats and frogs. His principal contention is that perception is 'action oriented', and goes on to develop analogous artificial models for some perceptual mechanisms.

5.1.4 COGNITION, PERCEPTION AND PATTERN-RECOGNITION

The vast amount of research and experiments carried out by cognitive psychologists in the field of 'perception' has provided an important source of knowledge for PR workers engaged with such problems. Discoveries, mainly, based on human perception have suggested ways which information could be classified, compared, generalised, organised, stored, retrieved, and in general perceived. Much of the general principles used in the construction of PR computer models have been adopted from these studies, especially from research on visual-perception - of course, other principles adopted from various empirical studies have also been incorporated in such models. Ideas on 'perceptual cues' and 'clues', and formation of 'belief-structures', are some of the principal notions incorporated in the work of PR model-builders.

Simply stated, to 'perceive', according to George (1973), is something more than 'to sense' and less than 'to know'; and while, the defined boundaries of terms 'perception', 'cognition', and 'learning' are to some degree arbitrary, such a segmentation has helped the investigation and analysis of behaviour. George (1986) emphasises that these terms should not be viewed as pieces of a "jigsaw puzzle" that fit together, but as concepts which overlap in a somewhat "fuzzy" way.

The distinction between psychological concepts of 'cognition' and 'perception' is vaguely reflected in another dichotomy, "parallel/sequential", in PR. If the trivial classification of patterns by, primarily, using information from the sensory level, the "cognition", is the objective, then the 'parallel' methods of PR could provide the appropriate examples. By contrast, if the more complex logical and deductive (knowledge level) aspect, the "perception", is to be emphasised, then the 'sequential' procedures of PR, mainly developed in the field of A.I., could be cited. However, it is hard to imagine the complete PR systems of future solely relying on only one of the above paradigms.

The close analysis of even the simplest recognition tasks carried out by man, and the identification of parameters and components of perceptual processes involved, reveals the daunting difficulty of trying to implement PR in machines, computers or robots.

Some perceptual laws of organization were outlined and discussed in chapter two. Human and animal perceptual systems show many peculiarities which have been extensively demonstrated by, the so called, perceptual illusions. Examples, such as those illustrated by Gregory (1966) or Lindsay and Norman (1972), clearly show that sensory information could be interpreted in erroneous ways. 'Degradation of images', 'competing organizations', 'organizations without meaning', 'motion paradoxes', 'sensory aftereffects', 'spatial and temporal illusions', 'impossible organizations', and 'context dependant illusions', are some of the labels used for particular idiosyncrasies of human visual system - which are, in fact, manifestations of implicit rules of perception embedded within the neural structures. The main question for a PR worker is that to what extent these issues are significant to the core of the PR problem. This is important since direct physical observations of the mechanisms involved will not, in general, account for or predict the oddities observed in perceptual systems.

In spite of a lack of precise explanatory hypothesis, there are strong evidence which point to the existence of some 'feature extracting' mechanisms in the brain; and, also, there are indications pointing towards the importance of issues such as 'visual cues', 'context', 'expectations', 'meaning', and 'attention' in perceptual recognition processes.

Recently, some solid theories have been developed in the field of perception, be it in a narrow sense, in an attempt to explain a number of above peculiarities and distortions of the human visual and perceptual system. For example, rules have been proposed which based on simple notions of convexity, concavity, or relation of faces, edges, and corners of a picture could decide whether the object is physically "possible" in configuration or not. Consequently, these theories are finding their way to the design of various computer programs dedicated to scene-analysis or PR.

Cognitive studies of the process of PR have, therefore, shown that some interpretations must be made on the sensory data reaching the brain. The rules of these interpretations have been modelled by some computer and A.I. workers. One class of models analyze scenes of group of objects or shapes. For example, Guzman's (1968) model or Winston's (1977) 'block-world' model. Another class of models try to simulate human cognitive processes, and provide a medium for testing perceptual observations, an example being Simon and Feigenbaum's 'Elementary Perceiver and Memorizer' (EPAM). Generally, such models are characterised by their dual processing levels; first, an elementary specific feature extraction level; followed by a more global concept forming level, in a vaguely analogous manner to the natural PR processes.

5.1.5 INTELLIGENT MACHINES AND PATTERN-RECOGNITION

PR has long been considered as one of the central problems in the design of intelligent machines and programs. A prime objective of designers of intelligent machines has been to give machines the capability to communicate with the outside world, more or less as humans do. Hence, implying the need to design vision systems that can 'see' and 'comprehend' the environment without the use of specialised intermediating switches, keyboards, or transducers.

As explained, PR encompasses a multitude of issues such as perception, cognition, and abstraction. This important consideration was acknowledged right from the outset of development of this subject, as stated by Selfridge and Neisser (1960): "...until programs to perceive patterns can be developed, achievements in mechanical problem-solving will remain isolated technical triumphs."

Perhaps, the impressive array of competent yet very specialised models developed in the computer related and A.I. subjects thus far reflects the above early skepticism and concern. Many indications point to an imbalance between the developments of 'acquisitional' and 'processing' aspects of intelligent machinery. The net result being that most developed systems only function proficiently in specialised domains with carefully prepared inputs.

Aleksander (1984) outlines three broad classes for recognition systems developed for use in intelligent systems or machines. Briefly, they are:-

(a) - 'Preprogrammed systems': that detect geometric properties (e.g., area, perimeter, etc.) of objects, and, hence, calculate their identity. Standardization and simplification of images is of prime importance, therefore, great attention is given to smoothing and enhancing the quality of images (from TV-cameras). High speeds of processing are required for this class of systems.

(b) - 'Adaptive systems': which incorporate more versatility (but less expertise) into their designs. These systems can, either under the external supervision of an operator or acting autonomously, interact with their environments; and, by 'seeing' samples of objects to be recognised, are able to make a recognition choice when confronted with an unknown object.

(c) - 'Intelligent systems': which can be attributed with some higher intellectual capabilities of humans, such as language and reasoning. The emphases in this class of models have been on the formation of knowledge structures (logical, linguistic), and on the representational aspects of 'concepts', rather than physical 'images'.

Aleksander also discusses how these three classes of models, having found their specialised applications in different areas of PR, A.I., expert-system, etc., could be combined and utilised in the development of 'intelligent automation'.

5.1.6 ARTIFICIAL-INTELLIGENCE AND PATTERN-RECOGNITION

PR is one of the principal pillars of the science of 'artificial intelligence', and its overlaps with logic and language, specially in the realm of perceptual models, have been some of the central issues in A.I. research.

PR is generally regarded as one of the four principal subdivisions of the field of A.I. - the other three being 'game-playing', 'language understanding', and 'problem-solving'. The importance of PR within the science of A.I. is due to the fact that most developments in the latter subject (e.g., problem-solving, concept-learning, theorem-proving, language-understanding, etc.) have to deal with some aspects of recognition of patterns; and it is implicitly assumed that any truly "intelligent" system should have the important capability of PR as a prerequisite of its operation.

Some examples of conceptual PR models developed by computer-oriented A.I. workers are those which are able to detect vertices, lines, and surfaces of input patterns, and identify the structural features of different geometric forms. In other cases the programs are based on detecting 'resemblances', and not absolute 'matching' of patterns to the originals. In addition, A.I., computer vision and PR models which concentrate on the higher level analysis of feature spaces have also served as frameworks for testing various conceptual hypothesis and issues on problem-solving or learning.

At the core of the field of PR is the process of 'recognition'. Recognition, according to Raphael (1976), is one of the two principle approaches involved in 'problem-solving' - the second being 'derivation'. This distinction (arbitrary and non-fundamental in some sense), in general, hinges on the relative emphasis given to 'search' and 'logical deduction' aspects involved. Examples of each type of process are: the 'recognition' of characters from a pool of predetermined catagories; and the 'derivation' of formal proofs for some specialised class of mathematical problems.

A typical scheme for PR in the science of A.I involves one or more of the following three stages:-

(1) - Sensory image processing: sharpening, noise-reduction, or normalising input patterns.

(2) - Pattern classification: naming objects, and formulating equivalence classes.

(3) - Scene-analysis: identifying components of objects, and describing how these components interrelate and form larger structures.

A problem which has increasingly come to the forefront of the area of scene-analysis research is the question of three-dimensional (3-D) vision. While, in the earlier models, it was thought that a simple two-dimensional representation of objects could be extended to solve this issue, today, the complexities of 3-D vision are better understood and appreciated, and such a simplistic progression of ideas is not considered viable.

In the PR techniques of A.I., classes are normally defined by structural relationships of a set of features. For example, the connectivity, proximity, and orientation of various segments of alphabetical characters are used to form structural classes in character recognition programs.

Some early scene-analysis programs (e.g., Roberts, 1960) could recognise simple shapes made up of straight lines, such as boxes, wedges, pyramids, or any combination of them. These programs used simple engineering methods, based on geometric properties of such shapes, to predict the identity of a new pattern presented to the model, even though in some cases parts of their outlines were obstructed. Allowances were also made for camera perspectives, and in some programs the background of the room containing the objects could be 'recognised' and segregated from the main patterns.

Guzman (1968) (and other workers) employed a more general approach, sometimes referred to as the 'heuristic approach', in analyzing scenes of straight-edge objects. His program would take simple rules based on some trivial properties of objects such as vertices, edges, or faces; and by testing, experimenting, and modifying these rules produce a criterion which worked for some specific class of problems, and, hence, was able to recognise particular class of objects. Problems of much higher magnitude are, however, introduced when pictures to be analyzed include curved objects, or have light and shades, or when the issue of three-dimensional vision is to be considered.

Slagle (1971) had devised an algorithm for finding a so called 'linear evaluation functions', which by acting on feature vectors was able to evaluate or recognise patterns. He anticipated that these type of functions would have applications in game-playing, learning, utility-theory, automatic feature extraction, programming, and other socio-economic systems. The evaluation of these functions for any given task was said to simulate learning, in so far as that it could approximate the process used by an individual, or a consensus of experts, in carrying out the same task. Various analytical procedures were proposed for finding a 'good' set of values for the coefficients of these function - such techniques were an extension of already well established methodologies of statistical decision theory for patterns with known distributions.

Here, it was implicitly assumed that features are readily detected by an expert pattern-recognizer or problem-solver, and that their expertise could be realised in terms of a function of these features. However, for most practical tasks, this assumption is grossly inadequate, and the traits are either not identifiable, or the procedures for decision-making involve many other unknown factors. Furthermore, these techniques do not incorporate facilities for generating and testing novel features, or evaluating new hypothesis. Hence, the relative lack of progress in these lines of research could be contributed to the narrowness of its methods, and to some of their underlying phenomenological assumptions.

Learning a language without perceiving the objects or subjects it is about would be a meaningless task. The complex interdependence of human faculties of vision and language has, hence, been depicted in other PR programs of A.I. Probably the best example is Winograd's work on 'block worlds', in the early 1970's.

Later on, the more conceptual, the 'semantic net' type, recognition models were developed by other workers such as Winston (1977). These semantic networks could augment the previous class of models, by interpreting the results of a scene-analysis program.

The above class of models, which heavily rely on the semantic representations or descriptions of perceived data, have sometimes been labelled as the 'top-down' solutions to the problem of PR. However, according to some critics, the top-down preprogrammed rule-based (rules arrived at heuristically or otherwise) PR approach, despite having some similarities to the mechanisms of human perception, has many shortcomings. Their contention is witnessed by the fact that, so far, such models have been unable to deal adequately with real environments (outside laboratory), and have usually been confined to recognition tasks involving specific classes of relatively simple objects with uncluttered backgrounds. Their proposed alternative approach to PR, the 'bottom-up', is dependant on the more fundamental properties of human and animal perceptual systems, and therefore, in principle, should be able to tackle patterns of much higher complexity. An underlying feature of systems in this, sometimes called 'adaptive PR', approach is their widespread use of the learning process as a means of achieving the objective. Techniques employed are, normally, different from the 'serial' algorithmic methods of conventional computer oriented researchers, and involve 'parallel' processes - more or less emulating the sudden simultaneous changes of neuronal states of the brain.

In our continuum of investigations of the area of artificial 'learning' systems, the higher conceptual PR models will be discussed more fully in the later section dealing with Artificial-Intelligence.

5.1.7 DESIGNING PATTERN-RECOGNITION SYSTEMS

A simple line of argument will be followed here, whereby, the basic issues involved in designing PR 'learning' systems, in particular the parallel systems, will be elaborated step by step. FIG.5.1. shows schematically the principal components involved in a typical PR system; however, in practice, some stages may be eliminated.

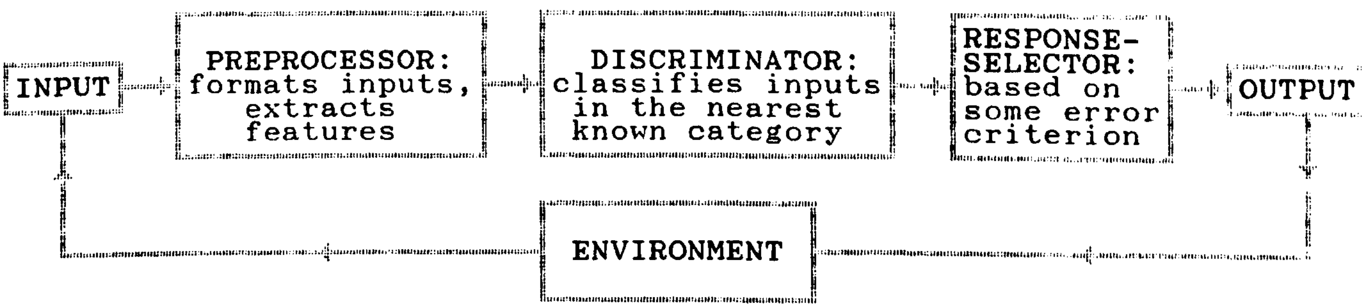


FIGURE 5.1. A general blue-print for a pattern-recognition system.

(i) - REPRESENTING and ENCODING INPUTS

The question of 'representation' of inputs, for models which do not exclusively rely on their raw sensory data, is of prime importance for PR model-builders. This issue, sometimes referred to as 'pre-processing', has been tackled by various mathematical techniques (e.g., probability distributions, fuzzy-sets, etc.), as well as some purely descriptive methods. The elaboration of a particular model and its range of inputs dictate the level of complexity and the type of representational characteristics that need to be considered. For example, whether it is necessary to consider the higher semantic or contextual issues or not.

In designing a simple PR system the first step is to present the input to the system in its most trivial components. This is normally done by breaking up the incoming data into an ordered array of segments. The digital nature of computers, and most neural network models, has meant that information in the majority of such models is presented in a binary form - giving a sequence of binary bits as the state of the matrix of input cells.

Moreover, it is normally desireable to reformat or change the initial basic binary input pattern in the latter stages of preprocessing (for the ease of subsequent analysis). Some of the methods used for preprocessing (or 'conditioning') the primary input patterns into a more convenient form are:-

(1) - 'Enhancing' or 'emphasising' certain characteristics of input data which are deemed important.

(2) - 'Providing invariances' in spite of minor changes, obstructions, perspectives, environmental changes, or spatial positions.

(3) - Reducing the amount of incoming information by 'noise suppression', 'normalization' of inputs, 'focussing', or 'segmentation'.

(4) - 'Reformatting' the inputs, for ease of processing by next stage of the system.

(ii) - TEMPLATE-MATCHING

Now, the most obvious and intuitive approach for pattern-recognition would be to compare the incoming information with a specific set of previously stored examples (or more commonly called 'templates'), and calculate their similarity to these examples on the basis of their optimal matching of bits. Hence, the simplest of all possible schemes for classification and PR is 'template matching'. Machines whose designs are based on this criterion are easiest to build, yet are most susceptible to error.

An arbitrary threshold N can be chosen, above which the correlations signify a recognised pattern, otherwise, the pattern is either stored as a 'new' template or discarded as 'insignificant' - here, 'significance' could either be defined on the basis of frequency of occurrence of inputs, or on the basis of distinctiveness from previously stored patterns. A simple illustration of this scheme is shown in FIG.5.2 (next page).

The process of template-matching can be seen to be present in most PR models. Since, even those systems which are based on the more complex processes of feature-extraction and feature-recognition use template-matching on portions of their total input patterns at some stage - to locally identify the nature of a feature (e.g., edge, point, angle, etc.).

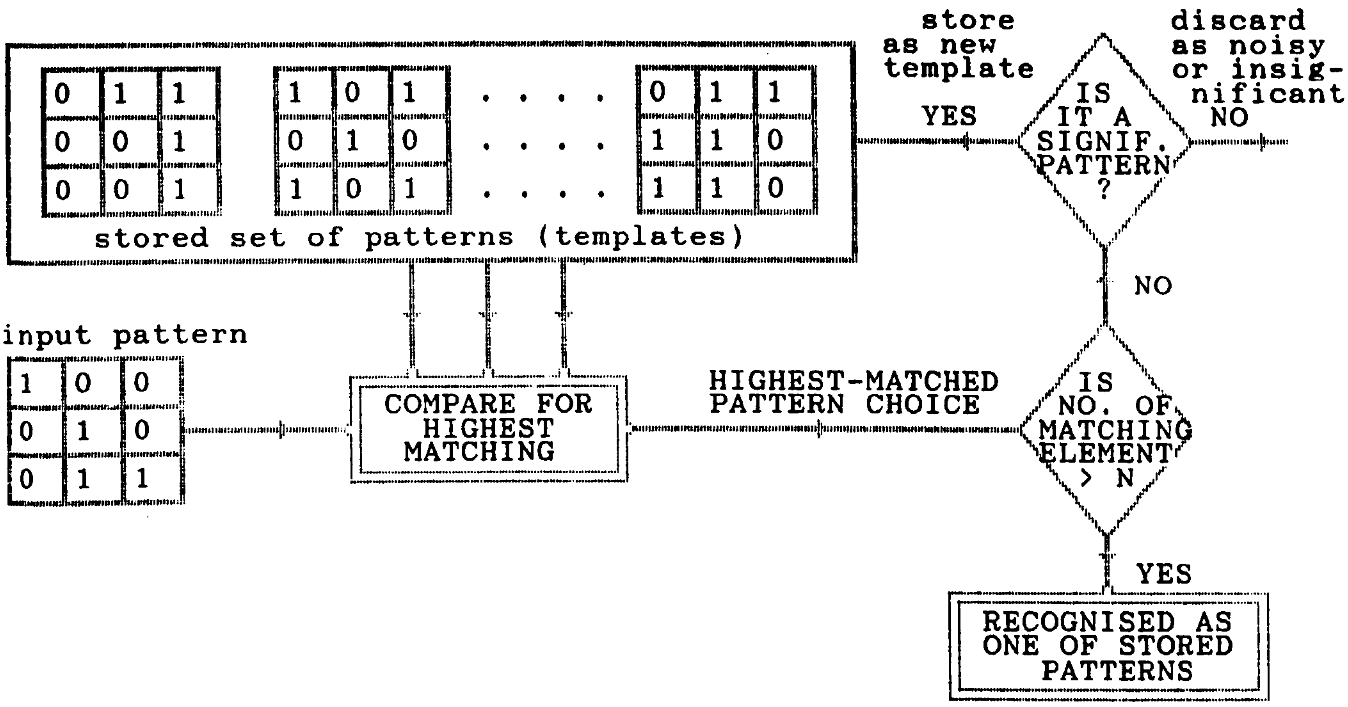


FIGURE 5.2. A simple scheme for recognition of patterns presented and stored as arrays of binary bits.

(iii) - PRESERVING INVARIANCES

In the basic form the above procedure is clearly not going to be very efficient, since patterns that are similar to the stored templates yet are rotated or shifted to a slightly different orientation, or enlarged/reduced in size, will not be recognised as same, because of their poor matching.

Ideally, PR systems should be able to identify objects in spite of their spatial transformation, partial inputs, or slightly distorted inputs. As stated by George (1973), the long term goal of a PR designer is: "the search for a system that preserves, certain invariances under transformations." (He also points out the clear interdependence of the processes of PR and learning, and contends that only the utilization of previously acquired information by way of learning could result in an efficient manifestation of the natural recognition process).

One way of preserving invariances is to carry out various spatial transformations on the input pattern, and try to convert it to a more standard form; alternately, all orientations and sizes of an input could be matched with the templates. In either case, we can see that an extra level of processing becomes necessary to reduce the otherwise enormous number of exhaustive classifications - that would have resulted from individual storing of all possible variations. A commonly used technique for standardization is to find maximum density areas of a set of patterns, or, alternately, find the average values of groups of data points. Similarly, various mathematical techniques such as transformations of coordinate systems or matrix-transformations could be used as a possible means of realising spatial variations of a standard pattern.

Standardization of input patterns is particularly useful for technological or commercial applications involving PR (e.g., character recognition, automated sorting machines, etc.).

'Smoothing' and 'sharpening' operations, which remove some of the noise and irregularities from inputs, are very common in most computer vision systems based on digitization of picture patterns. Mathematical processes which entail transformation of zeros and ones of picture arrays have been scrutinised extensively by numerous workers. In particular, studies involving "Life" transformations have yielded some interesting results about the behaviour of networks of interacting elements. The possibilities of using "Life" transformations as means of achieving standardization have also been investigated.

In one of the earliest PR models, Selfridge's (1956) computer based model, the input patterns were formed by an array of sensory elements, arranged in an NxM matrix. These primary inputs went through a series of transformations, in some cases through the same transformation more than once, to form 'secondary images'. Three basic transformation operations resulted in: (a) - an increased uniformity of patterns; (b) - the emphasis of differences (or contrasts); and (c) - the replacing of relatively isolated sets of elements by a single element. The secondary images were, then, compared and matched with some stored set of templates, and when a specific predetermined level of correlation was achieved, a pattern was said to be recognised. The most important aspect of this model was the finding of the correct sequence of the above three basic operations for reaching the recognizable patterns. The sequences were initially chosen at random, but after 'detecting' successful sequences, a 'matrix of transition frequencies' was biased accordingly.

Culbertson's (1963) PR model, based on a neural-net type representation, had the capability to 'translate' (shift), 'rotate', 'dilate', 'expand', and 'centre' the shape of an input image into alternate forms, and then compare these translated images with previously stored templates.

(iv) - CLASSIFICATION

'Classification' is one of the fundamental issues of PR, as well as being one of the most basic and common operations in science. The ability to make appropriate classifications is basic to all intelligent behaviour. Classification, in a trivial sense, involves taking a 'sample' and naming the 'category' or 'catagories' it belongs. The most simple case for classification is when classes are to be assigned on the basis of the values of some quantitative variables of a particular description space. For example, in the classification of height or weight measurements of a group of people. But, when the variables are of a subjective, incomplete, or unknown nature, then the task of classification becomes a much more difficult endeavour.

The way classes are described (or selected) is principally dependant on the use which will be made of them. Classification in its simplest form is only a basic task of recognition. But, generally, three basic methods are used for process of classification:-

(a) - Selection by a simple direct matching process (e.g., identifying a letter of alphabet by individually matching with a set of possible examples).

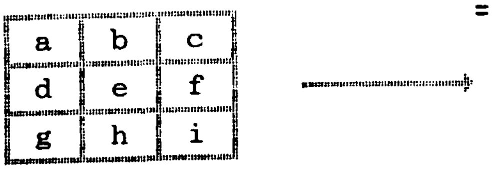
(b) - Utilising some logical information about patterns (e.g., a shape is a "SQUARE" if it has four equal sides and four right angles).

(c) - Choosing outcomes on the basis of statistical causalities of events (e.g., if event A is followed by B with high degree of probability then A is classed causally with B).

In type (a) and (b) classification we can see that samples have to be defined either by identifying all members or categories, or by defining their characteristics. In addition, similarities of an input pattern with the members of these sample spaces have to be evaluated. Once a novel input has been confronted, a decision must be made to either include it within a previously defined category, or to introduce a new category. But, care should be taken to not increase the number of classes unnecessarily, or have too few classes, or disturb the entire classification when adding new catagories.

In a broader sense, the process of PR has also been divided into a process of 'feature-extraction' and a process of 'classification'. Most methods of non-trivial feature-extraction are quite ad-hoc, while, for classification a variety of independent and well developed techniques are available for any particular type of problem - generally assuming a statistical independence between features selected. However, this simple division of the process of PR becomes quite inadequate for the more complex or contextually dependent recognition tasks, and a variety of other issues must be considered.

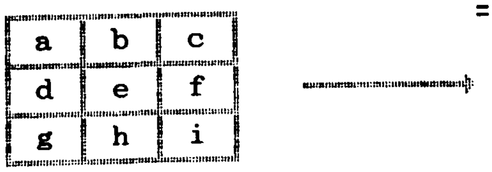
In the exhaustive extreme of classification, as depicted in the simple example of FIG.5.3(a), all possible inputs are taken into consideration. Hence, whenever a particular input symbol (e.g., bd, eh, eghi) is activated, we could recognise it from our complete classification system. The obvious futility of this approach for any non-trivial system suggests that only a selection of stimuli should be classified and stored, such as in FIG.5.3(b); these inputs could either represent the inputs that occur more frequently, or the properties/features that occur regularly in the inputs. Now, supposing the more feasible case of the latter, the real task is how to select these properties so that neither we are, again, confronted with a high number of trivial classes, nor general features that cannot adequately distinguish between inputs. In other words, to efficiently recognise an object we need to recognise its list of properties/features, but the shortest possible list.



COMPLETE CLASSIFICATION OF ALL INPUTS

a, b, c, d, ...
ab, cd, ...
abc, bcd, ...
. . . .
. . . .
abcdefg

(a)



NON-EXHAUSTIVE CLASSIFICATION OF REGULARLY OCCURRING INPUTS or PROPERTIES (FEATURES)

a, c
ab, ef, gi
efh, defghi, abefhi

(b)

FIGURE 5.3. (a) – The exhaustive classification of all possible combinations of firing on input elements.
(b) – An example of selective classification of sets of firing elements.

In our example, 'ab', 'ef', 'gi', etc. could represent properties (e.g., size, colour) of inputs; and for an input such as "abcef" we could make a recognition based on which features ('ab' and 'ef') are detected. However, the list of features may be extended or reduced based on other frequency considerations. For example, if new and consistent features are detected, or if too few features are defined, or if two features are coincidental on numerous occasions.

If we, now, look at some possible ways which decisions about the identity of the class of an object may be reached from the detected features, then we could see that, probably, the most promising methods are those which involve statistical probability considerations based on previous experiences. These methods could entail attaching probabilities (or conditional probabilities) to various features being associated with a particular object.

If a 'learning' is also to be incorporated, then the probability values could gradually be modified using some 'counting' mechanism which determines the frequency of successful associations; hence, after sufficient experience, we could decisively associate the occurrence of a particular feature, or a set of features, with an object. In other words, given such features, the probability of recognition should approach 'one'; but, normally, in practice, lower probability levels than 'one' are considered quite adequate.

Now, if the machine itself is to determine and generate its own pattern classes (or feature classes) autonomously, then it is imperative that some measure of utility (or hedony) should be also incorporated within its design - so that the system can judge the usefulness of its selected classes. Where, again, the statistical and other techniques such as 'cluster analyses' could be employed to achieve a simple realization of a hedonic measure - spatial or temporal clustering of sample patterns could indicate possible boundaries for classification, and groupings defined accordingly.

A problem common to most PR models which use classification of traits or features, and statistical decision procedures based on this classification, is that the inherent connectivity of a pattern is not preserved. Whereby, data is collected at distinct instances of time, and stored in locations having no direct spatial relationship or dependence on each other. While, it would be more logical not to disturb the information contained in the topological continuities of input patterns - specially, since evidence from some neural observations point to this type of preservation of topology of input (although the input patterns are also diffused non-locally to some extent).

It should be mentioned here that time related elaboration of above ideas on "classification" and simple "recognition" (i.e., sequences rather than individual inputs) yet adds further complications to the probability calculation. The connection of temporal ordering of events on the basis of a measure of 'expectancy' will be one issue influencing the probability modifications of such PR scheme. In addition, when the patterns to be recognised are of a more complex form, then problems of 'segmentation of patterns' or 'feature distinction' pose more difficulties.

(v) - FEATURE-DETECTION, FEATURE-GENERATION and FEATURE-EXTRACTION

The issues of "feature-detection", "feature-generation", and "feature-extraction" will be further considered here. Patterns presented to a system are not simply shapes made up of homogeneous points or segments, but, as in most natural stimuli, normally consist of parts which have varying amounts of significance. For example, in the recognition of song patterns in birds, or in the recognition of symbolic patterns (e.g., numbers or letters) in humans; specific 'features'/'traits'/'attributes'/'properties' give the majority of information about a pattern. These features can characterise a pattern without actually relying upon the precise point-by-point (which is in fact the trivial listing of pattern 'features') configuration of the pattern.

The process of defining 'feature classes' for a set of observed patterns is an arbitrary one, since a variety of different features might be used to construct pattern sets - ranging from individual points to topological regularities or various semantic and syntactic characteristics of inputs. But, generally, it is the shortest and the most efficient list of features which is of interest.

Additionally, a PR machine (or program) has a 'most primitive' set of measurements which it responds to (e.g., binary bits or voltage levels). Such measurements do not, normally, have a direct bearing on its feature sets. Here, if we consider Ashby's Law of Requisite Variety, it can be said that for implementing an adequate recognition process the variety in the 'most primitive measurement set' should be greater or equal to the variety of the feature set involved in the task.

Therefore, now, helped by clues from natural observations, and other parsimony considerations, we can improve our previous PR model to one that recognises 'features' of patterns rather than their exact shapes.

There are two principal ways these features could be detected, generated, or extracted and subsequently utilised in the system. Firstly, the system can, in a sort of self-organising manner, generate its own features and decision procedures. For example, a random network (such as those discussed in the previous chapter) by reinforcing its pathways on the basis of some performance criterion could determine its significant features. This method is probably the most ideal way of selecting features; yet, so far, models devised on such basis have not shown a significant degree of complexity or expertise.

Secondly, the designer or the programmer could using introspection define all the adequate features of the model, a priori. In this method very little or no real 'learning' takes place and the performance of the system will remain rigid, yet the system will invariably perform much better and faster, and will start with a high degree of proficiency from the outset.

In practice, however, the systems designed are a compromise between the above two extremes - mixing versatility with efficiency. Typically, features are defined from the analysis of some sample patterns, using part introspection and part investigation of sample properties.

Here, we will attempt to elaborate on three principal schemes that have been developed for the analyses of features in various PR systems:-

(a) - At the most primitive level of input measurement (i.e., the binary coding of patterns) features are assumed to be subsets of the basic binary patterns. The primary objective, here, is to define a useful set of such subsets.

In FIG.5.4(a), a 9x9 binary array of cells (sensors) is illustrated, each element either being in '0' or '1' state. If alphabetic characters are to be classified, then, typical 3x3 sets of points (features), such as those shown in FIG.5.4(b), can be used as possible elements for building each character - any character could be defined as a union of some feature subsets.

```
      9
  000000000
  001111100
  001000000
  001000000
9 001111100
  001000000
  001000000
  001000000
  000000000

     (a)
```

```
    3
  100   000   100
3 010   111   100
  001   000   100

        (b)
```

FIGURE 5.4. (a) - An example of a 9x9 matrix of binary elements used for representation of characters in a PR machine.
(b) - Three typical examples of features observed in various characters - many other angled or straight line segments could also be used as possible features of the pattern.

Methods used for extracting this type of features depend on determining local regularities of patterns. Mathematically, the problem can be represented by binary matrices and analyzed by Boolean Algebra techniques. However, because of the difficulties involved in solving Boolean equations, the more successful developments in this area have, mainly, utilised alternate heuristic feature detection methods.

A type of heuristic feature detection was used in Uhr and Vossler's (1963) PR model, which was concerned with the problem of recognition of 25x25 binary patterns involving simple 2-dimensional pictures. Arrays of 5x5 bit configurations were selected as features, on the basis of their frequency of occurrence. Thus, a vector of such arrays could describe a particular pattern; and simple clustering algorithms could determine the category (class) of an unknown pattern (if any). In addition, these feature sets could be modified, and a sort of 'learning' (or "generation") of new features, and 'forgetting' (or "elimination") of little used features was said to be manifested - based on the extent of utilizations of features. The system, indeed, outperformed human-subjects in discriminating tasks where patterns were of unfamiliar, abstract, random, or non-meaningful nature. Yet, it was not as competent in recognising other range of patterns, for example, different faces, letters or numbers.

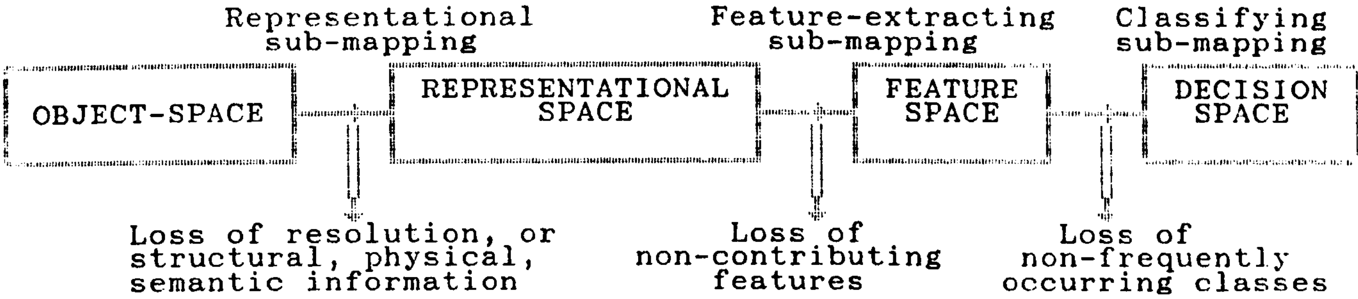
Other techniques have also been developed in this area. A more formalised approach to the above problem was proposed by Nilsson (1965) and co-workers which involved the use of set theory concepts and operations. The similarity of patterns were decided on the basis of their set theoretic commonality of features (or "masks").

(b) - A second distinct approach to the problem of feature extraction is that taken by engineering-oriented workers who, sometimes, regard their field to be a branch of applied mathematics. Here, patterns or objects are represented in terms of a set of basic (or primitive) measurements, which is normally denoted by a weighted combination of some basic features. The problem, now, becomes one of analyzing and determining various parameters and variables of feature (or 'factor') spaces. Methods used in this approach, usually, referred to as the 'multi-variable factor analysis' approach involve topics such as 'parameter-estimation', 'matrix-algebra', and 'statistical correlation analysis'.

Therefore, if we consider patterns as vectors in an n-dimensional space, then the recognition problem is the finding of regions of this vector space identified with particular class of patterns. Alternately, a process of PR can be described by a mapping from 'pattern-space' to 'decision-space', as:-

$$\boxed{\text{OBJECT/PATTERN-SPACE}} \longrightarrow \boxed{\text{DECISION-SPACE}}$$

The optimum mapping being one which has the minimum probability of error. However, treating such problems as a simple mapping by statistical decision methods is extremely cumbersome. Hence, in practice, this type of mapping is divided, quite arbitrarily, into sub-mappings, resulting in a loss of some sort of information at each stage of division. This is shown in the following diagram. The observation made is that these distinctions are not universally applicable to all PR models.



Most formal analysis of PR systems assume that a large class of features, which contain the necessary data for classification, are known a priori. The number of features could be reduced to a manageable size by: discarding the 'poor' features; selecting the 'good' features; or combining (generally linearly) a number of features into a single new feature. Techniques such as 'contour tracing' could be deployed to detect and extract features from input images or patterns. The algorithms used can detect lines, edges, etc. on the basis of the contrast of input picture pigments.

Similarly, the mathematical techniques of 'fourier analysis' have been used to define intensity functions of patterns, and have been used in some recognition problems of 2-dimensional images (Glorioso, 1975). The result of such transformations, and other similar techniques, is to redescribe patterns into a much more simplified form, while retaining some essential characteristics, and hence facilitate the use of well established formulas.

Particular difficulties have been confronted in devising general mathematical feature evaluation methods which can yield optimal feature-spaces for various PR systems. Some restrictions on the 'independence' or 'normality' of features have helped to simplify the problem in certain cases, and have resulted in fairly generalised solutions. But, on the whole, feature extraction methods are only defined for a particular class of problem in hand.

(c) - Thirdly, pattern-features could be detected and defined on the basis of criteria outside the immediate specifics of their physical forms. On the one hand, the designer of a PR system, using his expertise and intuition could devise a powerful, yet very specialised, method for dealing with a particular class of problem (e.g., in classifying a group of patterns such as those obtained on a chemical spectrograph). Here, no general techniques could be specified, and it is up to the individual designer to use his knowledge, and information about the characteristics of patterns under scrutiny, and propose an ad-hoc solution - the mathematical or other properties of patterns are used to simplify the task of feature detection or extraction. Many examples of these techniques can be found in the A.I. related work on PR. On the other hand, in the classification problems which involve semantic or syntactic elements different, and usually non-mathematical, feature-extraction methods are used. Here, feature-extraction is concerned with 'structural (grammatical) descriptions' of patterns, 'meanings', and 'relationships' between patterns. Such methods will be elaborated more in the A.I. section.

(vi) - PARALLEL vs. SERIAL OPERATIONS and PROCESSES

Assuming the essential features of a set of patterns are known, then there are two fundamental ways of comparing and matching incoming patterns with these stored features: (a) - 'sequential processing', (b) - 'parallel processing'. These two methods are outlined in the schematic examples of FIG.5.5.

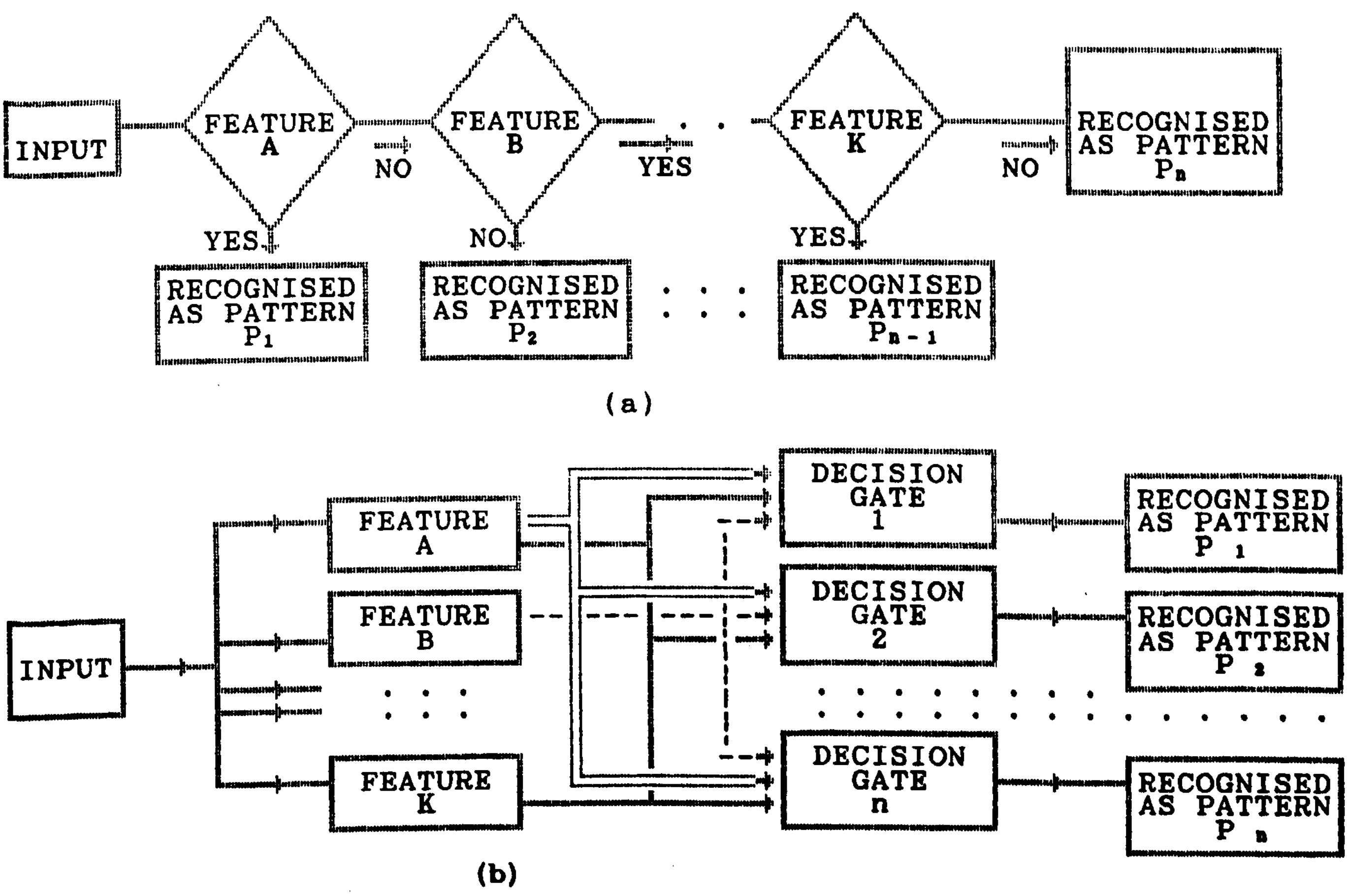


FIGURE 5.5. (a) - An example of a sequential recognizer of patterns $P_1,..,P_n$.
(b) - An example of an equivalent parallel PR system.

The sequential PR system tests each feature in turn, and the results determine the next steps. This is more or less the basis for the operation of most computers and machines in general. Parallel PR systems, on the other hand, function seemingly the way animals and humans handle information.

The historical course of development of PR models began with the design of simple parallel systems, involving statistical analysis of Boolean or Euclidean classifications of patterns, followed by serial systems which used sequential sampling methods. Later, these models lead to the introduction of computer-recognition (and vision) systems, based on structural descriptions and sequential classification/recognition techniques.

The particular advantages of parallel processes to sequential processes are their higher immunity to noise, distortion and error. Another advantage is the possibility of attaching adjustable weighting factors to parallel connections, whereby, slight adjustments could be made to the functioning of a system without disturbing the totality of its behaviour too much. While, in serial systems such small adjustments could lead to drastic variations in system behaviour. The implications of this final point is that gradual modifications and 'learning' could be incorporated much more easily in parallel systems - either by the system itself, or by changes brought about by an external operator (designer, programmer).

In addition, for any given parallel decision process, an equivalent sequential system could be designed with fewer steps; but, the speed of computation will, normally, be much higher in the parallel system. However, it has been found that the reverse is not true, and there are sequential procedures which do not have parallel equivalents.

Parallel classification systems could be, normally, represented by mathematical equations; but the serial decision procedures suffer from a greater complexity and, unlike their parallel counterparts, are much more cumbersome to analyze mathematically, and describe their formalised quantified representation. The serial classification rules are, often, in the form of sequences of conditional functions or probabilities, and, thus, can normally be only illustrated by a graphic 'decision tree' - as in the example of FIG.5.6. The criterion used for selecting the best rule could depend on notions such as the shortest path through the graph, or the least features used.

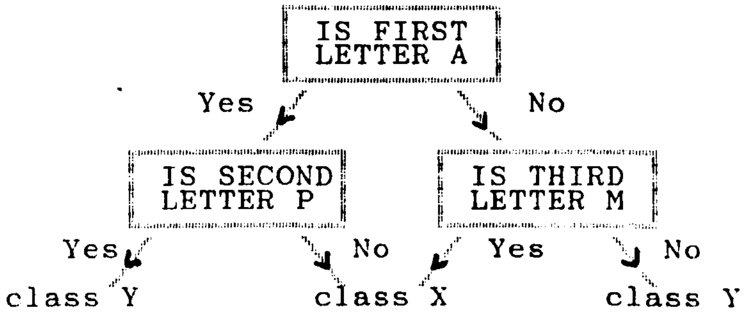


FIGURE 5.6. An example of a simple classifying rule (shown by the graph) for implementing the following problem:-

class X = { 'ABC' , 'ABF' , 'BHM', 'LGM' , 'AFT' }
class Y = { 'APM' , 'HBK', 'MTO', 'MPK' , 'MTA' }

In a limited sense, notions analogous to those used in formalisms of parallel systems have been defined for sequential systems to simplify their mathematical analysis. Concepts such as 'description spaces', 'samples', 'classes', 'classification rules', and 'classification probabilities' are some examples. Similarly, statistical techniques, such as 'Bayesian decision procedures', originally developed for parallel systems, have also been applied to sequential systems. Another important and widely used analytical tool for use in sequential type systems is 'dynamic programming'.

The various mathematical procedures that have been developed in the sequential PR approach are far less dependant on successive adaptive modifications of their classification rule, hence show far less 'learning'. This is, mainly, due to a lack of concrete general 'convergence' theorems, similar to those devised for parallel systems. Thus, unless a serial system is convertible to an equivalent parallel system, or has a small finite sample size, no general conclusions about a 'learning procedure' can be made. In other words, we could not find out whether a serial 'learning' algorithm converges to an optimal limit or not, therefore only specific ad-hoc solutions could be found.

Aleksander (1983) discusses the advantages of parallel hardware PR models on the basis of two principal arguments. Firstly, due to speed and capacity limitations it would be impractical to implement serial algorithms on present day computers (other than for very basic systems). Secondly, on the strength of similarities of parallel approach to some biological mechanisms of the brain. As an exponent of the, so called, 'bottom-up' view of PR he also attributes the relative lack of success of this, historically precedent, paradigm to the inappropriateness of supporting technological hardware devices.

5.1.8 MATHEMATICAL TECHNIQUES IN PATTERN-RECOGNITION

The interest of mathematicians and engineers towards PR was aroused in the late 1950's, particularly, in the area of automatic character recognition. The early endeavours were, mainly, occupied with ways of classifying input patterns of systems, and their devised early models (some hardware) basically tried to mimic natural vision mechanisms using clues from neuro-physiological studies.

Gradually a divergence from the main course of PR was established. An abstract mathematical PR related topic was developed which, generally, dealt with the property of automatic classification within the framework of formal (mainly statistical) systems. This field is sometimes considered as a subsection of computer sciences, however, its workers pursue research lines different (much more abstract) from those followed by computer scientists.

For the mathematically-oriented researcher in PR the problem is normally reduced to one of classifying vectors or points in multi-dimensional spaces - not the analysis of actual forms of patterns or images. Many aspects of such work are related to the classical techniques of control and systems theory involving multi-variable analysis.

Another diversification of statistical methods has been to devise systems whose classification rules are not determined from fixed samples of patterns. This class of problems are sometimes referred to as 'machine learning' or 'adaptive PR', and are closely related to the adaptive control systems (discussed in chapter 4). Here, the sample size is taken to be infinity, and the classification rule is continuously updated using successive approximations.

The algorithms used are normally called the 'learning rules', and their 'stability', 'convergence', and 'goodness' characteristics are of prime consideration. These 'learning' algorithms have been used in the implementation of some neural-net type systems - the best known example being the 'perceptron' (outlined elsewhere in detail). Various mathematical treatments and developments of such algorithms have been undertaken by numerous workers, for example, Nilsson (1965) or Minsky and Papart (1969). Typically, first, formalisms are developed for simple linear systems, and later generalised to higher order classifications, non-linear, or non-Euclidean systems.

Again, whether the psychological notions of 'learning' are emulated by such models or not is debatable. Yet, it is a truism that most of their designers strive for such an objective.

(i) - PROBABILITY THEORY AND PR

In current A.I.-oriented models of PR the use of probability theory is, mainly, limited to the manipulating and representing of uncertain information in game-playing or problem-solving situations. However, in the neural-net or the engineering approach the mathematical concepts and theories of probability play a much more central role, and, often, probabilistic criteria constitute the essence of techniques used. In particular, when the outcome of uncertain events is to be defined or approximated, or when it is desired to establish correlations of various hypotheses and events for a given set of evidence.

One of the theories that has been extensively employed for finding such correlations is 'Bayes' theorem'. Bayes' theorem provides a methodology for calculating the validity of hypotheses (events) for a given set of observed events. Bayes' theorem is based on the concepts of conditional probability, and it has also found applications in some A.I. programs for recognising patterns in terms of previously observed characteristics. But, the more widespread applications of this theorem are found in the lower-level classification (PR) models which use the trivial significances of input patterns. The basic form of Bayes' theorem has also been elaborated and modified for various complex or specialised problems, and in some cases other techniques of probability theory have been used in conjunction.

In addition, mathematical concepts such as "conditional probability" are used in 'Bayesian inference methods' to devise 'decision procedures'. In the more classical statistical approaches if the appropriate probabilities are not available, using information from samples of data and some tentative probability distribution functions, some parameter-estimation techniques can avail the value of probabilities of encountering a particular pattern (or class) from a class of patterns with specific features.

(ii) - STATISTICAL TECHNIQUES IN PR

As in the case of most other approaches to learning covered in chapter four, statistical methods form an important and integral part of pattern recognition techniques. Indeed, in some cases, the work of analytical PR researchers have been categorised within the body of mathematical discipline of statistics and measurement theory. The basic criteria of the statistical approach to PR are that an object can be wholly represented and defined by a set of features, and that such features should be known before the process of classification could take place.

The formalisms devised for these type of models are based on describing patterns (or objects) as points in a multidimensional Euclidian Space, whose axis constitute scalars for particular features. Ideas of proximity, categorization, or reorientation of patterns are treated in terms of distances, regions, or transformations of such description spaces.

The principal objective of a mathematical feature-extracting system is to generate an n-dimensional vector from the input patterns it encounters. This vector should be able to adequately 'characterise' the patterns that are to be recognised. However, the intuitive and vague manner by which this characterization has so far been manifested leads us to believe that feature-extraction should be regarded as one of the most complex aspects of the process of PR.

In general, the process of feature-extraction could be manifested in two basically distinct ways:-

(a) - 'Logically' extracted features: the designer includes features he deems to be important.

(b) - 'Statistically' extracted features: a sample of pre-classified patterns, or continuous input data, are used to change or supplement the existing feature list.

A mathematical elaboration of classification of inputs can be made in terms of features. Each class could be represented by a function C_i, a weighted sum of a set of features $\{f_1,f_2,f_3...\}$. Such functions could be of form:-

$$C_i = w_{i1}f_1 + w_{i2}f_2 + w_{i3}f_3 + \ldots$$

Where, w_{ij} represent the weights attributed to each feature f_j. The value of weights could be positive, zero, or negative, identifying the degree of significance or hinderance of that particular trait to the class C_i.

Once a feature-set is selected, then the problem of classification can be mathematically tackled by a variety of formal statistical or decision theoretic means. These methods, normally, involve either the minimising of an 'error function' or the maximising of a 'discrimination function'. But, a prior knowledge of the probabilities of occurrence of classes, and also the conditional probabilities of features (or feature vectors) is required. However, in most PR problems such data are not known in advance, or are incompletely known. Therefore, various techniques have been developed to estimate the values of these probabilities and their distributions, and hence enable the application of well established formal classification methods to a particular problem.

Techniques covered in the 'adaptive control systems' section of Chapter 4 are very similar to these estimation methods of PR. Samples of data can be used to gradually 'learn' the values of certain parameters of a probability or density function - Bayes theorems providing the principal theoretical groundwork for such techniques.

Alternately, when no clear functional form can be assumed about a probability density, then other general statistical procedures, such as 'least-square' or 'nearest-neighbour' techniques could be employed to estimate the density. However, here, a much larger set of samples and data points are necessary.

The need for this large number of samples can be reduced by introducing the, so called, 'adaptive classification procedures'. These techniques could be applied to various forms of distributions of pattern occurrences; and can, normally, yield an optimal classification rule after a series of weighting factor modifications (which use specific reinforcing or error-correcting criteria). A disadvantage of these techniques is that after the optimal classifier is derived from the analysis of a particular set of samples, then there is no guarantee that in the real mode of operation its performance will be optimal. The formalisms involved have been investigated by workers such as Rosenblatt (1962) and Nilsson (1965).

Irrespective of the method used for classification, it would be very desireable to construct machines (or programs) that could define their own classes. Classes could be defined by self-adjusting the parameters of simple features (i.e., f_i's). This feature is commonly referred to as 'unsupervised-learning' or 'learning without a teacher'; and is, normally, incorporated in designs where many analyses of samples are necessary to discover the underlying consistent traits. However, it is found that a good deal of prior information should still be incorporated in such systems.

Similarly, the issue of 'overlap' of classes has been a point of particular interest when designing algorithms for classification.

There are many ways of designing the 'classifier' of a statistical PR system. Such 'decision mechanisms' are discussed by Vigilone (1970) in a survey of applications of PR technology. A typical configuration of a PR system is considered, as illustrated in FIG.5.7. Here, the case for a two-class problem is shown - additional response units could be incorporated for multiple class problems.

After a pattern is pre-processed and represented to the feature detectors, then their outputs (b_i's) are denoted by:-

$$b_i(j) = \begin{cases} 1, & \text{if the j-th pattern contains the i-th feature} \\ 0, & \text{otherwise} \end{cases}$$

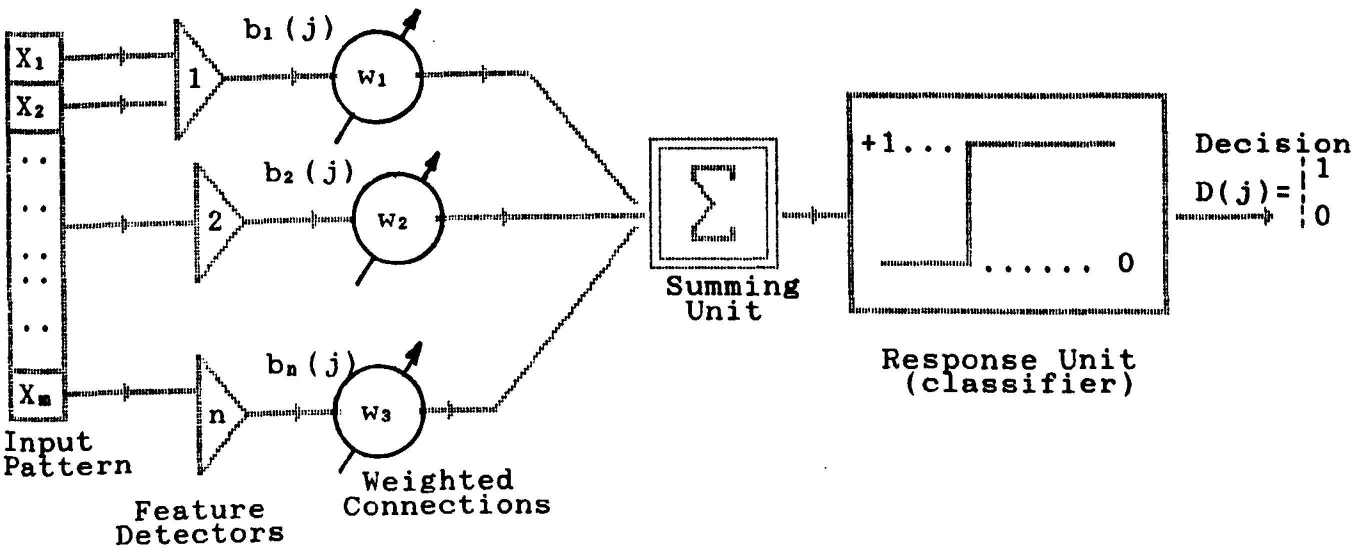


FIGURE 5.7. A schematic illustration of a typical PR decision structure from the engineering point of view.

Now, the response of a linear response unit for an input j could be defined by:-

$$D(j) = \begin{cases} 1, & \text{if} \quad \sum_{i=1}^{n} w_i b_i(j) - T > 0 \\ 0, & \text{if} \quad \sum_{i=1}^{n} w_i b_i(j) - T < 0 \end{cases}$$

Where w_i's are the adjustable weighting factors, and T is a decision threshold value.

There are also many techniques for the modification and updating of weighting factors (w_i's). Viglione embarks on a comparative study of some such methods, namely: 'Forced Learning', 'Error Correction', 'Bayes Weights', 'Iterative Design', and 'MADALINE'. Furthermore, the application of these different ways of achieving the decision function are discussed in various practical problems, such as the analysis of EEG wave forms or satellite photographs, and their relative merits investigated.

Another formal approach, distinct from those mentioned so far, has been proposed which uses the concepts of 'sequential decision theory' to find the classification rule or feature-space. Here, the principal difference is that decisions are taken on the basis of continuous evaluations of confronted data, and not a sample set. Both 'forward' and 'backward' sequential techniques, such as dynamic programming, have been utilised for feature selection and classification purposes.

The above techniques are predominantly designed for 'linear' problems, hence, the idea of 'linear-separability' of classes is of importance; however, various other techniques have been developed to transform a non-linear type problem into problems containing linear classes.

Other analytic tools such as 'decision theory' and 'filter theory' have also been utilised in dealing with the probabilities of occurrence of events, and in minimising errors between the actual responses and the desired responses of PR systems.

Finally, some theoretical solutions based on a compound decision theory have been put forward, to deal with the contextual aspects of recognition - omitted in the simpler isolated PR problems.

On the whole, although the above formal techniques have added a valuable general theoretical dimension to the science of PR, nevertheless, in reality most research in this field is directed towards practical ad-hoc (and non-global) solutions.

(iii) - INFORMATION THEORY AND PR

"Information" received from the primary levels of a PR process can also form the basis for PR models. Whereby, quantitative aspects of information contained within patterns is investigated using formalisms and techniques of information theory. Some methods deal with the raw data, others deal with feature spaces.

Rapoport (1955) analyses a neural-net type PR system in terms of some information-theoretic concepts. Various hypothesis are proposed regarding the way information is transmitted and transferred from retinal receptors to the brain. Similarly, the plausibility of different criteria for the transmission of messages during the process of recognition is investigated.

5.1.9 EXAMPLES OF PATTERN-RECOGNISING MODELS INVOLVING 'LEARNING'

The three main components of any PR process are: (a) - the 'description' or 'representation' of objects (or their classes) to be recognised; (b) - the 'classification-rule'; and (c) - the nature of 'prior information' supplied to the model (e.g., sample patterns, priming of parameters).

An important observation made about the majority of PR models devised so far is that they are more concerned about the process of 'recognition' than 'learning to recognise'. Now, if a process of 'learning' was to be manifested in any such system, then all or some of the above three components could be involved. But, more often than not it is only the modifications of the 'classifying-rule' which is the target of the designers interested in implementing some 'learning' or 'adaptability' aspects. Generally, a performance or error evaluation function determines how successful the 'learning' is. The classification-rule changes are governed by the previous results of system/environment interaction. Either a 'specific memory' of occurrences is kept, or 'statistical recordings' of occurrences (the average or typical cases) are used to update each tentative classification-rule.

An interesting case is when no direct error feedback from the environment is used for the modification, in such a case, it is assumed that the in built classification-rule of the system is correct, and only minor adjustments are made on the basis of some averaging of inputs; this type of 'learning' is commonly referred to as 'unsupervised learning process'. Other notions introduced in the area of rule modification are: 'convergence', 'optimality', 'computational complexity', etc. These terms, evolved from within the more analytical studies of the subject of PR, and refer to the attributes of the various algorithms involved.

In the following sections some particular examples of PR models where 'learning' is a prominent feature will be discussed in more detail.

(i) - PANDEMONIUM

One of the earliest 'learning' PR models, using the parallel principle, was Selfridge's (1960) "Pandemonium" system. Pandemonium was a computer based system which specified a sequence of events needed for a feature analysis of different pattern. For example, in recognising ten different hand-written characters. It could apparently 'learn' how to associate various features and patterns. Yet, it could not deal with the problem of spatial transformations of a pattern.

Pandemonium was made up of a hierarchy of relatively autonomous units, the so called "demons". At the lowest level, 'data demons' would encode the input information and pass them on to the 'computational demons', these, in turn, detected and combined features in patterns (for some simple systems it would be possible to exclude the computational demons). At a higher level, 'cognitive demons', one for each class of pattern, would indicate the relative likeness of an incoming pattern to different stored classes. Finally, at the topmost level, a 'decision demon' would give the final choice of the system for a given input pattern. A ch matic representation of this hierarchy is shown in FIG.5.8.

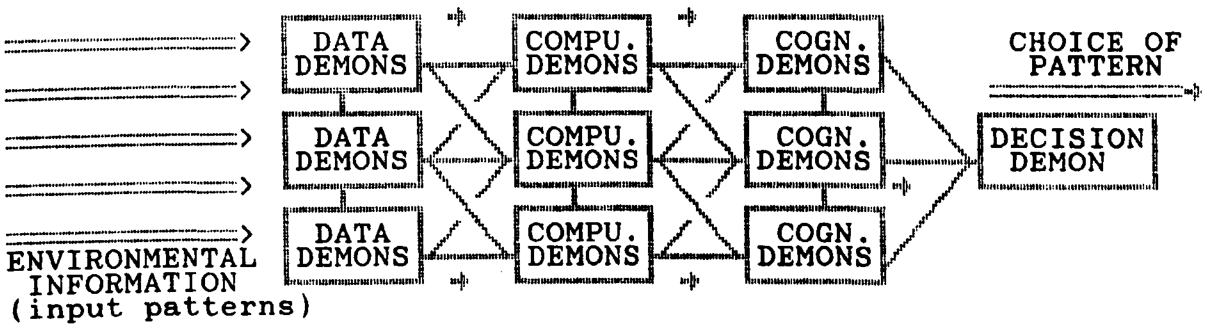


FIGURE 5.8. A schematic representation of different levels involved in a typical 'Pandemonium' system.

The Pandemonium, as a 'learning' machine, using a feedback from its environment makes adjustments to the connections between its 'demons', so that a better performance could be obtained from the machine. The evaluation of performance could either be done by the degree of some goal achievement, or judged by an external 'teacher' or 'supervisor'.

Techniques used by Selfridge to realise these adjustments entailed defining functions of the form:-

$$D_i = \sum_j w_{ij} d_j \quad ,$$

where D_i is the output of i-th cognitive demon and is determined by the weighted sum of outputs of all computational demons, the d_j's. Similarly, the 'worth ', the W_j, of the j-th computational demon to the whole system is determined by:-

$$W_j = \sum_i |w_{ij}| .$$

Once the weights were adjusted to a relatively successful set of values (using various algorithms, some described in chapter 4), it was possible to eliminate the 'low-worth' demons, or generate other new demons. The two following possible schemes were suggested by Selfridge for generating new computational demons, both involving some modifications of some existing demons:-

(a) - 'Conjugation method': combining two 'high-worth' demons.

(b) - 'Mutated fission method': forming new demons which are similar but not identical to the 'high-worth' demons.

At the first stage of a ten character recognition pandemonium program, inputs were presented on a 32x32-cell matrix; in the second stage, the patterns were smoothed-out, and then presented to the feature recognising sub-system. Some features, deemed to be significant by the designers, such as the number of intersections or the length of different edges, were initially stored in the program.

The initial 'learning' phase of the Pandemonium involved presenting various known inputs to the system and evaluating the relative frequency of occurrence of different features. A look-up table was formed which could give the probability distribution of outputs in terms of the defined features. Later, if an unknown input was fed into the computer, the system could test for each of the features, and depending on features detected it could guess the character with the highest probability of matching.

For this simple task domain the program was able to perform the task of recognising ten characters with a respectable efficiency (10% worst than a human reader). However, as conceded by Selfridge himself, the design of a comprehensive pattern recognition system would pose problems of many higher orders of magnitude. In particular, the problems of: separating or segmenting inputs into individual identifiable parts; devising systems that generate their own test features; and also introducing different levels of 'learning' to such systems, so that, as information is transferred from sensory to higher semantic and syntactic centres various refinements and improvements could be manifested.

(ii) - PERCEPTRONS

Probably the best known class of models within the PR approach are the so called 'perceptrons', initially introduced by Rosenblatt (1958). Perceptrons could be regarded as descendants of early neural-nets, since a continuum of ideas is apparent, and also their representational notions are closely related. Perceptrons were, like Pandemoniums, parallel feature detection systems with some 'learning' capabilities. Rosenblatt's contention was to introduce a general purpose functional model of behaviour which did not depend heavily on structural topologies, or specific logical realization of a system's function; instead, he was, mainly, concerned with the organizational properties of systems.

A simple perceptron refers to a class of network type models that could be characterised by: a 'stimulus-unit', an 'association-unit', a 'response-unit', and a 'variable interaction matrix' which depended on the past activity of the network. This matrix is sometimes called the 'structure matrix', and gives the states of coupling coefficients between pairs of units. Also, both stimulus and response units can generate 'internal' as well as 'external' signals.

Furthermore, various specific conditions could be imposed on the type, the strength, and the signal transmission characteristics of connections between different units. Transmission of signals along a path of a network can be affected by the 'value' of the connection and its 'transmission time'.

Other more elaborate forms of perceptrons have also been devised. Examples are 'experimental perceptron systems' (which are perceptrons connected to a reinforcement control system); and the more generalised perceptrons, the so called 'universal perceptrons'.

In addition, developments based on perceptrons have resulted in the introduction of various solid formal concepts such as 'linear PR systems' or 'threshold-logic systems' - constructed, respectively, on the primary units of 'feature-recognisers' and 'threshold-logic elements'. These mathematical concepts and their applications have been extensively investigated, both in abstract form, and also in specially designed hardware (analogue computer) devices.

(iii) - WISARD

Aleksander (1983) and other engineering-oriented co-workers have elaborated a PR system over the years. The basis of their parallel model, which has also been realised in hardware, is the RAM (Random Access Memory) unit of computer storage systems. The RAMs are considered as simple analogues of neurons, and various aspects of their operation is compared with a rudimentary view of neuronal firings or non-firings, and synaptic changes.

The simplest realization of this type of model is in its, so called, 'single layer net' configuration of RAMs - no interconnections between the RAM elements. Automatic visual PR is one area which could, according to Aleksander, exploit this type of system to its full potential. The basis of functioning of these single layer nets is to change the contents of a memory address, and later when that particular address is read, the existence of a '1' on the output would be equated with 'seeing' ('recognising') the pattern corresponding to the digital code of that address. Elaborations of this basic design idea, involving larger number of elements, could be carried out - whereby, more varied and complex responses could be elicited.

A more detailed analysis of the ways 'recognition' could be carried out by RAMs shows that combinatorial explosion will inhibit the design of systems that are solely based on single RAMs identifying each distinct pattern (of acceptable resolution). Hence, ways should be found to limit the storage requirements, yet enabling the system to deal with small variations and distortions of a basic pattern adequately.

Aleksander describes the, so called, 'n-tuple' method of segmenting the array of patterns into distinct areas, each segment being associated with a single RAM. The system is 'trained' to recognise different variations of individual segments, as a particular pattern is moved/distorted from its 'ideal' orientation. Furthermore, a generalization is introduced by defining the final recognition decision choice on the basis of a statistical consensus of segments recognised. A surprising versatility is manifested in the system by the adding of the above properties, and many 'novel' and 'untaught' patterns can be recognised with a high degree of success - a kind of 'universal' of a pattern is said to have been created. However, it was found that this segmentation of patterns could not be carried out to extremes, since gradually ambiguities would creep into the system, and although in the 'saturation' state of one

RAM per element of the pattern the capacity of storage will be at minimal, the discrimination will be very poor. Hence, a compromise should be made between the number of RAMs, the size of storage capacity, and the degree of recognition accuracy.

An interesting feature of this model is the way it has managed to emulate some biological properties of visual mechanisms. For example, in experiments on cats, it was established (by Hubel and Wiesel, and others) that some areas in the visual cortex are dedicated to detecting horizontal or vertical patterns, and they are only partially active in the intermediate angles.

An 8x8 artificial system made up of sixteen 2x2 segments is trained to recognise vertical or horizontal patterns of the form illustrated in FIG.5.9(a). Upon confronting a pattern similar to FIG.5.9(b) it will show the uncertainty of its equivalent natural system - 11 of 16 segments are detected for the 'horizontal', and 11 of 16 segments detected for 'vertical' (each segment is a 2x2 array).

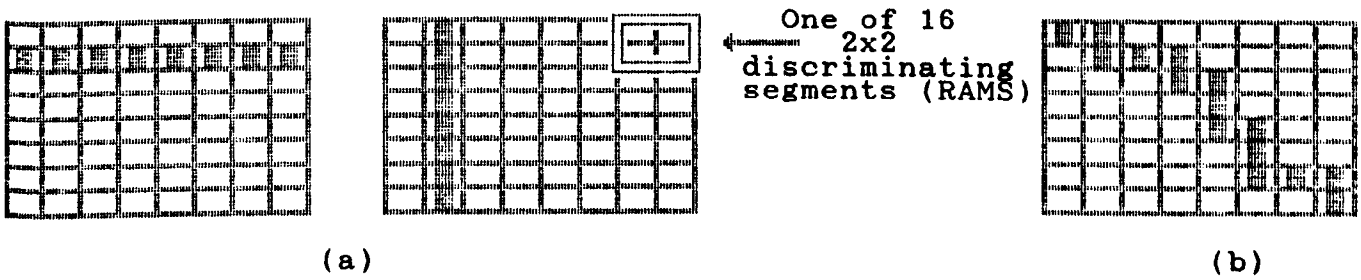


FIGURE 5.9. (a) - Two examples of 16 different perfect horizontal (8) and vertical (8) patterns used to train the system.
(b) - An example of intermediate pattern.

A problem is also envisaged if the segmentation of picture results in a set of discriminating arrays which all become active for two (or more) different training patterns (i.e., patterns having the same building blocks). Here, again, by observing the way connections between the receptors in the eye and the visual cortex seem to be randomly established a similar principle is deployed to randomly separate the array into groups of elements, rather than organised divisions. The result is, opposite to expectations, to enable the system operate in a more organised and superior fashion.

This 'bottom-up' parallel approach to PR resulted in models which were very impractical to simulate on computers for any reasonably large number of elements. Hence, in the early 1980's, Aleksander and colleagues built (at Brunel University) a hardware model based on the above general blueprint. The model is a 512x512 resolution machine called WISARD, the earlier versions of which, the MINERVA machine and the JANSYS program, were built during the early 1970's. Wisard is capable of discriminating between different patterns (e.g., faces) after an initial training phase which involves feeding various configurations of a pattern to the machine.

Aleksander (1983) also discusses the possibility of devising more intelligent type machines. He predicts that such machines could be created by combining the above single-layer type PR systems, that are powerful in discriminating or generalising patterns, with 'learning' automata, that are capable of inductively 'learning' state transition rules - using some form of feedback. It is also envisaged that autonomous machines could be designed that are not 'taught', but directly interact with their environments. Introducing 'feedback' to such systems could result in the strengthening of distinctions made between various patterns (subjects of discrimination); and, also, it can be said that a sort of 'short-term memory' has been manifested - by signals persisting within the system, even in the absence of their source.

Furthermore, by defining the response of a system in terms of a mixture of its present and previously occurred inputs it can be said that now the model has a non-trivial 'state-structure', and, hence, it can be thought of as a rather more complex type of 'automaton'.

By implementing (hypothetically or practically) these new ideas in the basic WISARD machine, its designers have been able to bring more complex dimensions into its behaviour. Images could be memorised and recalled, simple associations between patterns could be established, and new schemes could be devised for storing these images and associations. A single large discriminator is deemed to be able to carry out the job of 'learning', for example, the associations of names and objects; and can even carry out generalizations, such as relating a variety of different expressions to a particular face. Moreover, by adding feedback, sequence of patterns could be 'learned, whereby, giving one segment of a sequence to the machine as a cue, the whole of the pattern could be recalled. In its most advanced form, WISARD is given some capabilities for control of its environment, in particular, the control of its 'eye' (camera) movement.

However, it must be pointed out that this model and its particular design features (feedback, etc.) have basically been developed to tackle a class of practical PR problems in a more proficient manner, and, therefore, references to incidental similarities with the natural recognition systems, or other higher level explanations, should be appraised cautiously.

(iv) - OTHER NETWORK-BASED MODELS

The work of neural-net and logical-net modelers was a prime impetus behind the design of many PR systems; and the mathematical analysis and synthesis of logical threshold elements and functions lead to the introduction of analytic cellular systems, parts of which were modifiable for recognition purposes.

Various network-based models have been developed, either in abstract or in hardware (mainly electronic), to emulate some PR capabilities. For example, George (1973) discusses a hypothetical general model for perception (visual) which could be based on neural networks. Various schemes are proposed as to how the specific characteristics of human visual system (colour-vision, etc.) could be manifested in such models.

Rosenblatt in the early 1960's developed his hardware 'Perception' machine, modelling the basic workings of the human visual system. A simple conglomeration of input, output and associative cells together with adjustable weighting elements was able to 'learn' to recognise very basic patterns, by following three general rules of:-

(a) - increasing weights of non-active cells for incorrect recognitions
(b) - decreasing weights of active elements for incorrect recognitions
(c) - making no changes for correct recognitions

Culbertson's (1963) simple image processing model was also based on neural-nets. His model could recognise trivial shapes (e.g., squares, triangles) by a process of matching with a standard template of these patterns. The input pattern received by the 'retina' of the system underwent transformations (i.e., 'linear', 'dilation', 'expansion', and 'rotation'), this enabled the comparison of the unknown input pattern with different configurations of standardized templates.

In pandemonium and perceptron the emphasis was on adjusting the weights of a predetermined set of feature detectors. An alternate method, originally proposed by Uhr and Vossler (1963), was to look for 'good' features which could give rise to a suitable weighting rule. They devised a PR system that could develop its own features, by generating and testing the validity of new features on the basis of its discriminating value. It was, thus, possible to generate some useful and novel features which even the designers of the system had not anticipated.

The feature-extraction process of Uhr and Vossler's model (briefly discussed in section 5.1.7.(e)-i) entailed scanning along and across the unknown pattern with a 'feature-detecting operator', and determining 'matchings' with this operator at each region of pattern. The operators themselves were selected/generated by experiments. A 'learning scheme' was devised at the classification stage of the process. Whereby, a 'heuristic' method modified (on hedonic basis) weights of feature vectors for each 'pattern-class', using a 'difference-score' (i.e., similarity to a range of previously stored patterns) for an unknown input. This well elaborated functioning PR system established a very high standard for the category of models it represented, yet in spite of its 'learning' capabilities was only able to operate on a limited and simple range of isolated patterns.

Other classes of models, based on the neural-net and logical-net models of chapter four, have also been developed to depict various information processing aspects of the process of PR in the brain. Their designers, normally, characterise neurons as on/off devices connected by 'weighted synapsis', and arranged in certain conglomerations (networks). An interesting and powerful feature of these networks is that, by changing their synaptic weights, the behaviour of the system could be modified in a useful way, and, in some cases, 'learning' is said to be manifested.

In the 1960's, numerous researchers were involved in the elaboration of such logical PR network systems. For example, Singer (1961) proposes an electronic model for a 'learning' character recognition system, based on the human visual processes. The emphasis is on implementing a size invariance for the patterns of an object, as the angle of view or distance from an object changes. The technique used involves introducing a mathematical lattice in the form of a matrix of polar coordinates, upon which the transformations of input patterns take place. Various forms of alpha-numeric characters are tested in typical recognition tasks - whereby, the number of coincidences of outward transformations of an initially 'centred' patterns with previously established templates determine the class of a pattern.

A wide class of abstract PR models are also labelled as 'adjustable-weight majority logic' systems. These systems (perceptrons being some special examples) are designed on the basic premise that their inputs should be connected to the system through adjustable weighting factors. The recognition of patterns is achieved by running the system (or machine) through a distinct 'training' phase, where adjustments and reinforcements of weights are carried out, generally by human operators. The procedures involved should ensure a convergence to a final trained state in a finite number of steps; however, the existence of such procedures is a prerequisite to its convergence analysis. The separation of 'training' and 'operating' modes of such PR models has limited their potential; and, unlike the natural recognition systems where the inter-linked processes of training and operation are continuously active, their adaptability is rigid, and only ensures adequate performance for fixed environments.

An early version of such systems was the 'Adaline' (ADAptive LINear Element) based networks, introduced by Widrow (1962, 1973). Adalines were fundamentally different from the basic neural elements proposed previously. Each unit in a single layer Adaline network, as depicted in FIG.5.10., had independent weight adjustment provisions for its inputs to the threshold section. Basic Adalines were constructed in hardware and also implemented on digital computers. The weights were modified, either continuously or in discrete steps, according to some performance criterion. In the hardware models these adjustments were done by electro-chemical means, however, later other models used alternate electrical means for modification.

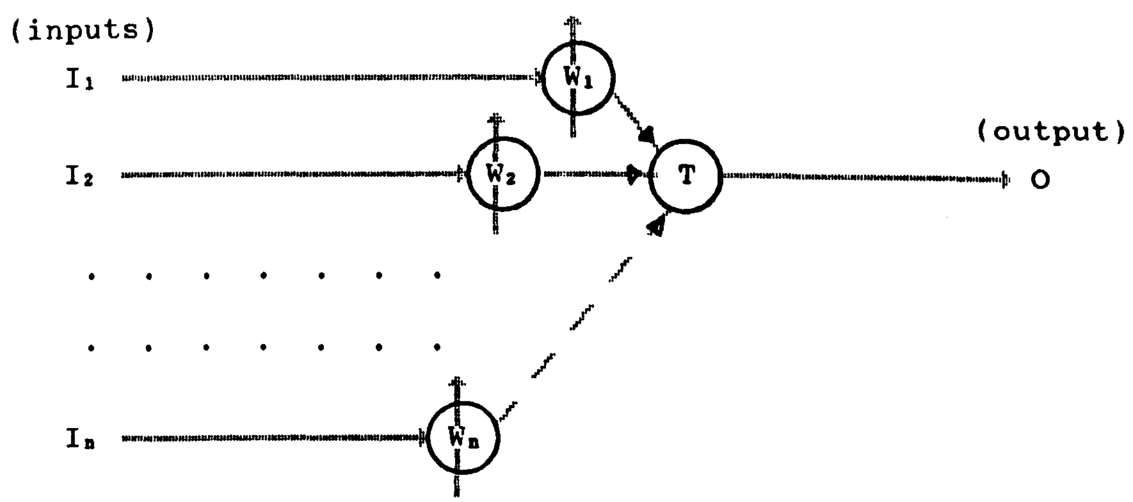


FIGURE 5.10. Diagram for an 'Adaline' element having the threshold value T.

For a more functionally complete configuration a two level unit, the so called 'Madaline' (Multiple ADAptive LINear Element), was developed; yet, the adjustments were only directed towards the first level - here, the close similarity of these systems and some SOS or logical-net systems (discussed earlier) should be recognised.

In these more complex systems the problems of simulating parallel networks using inherently serial or sequential machines and computers were magnified many folds, and posed much complications. In particular, many limitations of perceptron type devices were exposed by an exhaustive and thorough analytical investigation of the paradigm by Minsky and Papert (1969) in the late 1960's, which resulted in a significant recession of interest along these particular lines.

Studies of the behaviour of 'cellular-automata', as a sub-section of automata-theory, have been a useful source of theoretical ideas for the designers of network type PR systems. This discipline has provided some important insights into the computational powers and limitations of networks in general; and has also greatly helped the designers of parallel computers and machines. A cellular-automata can be defined as any collection of cells (or elements) arranged and interconnected in some regular organization (in two or more dimension). A simple example of a cellular-automaton (also mentioned previously) is the realization of "LIFE" transformations which has been investigated by various workers. Von-Neumann's 'self-reproducing' automaton was another example of these machines, and many of its concepts have been used in various modelling exercises (both mathematically and in hardware). Arbib's cellular abstract machines could also display many interesting computational capabilities.

(v) - STATISTICAL/MATHEMATICAL 'LEARNING' PR MODELS

Mendel and McLaren (1970) discuss various aspects of the application of the learning process in PR systems. 'Learning' is said to take place when the system's past experience can effect an improvement in performance. This implies that for a given decision a feedback of information about the 'performance' of the system should have taken place. 'Learning' PR systems are classified as 'parametric' or 'non-parametric', 'supervised' or 'non-supervised', 'on-line' or 'off-line'. Various 'learning algorithms' are considered, especially those related to some mathematical learning theories developed in psychology. The principle of reinforcement is incorporated in the design of a class of algorithms, the so called 'reinforcement-learning algorithms', and applied to different PR problems. It is found that the probability of occurrence of outcomes could be updated favourably by changing the values of decision parameters (or certain other parameters) of the 'classifier' on basis of a reinforcement criterion. Similarly, reinforcement algorithms could be applied to feature-selection.

Furthermore, 'learning' PR systems are considered as 'goal-seeking' systems, and various system-environment-interaction assumptions or other issues (e.g., 'linearity', 'memory', etc.) are discussed. Comparisons are made between such systems and 'learning control systems' (discussed in Chapter 4); and also many equivalences are established between the elements of a proposed 'reinforcement-learning' PR system (using a stochastic automaton model) and the elements of psychological stochastic learning theory.

5.1.10 AN OVERVIEW OF PATTERN-RECOGNITION APPROACH

On the whole, the discipline of PR has been characterised by three distinct approaches. The distinctions approximately reflect the three major developmental periods of this subject, as signified by the relative popularity of each viewpoint within the past three decades. These approaches are:-

(1) - 'The neural approach': characterised by pattern-matching models which are effective in rigid and exact domains, yet, are very inflexible and impervious to small variations of inputs.

(2) - 'The engineering approach': an elaboration of neural approach which involves defining further characteristics of input patterns. Various mathematical methods and combinatorial techniques are developed and utilised.

(3) - 'The A.I. approach': models in this more recent and flexible approach are, mainly, concerned with definition of features, and description of relationships of features. They are, generally, related to some human cognitive processes, and their designers try to incorporate some perceptual notions such as 'knowledge', 'meaning', 'understanding', etc. into their designs. However, learning from experience is not featured to the extent it was considered in the earlier paradigms.

Many scientists have voiced their criticism over the engineering only or abstract only view of this subject. The stalemate confronted, when pursuing a too narrow an approach, is the main witness for their argument. It is also felt that possibly more fruitful, and universally applicable, results would be obtained if first the nature of the process of PR could be better understood in humans and animals.

As far as the human PR is involved, in spite of a great deal of work on the subject, it is still not clear yet how this process is exactly manifested, or what its underlying mechanisms are, and many of the principal questions raised in the early days of this science have still not been answered adequately.

In a critique of the current trends of A.I., Dreyfus (1972) rejects any similarity between the way, presently, humans and machines recognise patterns, and attributes the slowness of progress in this field to ignoring such similarities. In his view, PR models which perform their tasks on the basis of: normalization of inputs, by means of transformations; detection of discriminatative features, by means of a decision tree-search or a statistical choice procedure; or recognition of resemblances; cannot be equated with human perceptual mechanisms. He also speculates on the nature of some properties which characterise human PR, and proposes that:-

"Any [artificial] system which can equal human performance, must therefore, be able to

(1) - Distinguish the essential from the inessential features of a particular instance of a pattern;

(2) - Use cues which remain on the fringes of consciousness;

(3) - Take account of the context;

(4) - Perceive the individual as typical, i.e., situate the individual with respect to a paradigm case."

Efficient technological tools for PR which do the actual sensing of the outline of an object have been available for some three decades - mainly in the form of hardware devices which translate an image (obtained by TV-cameras or otherwise) into an array of firing or non-firing elements.

Many computer or robot-vision systems have also been developed within the past two decades, and the prospect of extensive industrial application has put a great deal of emphasis on research in this subject. In recent years, researchers on PR in partnership with workers in Robotics have introduced a new specialised dimension to machine recognition. However, models devised in this principally industry-oriented area are quite distinct from the earlier endeavours of PR workers.

Hence, while the progress of the technological side of the research, that of designing sensory mechanisms or faster and more powerful computers, has been very spectacular, research on the inference making aspects, which involves the information processing level, has not shown the same degree of achievement. The main endeavour is, therefore , to devise an appropriate general classification system to operate on available data obtained by hardware peripherals.

Today, the initial optimism of computer scientists have been substituted by an appreciation of the complexities of the problem. Some of the principal difficulties and stumbling blocks encountered are: (a) - devising appropriate algorithms for three-dimensional vision; (b) - separating and classifying different objects in a scene; (c) - preserving the constancy of size of an object within different percepts. In addition, some proposed sophisticated algorithms for real-time scene-analysis require computing speeds and data handling capabilities far in excess of present computers to function in non-trivial environments. The use of inherent 'continuities' of patterns is one possible way of limiting the need for such extensive computing power.

When assessing the performance of a PR model, the consideration of 'meaning' and 'intentionality' aspects incorporated within the design of the system is extremely important. We should determine how much the system is capable of 'understanding' - whether it operates blindly according to a preprogrammed set of rules, or works towards a specific 'objective'. Similarly, we should enquire if it makes 'generalizations' or 'inferences' (if any) on its experiences.

Many important issues have also been raised by considering distinctions between 'serial' and 'parallel' processes. Although parallel processes could be simulated (but more slowly) by sequential machines, such as digital computers, recently, a great deal of attention has been focused on the design of parallel computers and machines. Typical reasons cited by designers of such machines are: the limitations of serial machines; and also the desire to incorporate and simulate some aspects of parallel biological neural systems (such as redundancy), which are deemed to possess the optimal forms of information processing capabilities.

The various parallel PR models elaborated so far have shown that, indeed, machines could be built to display some facets of intelligence as a consequence of their particular structure. In particular, they could have the added capability of learning from experience. Unlike most preprogrammed machines of A.I. where systems involving 'learning' either rely on the 'learning' of sequences of patterns or the rules which bring about such patterns. In this light, such PR machines (and systems) should be regarded as challenging alternatives to the algorithmically based models that presently dominate the field of machine intelligence.

However, the parallel PR models (based on networks) which function quite proficiently in single-layer configurations, dedicated to simple recognition tasks, have not been very successful when extended to higher order layering - where there is feedback and interconnection between elements. The main reason is that the complexity of tasks, and the processes, which these higher order systems are trying to convey are of a different degree of sophistication. Most proposed higher-layer solutions have, hence, involved limiting or defining new constraints on the range of acceptable patterns, or adding extra structural features to conform to a particular design specification. Thus, suggesting that, perhaps, models which are originally designed for achieving simple engineering objectives, or are solely based on particular simplistic hardware configuration, will not provide an adequate vehicle for representing complex 'learning' PR processes - their potential will be too limited.

5.2 THE ARTIFICIAL-INTELLIGENCE or "TOP-DOWN" APPROACH TO 'LEARNING' SYSTEMS

In this section we will attempt to survey the approaches to the modelling of learning from the, so called, "top down" view of the subject. This trend of enquiry, which is generally accepted as the most recent, is normally labelled as the 'Artificial-Intelligence' (A.I.) approach; but, in some instances, also referred to as the 'Information-Processing' or the 'Machine-Intelligence', approach. A dominant (and distinguishing) feature of all models devised in this area has been the importance of role of computers in their manifestation and representation. In fact, if disciplines such as A.I. were devoid of their 'computer angle', then, it would be difficult to differentiate between many of its research fields and some areas of disciplines such as cognitive-psychology.

Although, in some of the following specialised subjects, such as the 'evolutionary programming' approach, the 'connectionist' approach, or some cited examples of generalised models of learning, a kind of "bottom up" tendency is prevalent. Nevertheless, because of their predominant reliance on the use of computers in design, and the relative recency of their appearance, these topics are all included within this final categorization of 'learning' systems. But, they should really be regarded in more independent terms; even, in some instances a closer kinship with our other previously covered paradigms could be established.

Additionally, because of the diametric distinctions in emphasis and approach from the rudimentary cybernetic undertones of our thesis, here, the A.I. 'learning' models, will be discussed and analyzed fleetingly - without scrutinising their various underlying concerns, or typical examples, to a great extent. However, the large collection of literatures and papers cited in the reference section of our thesis can provide details of many research endeavours on this type of 'learning' models.

5.2.1 THE SCIENCE OF ARTIFICIAL INTELLIGENCE

In some of our previous chapters, we discussed different aspects of the discipline of A.I.; indicated its characteristics; and considered some typical examples of its work.

The discipline of A.I. is usually defined as the science of 'intelligent' behaviour by automata and computers for the purposes of study and simulation of human intelligence, or construction of bigger and better automata. The principal techniques used in A.I. are 'introspection' and 'intuition'. The various mathematical and computing techniques used are, generally, considered as secondary in importance, and non-fundamental to the methodology. Philosophy, Linguistics, and Cognitive Psychology have been some of the major contributors to A.I.

Ever since its introduction, the subject of A.I. has entertained a great deal of self-analyses and self-criticisms regarding its 'validity', 'justification', and 'practicality'. The principal phenomenological difficulty of A.I. is that there is no clear cut definition of intelligence (human or otherwise). The normal resolution of this difficult 'analysis' task has been to breakdown the notion of intelligence to some more manageable components; and redefine the problem in terms of concepts such as 'problem-solving', 'theorem-proving', or some other currently pursued subtopic of A.I. Yet, paradoxically, it is also possible to envisage all these secondary notions in terms of 'non-intelligent' behaviour, as attested by a variety of 'clever' programs of today. It is also true to say that although great many questions are asked about the nature of intelligence, nevertheless, not many researchers dwell upon the viability of equivalence of an 'intelligent' and 'non-intelligent' behaviours.

Computers are made to exhibit more and more complex tasks, yet, once the exact means of achieving these tasks are known they are no longer given attributes such as 'creative', 'thinking', 'free-will', etc. However, unless all these attributes are defined precisely, or machines could be constructed which can exactly imitate all human mental processes, this dilemma will continue.

Today's endeavours in A.I. are mostly directed towards particular task domains. Models devised, normally, have some aspect of human cognition incorporated within them; and the realization and manipulation of such models occupy A.I. scientists to a much greater degree than 'pure' analytical studies. However, the actual synthesis of 'intelligence' was not the primary impetus of this paradigm; but, it was a desire for a better understanding of the human mental functions which initially instigated the scientific quest in the field of A.I. - hence, its kinship with cognitive-psychology.

5.2.2 COGNITIVE-PSYCHOLOGY and ARTIFICIAL INTELLIGENCE

Cognitive psychology is a very broad subject, covering disparate areas, which at times make it very difficult to represent its views uniquely. "Intelligence" and "cognition" are also considered as closely related notions; and, hence, by and large, the task of simulation of cognitive processes is taken up by A.I. researchers involved in such aspects.

In cognitive psychology the human brain is vaguely regarded as a kind of "information-processing machine" (more or less like a computer); with elements such as 'sensory receptors', 'effectors', 'memory store', and a 'central processing' unit. As we have seen previously, "perception" and "memory" are two of the central pillars of cognitive studies. Hence, the majority of the cognitive models of human learning have revolved around these two issues. Particularly, the concepts of 'Long Term Memory', 'Short Term Memory', or 'Working Memory' have been utilised in many such models.

Principal concerns of cognitive psychologists have been the ways which memory is organised during the learning process, and the strategies used for storing data - not the actual recording of data. For example, the information accumulated in LTM (regarded as main source of learning) is categorised into: (a) - 'Sensory-Perceptual' knowledge, which contains information extracted from sensory inputs; (b) - 'Procedural' knowledge, which contains sets of rules, or stimulus-response rules of association; and (c) - 'Propositional' beliefs, which contain value judgements or beliefs about subjective truths.

In cognitive psychology the general view is that relationships are more important than individual elements; and that cognitive mechanisms are a kind of filters on percepts. The notion of "perception" is referred to as the process of interpreting the stimuli encountered by sensory systems from the environment. Additionally, cognitive psychologists define the process of "learning" in terms of changes brought about in the cognitive systems; and consider the system's structure (not its elements) to be the critical agent in bringing about such changes.

Simulation models of human learning which involve 'memory formation' have been devised both for the developmental stages of the growth of a child, and also for the adult state of humans. The investigations of these models have, normally, featured aspects of memory such as 'forgetting', 'recall', 'rehearsal', 'information coding', etc.

The nature of architectural principles of human mind, and the investigation of similarities (and differences) of such structural organizations with their artificial counterparts in abstractions, programs or machines are other major issues in cognitive psychology. Some workers have attempted to explain the various cognitive processes in a unifying fashion, contending that there are equivalent principles at work. However, others have proposed the so called "modularity hypothesis" in explaining various facets of cognition; whereby, each cognitive function is contended to be governed by distinct principles. For example, workers such as Chomsky or Marr have tried to theorize about the faculties of 'language', 'reasoning', 'problem-solving', or 'vision' in an independent and externalised manner.

5.2.3 'NATURAL' vs. 'ARTIFICIAL' ASPECTS OF INTELLIGENCE

Two basic fundamental assumptions are made in A.I. Firstly, the faculty of human "intelligence" is regarded as an 'universal' phenomenon, independent of culture or individual variations. Secondly, it is assumed that we know enough about this natural process to consider its simulation. However, both assumptions can be said to be over-optimistic simplifications.

At the beginning, a proficient chess-playing program was considered as 'intelligent'. But, today, the common view is that once a program manages to achieve its goal then it no longer is seen as 'intelligent'. If "intelligence" was simply defined on the basis of apparent similarities with human intelligence (as is the case in 'Turing test for intelligence'), then, it could be argued that each instance of such exhibited trait in a machine is only a case of mimicry. Yet, diametrically, it could be argued that any dissimilarity with human intelligence could be detected, measured, and hence incorporated in such machines; leading to the narrowing of the gap between the 'natural' and the 'artificial' manifestations of intelligence.

An important theoretical discovery by Turing was that: "For any deterministic automatic formal system whatever, there exist a formally equivalent Turing machine". The implications of this theory was that no automatic system can do anything that Turing machines can not do, and also that the Turing machine is the only automatic system we would ever need.

In the same abstract sense, Gödel's theorem, which states that: "In any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless the system is inconsistent", has been cited as an argument against the possibility of constructing truly 'intelligent' learning systems or artifacts, without encountering inherent inconsistencies. Yet, many workers in the field of "machine intelligence" have rejected this type of reasoning on bases of two arguments; firstly, human intellectual powers themselves are not free from fundamental errors or inconsistencies; and secondly, it would be possible to construct 'open' artificial systems to which such objections would not apply.

The above arguments and counter-arguments are typical of the disciplines which involve the modelling of 'natural' aspects of human mental faculties, such as learning or intelligence - course of their developments characterised by numerous mind/body or man/machine controversies and debates. Here, we will not attempt to linger upon the controversies and arguments which the meaning and the definitions of "intelligence" has brought about. Since, this kind of philosophical altercations about human mental attributes sometimes regresses towards highly ambiguous debates, and will not be scrutinised deeply. The notion of intelligence and its facets (e.g., learning) will be considered as natural properties of living animals, whose descriptions (at some level) are the bases for their implementations in artificial systems.

5.2.4 COMPUTERS and ARTIFICIAL INTELLIGENCE

Computer is the principal tool used in A.I., because it is currently the only usable logical manipulative system. Ever since its introduction, the computer has been attributed with 'living' characteristics, and often the scope of its potential misjudged. As an 'infant progeny' of technology, questions are asked about the possibilities of its future developments. What will happen when this 'infant' grows up? Will future computers be our masters or slaves? At what stage in its development can we regard them as 'intelligent'?

History has shown that every 20 years the computing power has increased by a factor of 1000; and while their future technical enhancements have been consistently underrated, their scope of capabilities have been overrated. It is generally envisaged that within the next 40-50 years truly 'intelligent' ('thinking') computers can be built which are able to compete with humans, and solve many of their intellectual problems. These 'intelligent' or 'super-intelligent' computers will, probably, not be thinking as the humans do, and might have very little in common with us; but, indeed, it is possible to imagine a coexistence with such machines. Although, the actual pathways of these developments are not clear, and most research is currently focussed on digital computers; nevertheless, some alternate possibilities are also scrutinised. For example, self-reproducing computer architectures have been investigated; or parallel processing structures, that do not rely on absolute computing powers but on the intricacy of their connections developed.

Computers are presently in their fifth-generation, with blue prints of their next two generations already sketched. The essence of today's 'intelligent' computers is in the way they are able to use a data-base of information (or rules) in conjunction with an inference program, and interact with a human operator. Some are able to 'talk', 'voice-operate', 'read', 'diagnose', 'translate', etc. Others employ alternate computing architectures (e.g., parallel) or languages (e.g., LISP, PROLOG, STRIPS, PLANNER) as a more efficient way of dealing with particular problem domains. In any case, normally, the questions posed about the level of 'intelligence' of a computer program, such as a chess playing program, should be redirected towards its level of 'performance'. Since, it is the evaluation of the performance of the program which signifies whether it is behaving intelligently or not.

Additionally, many attempts have been made to 'simulate' the various aspects of intelligence and learning on computers. For example, Friedman (1967) describes an elaborate computer simulation of instinctive behaviour, based on traditional psychological theories; Findler and McKinzie (1969) apply a general computer simulation technique to a wide range of goal-seeking biological phenomena, such as learning and self-preservation; Kent's (1978) work concentrates on the computer modelling of the brain's neural mechanisms; and Albus's (1979) computer models try to depict the cognitive functions of the brain.

5.2.5 KNOWLEDGE REPRESENTATION and ARTIFICIAL INTELLIGENCE

Considerable work has been done in the fields of A.I. and cognitive psychology on the problem of representing knowledge. One of the most common methods is to represent knowledge in terms of 'propositional' relationships. However, work has also been progressing on other modes of knowledge representation, such as 'linear ordering' or 'categorization'.

Some knowledge representation tasks rely on computing or A.I. high level languages (e.g., PASCAL, LISP, PROLOG). However, many of the cognitive structural models use "production systems" for organising their knowledge bases. Production systems have also long been advocated as a suitable medium for modelling learning. For example, in acquisition of skills, language, or in development of reasoning. Additionally, many algorithms have been devised for extracting rules from data, some by organizing information to recognize patterns; others by more formal means of reasoning or induction.

One of the knowledge intensive fields of A.I. which has attracted some learning related research, and is deemed to benefit enormously if 'learning' was to be incorporated within its programs, is the area of 'Expert Systems'. Expert systems are currently limited because the process of extracting appropriate knowledge is the arduous task of their programmers; and also because their knowledge bases, normally, do not change (and improve) with experience, unless, their human operators add new information to them. A recent trend in expert systems has been to develope 'learning' programs that are able to operate on a data-base and "grow" richer in context. Typical elements of such programs are: a 'descriptive language'; an 'interpreter'; an 'associative memory'; a 'generalization procedure'; and a 'learning strategy'.

5.2.6 SCENE-ANALYSIS (PATTERN-RECOGNITION)

Various 'scene-analysis' models have been devised in A.I. to deal with quite complex real situations, specially, in limited problem domains. Objectives of such endeavours have been to find the 'existence', the 'description', the 'orientation', or the 'position' of an object of interest in a natural background. The more global objective of defining 'real-world scene analysis strategies' has, also, been pursued. Yet, in reality, many 'general' aspects are sacrificed so that working models could be devised; and applied to practical problems of 'detection', 'location', or 'navigation'.

An example of this type of model is described by Bullock (1976). Whereby, a generalized theoretical scene-analysis strategy is applied to the more specific problem of finding an object in an outdoor scene, by identifying its position and orientation. The necessary trade off and simplifications required for such a task is discussed, and various experimental results scrutinised. In this model the features are analyzed in three levels: 'point' (pixel intensity), 'local features' (lines, edges), and 'global' (shape descriptions). Additionally, some well established 'feature-extraction' techniques are utilised in the model.

The two principal approaches to the problem of 'feature-extraction' are: the 'statistical/mathematical' approach, and the 'structural' approach. In statistical feature-extraction, normally, a mathematical performance measure evaluates whether a particular feature should be selected or not. These type of features may not have a real physical meaning, yet, they represent an efficient and computable set of attributes of patterns. Such techniques are particularly suited for 'noisy' patterns, or when no obvious organisation is present in the pattern. However, there are, also, many difficulties associated with such techniques. For example, the 'availability', the 'number', or the 'quality' of samples used during a 'learning' phase severely effect the choice of features selected. Similarly, there are various problems involved in segmenting images into parts; or devising nonoverlapping independent features.

On the other hand, the 'structural' feature-extraction preserves all the richness of contextual information, and uses the close topological relationships of parts of a pattern, or various prior knowledge, in the task of selecting features. The weakness of this approach lies in the absence of generalised mathematical techniques for abstracting structural features. Therefore, it is, by and large, the intuition of the designers of such systems which governs the choice of features, from the endless list of possible contextual features of a set of patterns. Chen (1976) discusses the relative merits of the above two techniques, and argues the case for a mixed (structural plus statistical) feature-extraction system, which is contrary to most current efforts.

5.2.7 LEARNING and ARTIFICIAL-INTELLIGENCE

A typical view of learning from the A.I. perspective is expressed by Arbib (1970) in his analogy of computers and brains:-

> "We can gain much insight into the process of learning if we think of past experience as providing a repertoire of programs of activity, which become units out of which new routines are to be fashioned by further learning - even higher order routines may then become the units of later learning."

In behavioural sciences, such as psychology, learning or adaptation are regarded as a kind of fundamental perpetuating force which give rise to all behaviour, more or less, as the sub-atomic forces are regarded in physics. However, when we look at the direction which current computer related research, such as A.I., has taken we see that the prominence of this central phenomenon has been lost. No longer it is deemed necessary to look at behaviour the way it manifests itself in nature, starting from a raw state and gradually developing to its fully fledged form - using experience, education, instruction and introspection on its course of development.

In these new paradigms 'knowledge' itself, and not the way it is 'acquired', has become the domain which has attracted the majority of research. This tendency is in support of the belief that the scrutiny of knowledge structure is a less cumbersome endeavour than finding out how it was formed. The metaphor we can compare to here is to look at the structural organization of leaves and branches of a fully grown tree and try to predict its future changes on such basis, without considering its botanical mechanisms and developmental processes. In any case, the complexities encountered in trying to tackle the structural knowledge on its own, even at a simple logical common-sense level, impels us to believe that the isolation of 'knowledge' and 'learning', and their separate analysis, is not a viable premise for the task of understanding and investigation of human and animal learning processes. Although, the main reason for the apparent slowness of progress in designing machines that can act 'intelligently' is the complexity of the task. Nevertheless, recently, even some of the proponents of A.I. have come to appreciate the need for reactivation of 'learning' as a central research topic - Schank (1983).

The process of learning has been categorized in numerous ways by workers in the subject of A.I. These classifications, in some sense arbitrary, have involved distinguishing between 'rote learning' (e.g., learning by memorization, or mimicking), 'learning by example' (e.g., a scene-analysis program learning from samples), 'learning by being told' (e.g., a natural language program making inferences on a knowledge base, or extracting 'meanings'), 'learning by doing' (e.g., a game playing program learning from its previous mistakes), 'learning by analogy' etc.

A.I. or information-processing models of simple learning behaviours (mainly in animals), such as 'conditioning' or 'trial-and-error' learning can easily be implemented. The precision of these type of 'simulations' can, indeed, be increased, by elaborating the model into one which depicts the empirical observations more faithfully. Yet, workers in A.I. are seldom interested in the simple modalities of the learning process outside the human domain.

Cognitive psychologists were occupied with devising models of human learning processes long before the advent of computers brought the A.I. and information-processing models to the forefront of this paradigm. Later, a variety of A.I. computer based models were devised to exhibit some higher 'learning' capabilities within particular domains (e.g., language, expert-systems, problem-solving, etc.). Similarly, various hardware models were elaborated to display some 'learning'. But, always a crucial question has been asked: whether such systems are 'intelligent' or just 'clever'.

Some of the early hardware systems bordered closely with 'adaptive control' concepts, and the 'learning' interpretation, mainly, depended on the descriptive level the model was looked at. However, the later hardware models (e.g., robots, turtles) were designed in conjunction with the knowledge base of a computer, and could show much more advance 'learning' capabilities.

Although, 'learning' models in A.I. display a wide range of interesting behaviours, nevertheless, they almost universally are dependant on a "teacher" or "guiding" external element. Firstly, to evaluate their actions; and secondly, to provide procedures for development or initiation of 'learning'. Very few are in the natural sense oriented by basic 'drives' or 'needs'; even the 'goals' defined are, generally, non-elementary, and closely relate to the task in hand in an ad-hoc fashion. In a sense, their artificial portrayal of 'intelligence' or 'learning' will be only significant to a subjective human observer, having no connotations for the machine domain. In the following the various sub-divisions of A.I. will be outlined in a summary form. The emphasis will be governed by the relevance to our underlying fundamental cybernetic bias of the thesis.

5.2.8 GENERALISED 'LEARNING' MODELS IN A.I.

Here, we will briefly describe some A.I. work which do not concentrate on particular facets of intelligence or learning, but, provide a framework for analysis and simulation; or a methodology for synthesis, of much broader range of processes and behaviours.

- Friedberg (1958, 1959) describes a learning procedure used in a computer program, and attempts various experimental implementations. However, this program was more akin to the previously discussed cybernetic learning systems; and its approach was general and non-task dependant.

- Hormann (1962, 1964) also describes an abstract scheme for machine learning; 'tasks', 'problems' and other aspects of the 'learning' system are discussed in a generalized fashion. But, a specific problem is also addressed as an illustrative case.

- Andreae (1964...) (and co-workers) is one of the principal researchers in the field of 'machine learning', but his approach has not followed the mainstream of A.I., and has been, generally, treading an interdisciplinary line. His comprehensive coverage of the subject has traversed an evolutionary developmental path. Ranging from very abstract attempts at the unification of concepts of learning machines, to elaborate discussions of underlying philosophical issues, to practical implementation problems. The initial endeavour in design of a 'learning' automaton involved the construction of a simple mechanical tortoise. His general purpose learning scheme, STELLA, was based on 'pain-pleasure' reinforcements; and was realized by a hard-wire machine. Later, a symbolic manipulative dimension (a monologue) was added; and various conceptualizations, abstractions, theoretical-analysis, and computer simulations were undertaken. Further developments of this model was the PURR-PUSS learning machine, which could 'learn' by a teacher specifying patterns and actions; yet, it was still based on rudimentary learning criteria. The software devised were tested for various task classes, including robot problem solving.

5.2.9 GAME-PLAYING

Turing (1953) had suggested that games were prime domains for experimentation on design principles in the quest for intelligent machinery. Some of the consequent models devised have, in fact, demonstrated (with or without 'learning') how close they come to our definition of intelligent activity - chess or checkers playing programs.

Cybernetic machines and artifacts which display interesting 'game-playing' characteristics have been constructed for many centuries. Yet, it was the advent of computers which changed the emphasis from pure 'mimicking' of game playing behaviour to the mechanization of underlying thought processes.

In some instances, the mental skills required for the playing of simple games, such as noughts and crosses, can optimally be programmed into a computer. But, for less trivial games, like chess or checkers, a higher degree of elaboration is necessary, to depict the 'planning' and 'reflective' aspects of the human thinking. The game-playing programs and machines developed within the past few decades have managed to accomplish a high degree of proficiency in their particular task domains, some achieving competence standards comparable to an expert player.

One of the most prominent, and best thought out, early game-playing programs was Samuel's (1959, 1960) 'checker-playing' program. This program was able to display an intriguing 'learning' behaviour, and develope an expert performance level. Yet, its 'learning' was confined to parameter adjustments of some mathematical functions - following the trend of other contemporary work on adaptive control techniques. Also, a great deal of expertise and knowledge was primed within the 'non-learned' initial state of the program.

Samuel (1959) used an ingenious method for optimising the performance of his checker-playing program. Whereby, two similar programs were able to play each other, and upon three successive defeats a program would undergo a set of (fairly arbitrary) changes; hence, gradually converging towards its optimal performance. However, some later work has shown that for more complex systems neither the existence of such a convergence can be guaranteed, nor, the utility or efficiency of such techniques can be justified.

Samuel's checker-playing program involved many executive or house-keeping routines, but, its 'learning' features were incorporated within two 'rote-learning' and 'generalization-learning' routines. Rote-learning routines were used to store, organize, and evaluate past incidences; and generalization routines were used for choosing heuristics, and making weight adjustments. The 'utility' or the 'worth' of actions were chosen on basis of an "evaluation-function" (a multi variable polynomial with adjustable coefficients) which indicated the 'value' of a specific move in terms of a single real number. This type of function, expressing the 'value' in a single or multi-dimensional manner, has been a feature of many later game-playing, problem-solving, or heuristic-search programs. A comparative study of various 'evaluation' and 'learning' procedures is undertaken by Griffith (1974).

Another aspect of game-playing programs, such as Samuel's, is the 'search techniques' used for 'forward' or 'backward' analysis of possible alternate moves. The usual means of representation is in a "tree" format, each 'branch' indicating the choice of action, leading to 'nodes' where the evaluations of utility of actions is made. The task is to find the 'best' sequence of moves. For this purpose, a wide range of mathematical procedures, such as "pruning" or "minimax" have been developed. A broad analysis of the techniques used in game-playing programs is carried out by Marsland and Rushton (1974).

Yet, from our point of view the principal interest is the examination of the 'learning' aspects of this type of programs. The limited 'universe' of a game provides many advantages for a worker interested in tackling the problem of learning in machines. Firstly, games such as checkers, GO, or chess involve astronomically high possibility of movements (estimated at 10^{40} for checkers), which implies that deterministic techniques cannot be practically employed for their analysis. Secondly, they have a set of definite 'goals' and 'rules' which determines their behaviour. Thirdly, they can easily be programmed within a computer. Finally, they are widely familiar, and a good background of knowledge exists to verify the usefulness of their models. But, at the same time, the limitations of the game domains can lead to restrictions on the type of learning which can be investigated.

A survey of the work done in this area shows that, in fact, only few game-playing programs involve modifiable 'learning' or 'adaptive' components. Some 'rote-learning' or 'generalization' procedures developed in game-playing programs have also been applied to other real or abstract problem domains; for example, in solving control problems. Such tasks are sometimes described as "games against nature". Games are also excellent domains for utilising one of the principal tools of A.I., namely 'heuristics'.

5.2.10 PROBLEM-SOLVING

Many 'Problem-Solving' programs have been devised in other fields of A.I. (e.g., robotics); but, here, we will discuss some of the more prominent 'generalised' problem-solving techniques developed in this paradigm.

- Newell and Simon (1959, 1972) describe their 'General-Problem-Solver' (GPS) program which uses 'means-ends analysis' and various 'heuristic' techniques; in essence, based on human problem solving faculty. However, their system is general enough to be applied to other task domains.
- Doran (1968,1969,1970) develops a simulation program for a simple automaton-environment system; and also devises various algorithms and procedures which would enable the automaton, called the Graph Traverser, to display 'problem solving', 'planning' and 'generalization' features.

- Winograd's (1973) work on understanding natural language included a problem solving sub-system that was able to proficiently carry out many tasks within its limited domain. Yet, its techniques were general enough to be applied more widely.

- Fikes, Hart and Nilsson (1972) developed a well thought out methodology, STRIPS, for solving problems, particularly for use in robot problem domains, such as finding boxes in a room environment. Later, the learning and problem-solving schemes used were elaborated, and applied in broader contexts. Examples of such work are: Siklossy & Dreussi (1973); Sacerdoti (1973); Stepankova & Havel (1976); and Banerji & Ernst (1977).

5.2.11 KNOWLEDGE-DIRECTED CONCEPT/PATTERN/LANGUAGE LEARNING

A large proportion of contemporary A.I. research covers the problems associated with searching, analyzing, structuring and organizing of 'formal' or 'linguistic' knowledge bases; also, the majority of recent 'machine learning' work is, seemingly, directed towards these cognitive levels. However, since the class of learning processes targeted are at the top of the hierarchy of learning, unlike the domain of interest of our cybernetic approach, we will only engage in a brief discussion here.

Broadly speaking, symbolic description or representation can be viewed at five levels: 'message' (strings of words with no structure); 'syntax' (grammar or structure); 'memory' (formal notions of accessibility and organization); 'belief' (conclusions and inferences); and 'external' (subjectivity) level. Various knowledge structures place different emphasis on 'language', 'uniformity', 'consistency', 'labelling', 'accessibility', 'partitioning', 'growth', or 'change'. Some of the principal methods of such representations are: 'Plans'; 'Scripts'; 'Frames', 'Logical Nets', 'Semantic Nets'; 'Production Systems'; 'Predicate Calculus'; or other specialized representational languages.

Numerous Programs have been devised to work on a special class of data-basis and problem domain; and to 'extract', 'form' or 'learn' concepts. Similarly, 'language learning' programs have been designed to operate on semantic type knowledge basis. There are three principal methods by which concepts could be formulated. Firstly, by 'analysis', which involves applying existing concepts to events and processes in a new situation (customizing). Secondly, by 'induction', which means hypothesising about new concepts on basis of older concepts which are inadequate. Thirdly, by 'generation', which involves creating new concepts from more fundamental notions.

Some examples of research in this field are: Buchanan's (1978, 1985) DENDRAL, META-DENDRAL and MYCIN systems for, explaining empirical data and medical consultation; Soloway & Riseman (1977), Winston et al (1983), Torrance (1984), Connell & Brady (1985), and Phelps & Musgrove's (1986) knowledge-based 'pattern recognition' or 'description learning' systems; Winograd (1973), Brown (1975), Akama & Ichikawa (1979), and Gause & Rogers' (1983) [Kellerman,1972], knowledge-based 'language-understanding', 'language-learning', or 'question-answer' systems; and Hayes (1970), and Kochen's (1974) abstract 'learning' and 'problem solving' systems.

Finally, even a higher level of analysis of information contained within a data-base is undertaken in the 'Theorem Proving' or 'Automatic Deduction' research areas of A.I.; which involve drawing conclusions from a body of knowledge represented by logical statements, using deductive inference methods. This line of enquiry which was initiated from the original work of Newell and Simon (1959) into "Logic Theorist" has developed into many different application domains, such as 'logic programming'.

5.2.12 ROBOTICS

The introduction of term "robotics" is generally attributed to Czech play writer Karl Capek (term "robot" is a derivation of Czech word for 'workers'). During the past few decades, many unconstrained speculations of robots and their capabilities have been put forward by science-fiction writers and cinematographers. But, only when the advent of digital computers helped to realise these artifacts for scientific or industrial purposes was their limitations fully appreciated. Nevertheless, the science of robotics has suffered, perhaps irreparably, from the wild flights of imaginations which render most scientific achievements in this field as mundane or predictable, when compared to those seen or read in science-fiction films or books.

Various early cybernetic machines and robots have already been discussed in our thesis. But, during the past few years, many impressive achievements have also been attained in the science of 'robotics'. Intelligent robots, now, use powerful 'processing', 'manipulative', and 'vision' systems for perceiving and acting upon their environments; and, hence, have been utilized for solving many real-time practical problems. Some robots also use speech 'synthesis' or 'recognition' systems for ease of communication with their human operators.

Similarly, many educational or experimental robots have been designed in various research establishments. A state-of-art snooker playing robot has been developed in Bristol University, which incorporates many of the ingredients of industrial robotics research. Other 'walking' or 'hoping' robots, or 'mobile robot vehicles' have also been devised, and many aspects of control or dynamic characteristics of locomotion investigated theoretically. Another area of robotics research is 'transportation robots' and their networking control systems, for use in industrial manufacturing/warehousing.

Some of the principal application areas of robot research are those which involve hazardous and inaccessible environments, or monotonous and laborious tasks. For example, robots are widely used in industrial assembly plants, space research, oceanic research, nuclear-plants, or security applications.

A trend in robotics which is of most interest to us, and which has also been gradually becoming more utilised in industry, is the fusion of A.I. schemes with hardware robot realizations. The prevailing approach of the researchers involved with this problem (inherited from early cyberneticians) has been to construct 'mobile robots' for experimentation, verification, or demonstration of some hypothesis. Although, some 'hand-eye' configuration of manipulative vision systems have also been used.

Some research in robotics has clear biological undertones; and not only it is obvious that their protagonist's ideas originate from some natural sciences, but, in fact, one of their stated aims is the better understanding of the human and animal processes. Examples of this type of work are Friedman's (1969), Koplowitz and Noton's (1972), Kent's (1978), Filo's (1979), and Albus's (1979) models; in each case, robot 'learning' or 'intelligence' is discussed in terms of its equivalent biological underlying considerations. Yet, various engineering and A.I. techniques are also utilised in the work. A related area which explicitly declares this interlinking of natural and engineering concepts is the science of "bionics" - briefly discussed in an earlier chapter.

(i) - EXAMPLES OF EXPERIMENTAL MOBILE ROBOTS

In next chapter we will discuss our mobile experimental robot which was built as an exercise in designing a cybernetic 'learning' model. However, here, we will attempt to enumerate some similar hardware design endeavours, in each case describing briefly the main facets of exercise:-

- Rosen and Nilsson (1967) describe a mobile robot equipped with a TV camera and a simple retractable arm. The software implementations involve a hierarchy of computer programs which can solve simple problems, and devise 'plans'. Also, some scene analysis capabilities, and simulations of its environment are incorporated.

- Raphael (1968) describes one of the earliest, and best known experimental robot exercises, the Stanford Research Institute mobile robot vehicle (also referred to as SHAKEY). This model had an on-board TV camera and was connected by cable to a computer. Various software developments, involving different high level languages and A.I. techniques, were devised to demonstrate some obstacle avoidance or problem solving behaviours.

- Lewis and Bejczy (1973) discuss the planning considerations in the design of an elaborate autonomous roving type robot with a manipulator arm - such as those used in space missions. Principal software concerns are the control of the arm, and the navigation of the robot.

- Smith (1973) describes a basic implementation of a three-wheeled robot model, equipped with a sonar sensor and in radio-contact with a computer. He also outlines some executive control problems of the design.

- Heiserman (1976); Loofbourrow (1978); Gupton (1979) [separate works] describe detailed accounts for construction of 'turtle' type robots, with various electronic features (e.g., sonar, light-sensing, computer-control), that could be used for simple experimentation (e.g., tracking).

- Hollis (1977) describes a mobile robot equipped with a clasping manipulator and a simple image sensor. He, also, discusses some underlying design and control aspects of such models; and outlines a 'charger-seeking' program.

- Allen and Rossetti (1978) describe a fairly sophisticated mobile light seeking robot connected with a PDP-11 computer, which is able to track a light source while avoiding obstacles by use of a sonar detector.

- Marc, Julliere and Place (1980, 1981) describe a computer controlled mobile robot with a simple manipulator, a tactile sensory system, and an infra red position measurement mechanism. They also describe the software implementations, and schemes used for navigating and guiding the robot.

- Moravec (1982) describes a three-wheeled mobile robot equipped with a TV camera, and later to deploy some A.I. scheme (e.g., 'production systems').

In addition to the above hardware designs many researchers have been involved with developments based on mobile experimental robots with emphasis on the software aspects, examples are:-

- Nilsson (1969) describes some A.I. procedures devised for implementation of 'problem-solving' tasks in mobile robots, using "Q-A techniques"; also, outlines and discusses some other higher issues in designing intelligent mobile automata, such as 'theorem-proving', 'perception' and 'modelling'.

- Lasker (1974) engages in a theoretical discussion of the 'theory of mobile automata'; formulating an algebraic notation for the representation of robot movements and trajectories in its environment.

- Chan and Phillips (1975); Miller (1977); Thompson (1977); Giralt et al. (1979); Moravec (1979,1981); Shih (1982); Witkowski (1983); Thorpe (1984) [in separate work] describe research on various theoretical and software facets of the problem of navigation of mobile robots. The principal topics involved are: 'robot control', 'robot stability', 'obstacle avoidance', 'range detection', 'sonar signal analysis', 'visual mapping', 'simulation of environment', 'search techniques', 'path planning', 'decision procedures', and 'parallel route-planning algorithms'.

- Coles et al. (1975) devise a problem-solving program (in FORTRAN) for a computer controlled mobile robot, which uses 'decision-analysis' procedures.

- Rushby et al (1975) describe a simple computer controlled mobile robot equipped with a plotting pen and confined to a limited environment; and was devised to manifest the PURR-PUSS generalized 'learning scheme' introduced by Andreae (1972,1976). This work pursues a "non-engineering" approach, in essence different from most other A.I./robotics exercises; and is concerned with the more fundamental aspects of machine intelligence.

- Bond & Mott (1978,1981) describe research based on a computer controlled mobile robot, involving experimentation and development of a software language that can be used to implement a 'learning system' - established A.I. techniques and languages are utilized (e.g., SCHEMAs; PLANNER).

- Iijima et al. (1981); Elfes and Talukdar (1983); Prendergast et al. (1984) [separate works] describe and discuss various 'locomotion' and 'control' routines, 'control systems', and 'specialised robot-control programming languages' which are used in conjunction with mobile robot research.

- Kanayama (1983) describes a mobile robot control technique ('concurrent programming') which enables various tasks to be performed simultaneously, for use in perceptual recognition problems in such machines.

- Laumond (1983) describes a methodology for providing a mobile robot with 'learning' capabilities. Two aspects of 'concept learning' and 'procedure learning by generalization' are tackled. Most of the mathematics used are adopted from graph-theory. The 'learning' mobile robot is able to improve its navigational capabilities by modelling and analysis of its environment.

(ii) - OTHER EXAMPLES OF ROBOTICS RESEARCH RELEVANT TO 'LEARNING'

Many other research fields of the science of robotics have also been pursuing the objective of designing 'intelligent robots'. In the following some of the areas which have some relevance to the problem of modelling of learning will be outlined, and briefly described:-

(a) - COMPUTER SIMULATIONS OF ROBOTS AND THEIR ENVIRONMENTS: Some workers such as Nilsson and Raphael (1967); Uhr and Kochen (1969); Jacobs and Kiefer (1973); Uragami et al. (1976); Webster (1978); Heiserman (1981); and Rosenberg and Rowat (1981) have been involved in designing simple robot-environment (or organism-world) abstractions. Several mathematical techniques are used for depicting physical laws, and representing interactions. These models are either used for experimentation or validation of various hypothesis; or, simply, they are exercises in 'simulation' of a physical robot. Normally, the robot and its environment are represented on the computer screen by very simplistic images, and various 'obstacles' or 'boxes' placed within the environment; also, 'tasks' are defined in terms of movements or rearrangements of these boxes. Examples of issues tackled, using some A.I. techniques (e.g. LISP, Fuzzy-Logic), are 'navigation', 'obstacle avoidance', 'random-walk', 'spatial perception', 'adaptation', 'habit formation', 'pattern recognition', 'purposive action', or simple 'learning'.

(b) - MAZE SOLVING ROBOTS: The early 'maze-running' cybernetic machines (e.g., Shannon's mechanical mouse) were intriguing artifacts. Even, today this class of problem-solving robots attract much attention, although, not always from established scientific circles - yearly contests are held amongst the designers of micro-controlled maze-solving machines. Examples, of the more serious analysis of such problems are Stanfield's (1979) simulations of maze-solving on computers, or Allen and Allen's (1979) discussions of maze traversing algorithms. Maze-solving is also one of the main problems tackled by LOGO programs and their 'turtle' robots.

(c) - "PLANNER" ROBOT LANGUAGE and PLANNING AND GENERALIZATION IN ROBOTICS: "PLANNER" programming language was developed by Hewitt (1969) as a 'deductive logical system' which could best be used for manipulating, problem-solving or theorem-proving in robots. Various goals are established or dismissed, using a hierarchical control structure in conjunction with a set of assertions (statements). Many researchers have, hence, developed 'planning systems' or 'generalization systems' based on PLANNER, or other variations (e.g., STRIPS, PROLOG), for different classes of robots. For example, Nagata et al. (1973); Siklossy and Dreussi (1973); Kuzin et al. (1975); Hayes (1975); and Ferguson (1981) discuss such techniques in the context of specific task domains.

(d) - ROBOTS IN UNKNOWN OR PARTIALLY KNOWN ENVIRONMENTS: The control and manoeuvering of robots in unknown or partially known environments are problem areas of robotics which have attracted a lot of 'learning-related' issues. One solution is to use established mathematical procedures, such as 'dynamic-programming', for solving these control problems. Yet, some workers, for example Keckler and Larson (1970) or Friedman (1977), have proposed and discussed alternate 'heuristic' methods in formulating 'learning' or 'inference' systems.

5.2.13 TEACHING MACHINES

Although "teaching machines" are included within our A.I. (information processing) categorization of the 'learning' systems, nevertheless, they are a fairly independent area of research, normally associated with education. Principally, these teaching machines and systems are used in conjunction with some sort of "programmed learning", as tools for assisting learning, or evaluating progress. Yet, in spite of their early promise, these machines have not had a major impact on the educational system. The most elaborate, only, having few controllable parameters, and involving simple hardware/software.

An early example of teaching machines is described by Hoffman (1962), which was a simple device to be used as a supplementary tool in training behavioural scientists. Similarly, Pask (1970) discusses the mechanization of teaching from a broad cybernetic systems point of view, and in an elaborate survey of the field outlines some computer assisted instruction schemes.

The programming language LOGO was specifically designed, by S. Papert, for educational purposes; and has found popularity amongst the teaching establishments from a varied range of cultural backgrounds. LOGO is intended to be a compact experimental framework for teaching the concepts and skills of analytical and heuristic thinking. One aspect of the developments of this programming language has been the introduction of a complementary hardware realization, namely, the simple 'turtle' robot. These robots are only regarded as an extension of the notions used in LOGO; and they are considered as valuable tools for developing problem-solving or other perceptual skills of children. Feurzeig and Lukas (1974) discuss various aspects of use of LOGO and 'turtles' as programmable teaching machines.

5.2.14 EVOLUTIONARY-PROGRAMMING

We have previously discussed the kinship of the processes of evolutionary adaptation and learning. Various early cybernetic models (e.g. Pask,1962) had tried to depict the process of evolution through certain developmental stages. More recently, some A.I. researchers, in attempting to synthesis intelligence, have chosen the 'evolutionary programming' approach. This distinct avenue is away from the mainstream research of A.I., and aims to achieve artificial intelligence through the computer simulation of the process of evolution. Their protagonists cite various inadequacies of traditional A.I. methods in dealing with certain classification tasks, or in tackling a range of problems 'globally'. The basic technique is to abstract processes which can 'reproduce' with 'mutations' amongst the new 'population'; and, also, by using a process similar to 'natural selection', can evolve many generations of such 'organisms'.

Fogel et al. (1966) describe some attempts in the simulation of the process of evolution. Holland (1970) develops a mathematical basis for the analysis of such adaptive systems, which uses 'adaptive algorithm' in searching for 'good' solutions. Hand (1979) describes a simple computer simulation of organismic evolution. Schaffer & Grefenstette (1985); Davis (1985); Frey (1986); and Schrodt (1986) describe various research work based on Holland's 'adaptive algorithm'/'classifier', and 'genetic algorithms'; also discuss the concepts of 'learning', 'problem solving' and other related issues in such implementations.

5.2.15 CONNECTIONISM

In the past few years, once more the notions of neural networks have come to the forefront of the science of machine-intelligence. This type of research initiated from the 'neural-net' models of McCulloch and Pitts; and later 'logical-nets', 'Self-Organizing Systems', and 'pattern recognition' models tackled the notion of constructing 'intelligent' automata from conglomeration of identical interconnected elements. Meanwhile, other inter-disciplinary researchers had been occupied with developing various computer models based on 'associative memory' (Kohonen, 1977; Hinton, 1981; Palm, 1982; Barto, 1981; Drozen, 1970). However, the popularity of all these subjects had been waning for almost two decades; and it was only the advent of some relatively recent developments in 'parallel computers' (and other technological progress) which signified a revival of this approach.

The new, 'connectionist', discipline is fundamentally same as those discussed earlier. Yet, the influence of various A.I. related developments are evident in this area. The principal contention is still to depict the brain's mechanisms in some way, simulating the massive parallelism of neuronal networks. Learning has featured greatly in connectionism, and 'recognition' or other tasks have been tackled by various researchers. Examples of work in this area can be found in: Hinton, 1981; Palm, 1982; Kohonen, 1984; Feldman, 1985; Flynn, 1985; Shaw, 1985; Kibler, 1985.

5.2.16 AN OVERVIEW OF THE A.I. APPROACH TO THE MODELLING OF LEARNING

The various 'Learning' programs developed in this paradigm, and their associated robot realizations, display some interesting (and clever) behavioural learning patterns. Yet, there are some ideas expressed, or concepts and procedures defined, in their simple system/environment configurations that could have no real physical significance or sense for the model itself. Additionally, since there are no natural correlates of the learning behaviours observed, exact inferences cannot be made about any aspect of the natural learning process from this type of synthesis of learning. The only definitive conclusions made about the efficiency, accuracy, or complexity of the model will be confined to the particular physical or abstract implementation.

Additionally, most work in A.I. take certain primitive underlying logical organizations for granted. For example, the structured logical basis of digital computers or algorithms. Yet, when we look at the gradual developmental stages that the neural mechanisms of natural intelligence and learning have passes through, then, a phenomenological question becomes pertinent: could we truly expect to realize 'machine intelligence' if we bypass this apparently fundamental property of life, namely 'self-organization' from randomness, and attempt to 'synthesise' intelligence from well-structured beginnings.

When we look at the resulting entanglements of various robot or computer-based machines (e.g., vision-systems) in confronting the natural environment, or when the deep mathematical obscurities of some abstractions of a so called 'natural' process is observed, then a need for a rethink (refocus/reappraisal) of the problem becomes more evident. Today's robots and 'intelligent' programs seem to be on a path of ever increasing complexity; faster and technologically more advanced devices are utilised, and seemingly clever (yet ad-hoc) methods used in conjunction with large data bases to manifest proficient programs. A machine which has the capability to start learning from 'zero' or 'little' knowledge is the logical answer.

CHAPTER 6

DESIGNING A CYBERNETIC 'LEARNING' MODEL

6.0 INTRODUCTION

In this chapter once again we turn our attention to cybernetic models as tools for synthesis and simulation of learning. In particular, we focus on the problem which, as explained initially, was the root motive for our broad investigation of the field of learning; namely the design of a general purpose cybernetic model for manifestation of simple 'learning', which could also be used as an experimentation tool on various learning schemes.

We have, in fact, traversed a full circle. Starting from a practical specific problem, it was found that, to truly appreciate the underlying fundamental considerations and issues involved, a much broader perspective was necessary. Specially, in view of the rigid and unitary approach of almost all research into design of 'learning systems', it was deemed important to look at every possible aspect of learning. Hence, we embarked on an extensive scrutiny and discussion of the many facets of the ubiquitous phenomenon of learning. Now, we return, again, to our original problem equipped with a clearer understanding and a more qualified evaluation of its various aspects.

In the first part of this chapter we will attempt to briefly outline our hardware model. The principal intention will be to point out the general considerations involved in the design of the particular model, and also the whole class of this type of models. Specific technical features of the model will not be examined deeply, neither will the subsequent software developments be described beyond their generalised flow-charts.

In the second part of this chapter a more subjective analysis of the topic of our interest will be undertaken. Discussions will revolve around the general problem of devising simple 'learning systems' and various important related issues involved, keeping in mind many of the diverse subjects covered previously. Additionally, the blue-print and elements of some hypothetical 'learning schemes' will be proposed and scrutinised, particularly in the context of our hardware model.

The principal aim is to systematically approach the problem of design of simple cybernetic type 'learning' models, which have been appearing in various research fields as aids for demonstration or explanation within the past 3–4 decades.

6.1 A COMPUTER-CONTROLLED MOBILE ROBOT AS A TOOL FOR MODELLING SIMPLE LEARNING

The case for designing "hardware" models as an aid for experimentation, demonstration and better understanding of abstract theories or ideas was argued in the first chapter. Subsequently, various other aspects of such physical realizations were discussed, and many examples of hardware cybernetic or robotic 'learning' models were cited and examined in Chapters 4 and 5. In some instances it was also seen that a particular hardware configuration or device was, in fact, either the starting point for the design of a 'learning' system, or dominated over the software aspects; in a sense, the 'physical' features of these models overwhelmed all its later developments.

The degree to which the two aspects of 'simulation' and 'synthesis' have featured in the hardware models discussed so far have been quite varied. But, generally, the historically precedent physical models have been more involved in simulation of a particular learning trait; and the more recent models, with their enhanced processing capabilities, have inclined towards designs which try to synthesise a 'learning' behaviour using some non-physiologically based criterion.

Although, it must be emphasised that the distinction of these two features in models is not very clear cut, and an intricate mixing of simulation and synthesis is often seen in 'learning' models. Whether, a hardware model is trying to solve a 'maze finding' problem in the human fashion (simulation), or it is involved in an 'artificial' portrayal of problem solving behaviour (synthesis) are kinds of questions which can only be answered by considering the details of the design, the level of description, and the aims or intentionalities of the model's designer. Hence, such characterizations are often subject to interpretations, and are to some extent arbitrary.

Moreover, to make good use of 'simulation', it is necessary to have a certain degree of knowledge of the organizational aspects of the structure of a system and details of its mechanisms. Therefore, it is not surprising that in view of the complexity of underlying neural functions involved in learning, and a lack of precise neurophysiological postulates about learning, all attempts at its simulation should involve tentative synthetic criteria.

6.1.1 CONSIDERATIONS IN DESIGN OF A GENERAL PURPOSE MODELLING HARDWARE TOOL

The first consideration of the design was how to manifest the model/environment configuration, as represented schematically by FIG.6.1 in its most general form. Once the boundaries of a model are fixed within an environment, then the outputs can be defined in terms of changes that are brought about to the environment, or movements which are made by the model; and, conversely, inputs can be defined in accordance to the percepts that the model makes of its environment, or the physical influences that are impinged upon it. Of course, the subjectivity of such definitions must be appreciated; since, at any instant of time, the machine may be affecting its environment in a variety of ways; and also is being influenced by numerous inputs.

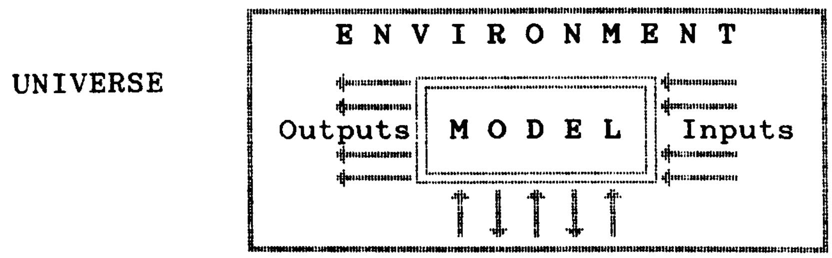


FIGURE 6.1. A general configuration of Model/Environment.

Now, the objective was to devise a hardware device capable of modelling the simple modalities of the learning process, as that which might be displayed by a hypothetical organism with basic sensors and effectors interacting with its environment. Additionally, the design of the machine was not to include any task-oriented consideration; in other words, it was not to be designed for displaying or achieving a preconceived pattern of behaviour - which might have limited the choice of approach to its later programming.

Various alternatives were investigated. The "environment" could be either fixed or changeable. A fixed environment can be used for simplicity and convenience to match some particularities of the machine, or limit the amount of inputs to the device. For example, a confined box with simple contours, or a meshed surface which could indicate the coordinates of the moving device, could be considered as a possible simple non-varying environment for the model.

However, in view of the general interactive way which the model was to realize the process of learning, it was decided that the richness and changeability inherent in a normal physical environment (the laboratory) would indeed be a desirable attribute. Any limitation on the amount of perceived information by the machine would be imposed by the choice of its inputs.

Other desired features of the "model" were considered. If a 'learning' was to be observed, and evaluated, then it was necessary that some controllable dynamism should be built into the model - a static model, or a model which modifies itself by virtue of simple thermodynamic exchanges could not convey any learning. Again, since the machine was to be used for the synthesis of basic learning criteria, then 'movements' rather than 'manipulative actions' were deemed to be more appropriate as outputs for the model. The system was to have the potential to even determine its own primitive goals, and any a-priori inclusion of specific patterns of action (innate behaviours), or use of manipulators which would require previously defined routines for operation, would hinder a true manifestation of rudimentary goal-determination.

Hence, it was decided that a free roving mobile robot be constructed as the 'body' of the model. A specific environment for its operation, as previously mentioned, could be devised. For instance, grid type surfaces similar to the examples illustrated in FIG.6.2 could convey the x-y position of the robot.

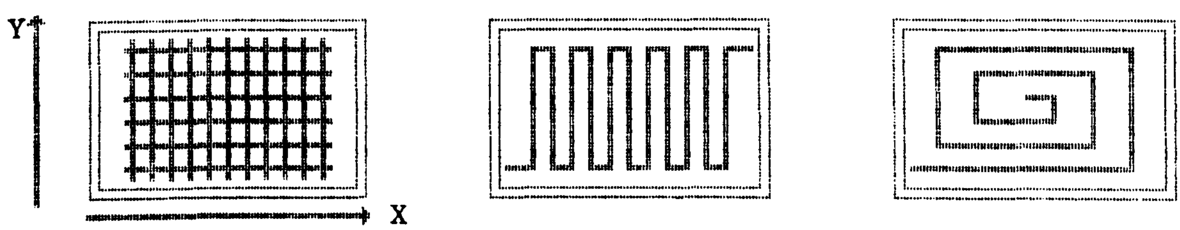


FIGURE 6.2. Three examples of intersecting the floor surface for monitoring the position of the robot.

Yet, these methods of determining the position of the robot in relation to its surroundings had the disadvantage of needing a lot of mathematical preprocessing, and subsequently lost all (or most) information contained in the connectivities of their physical world.

Other methods of 'perceiving' the environment were investigated, such as using TV-cameras, infra-red detectors, C.C.D.'s (Charge Coupled Devices), Photo-diodes, Sonar, etc. Eventually, it was decided that it would not be necessary for the model to realize its spatial coordinate within an environment, since this itself is a higher order learning stage.

Hence, a simple ultrasonic method for obstacle sensing was adopted, for reasons of simplicity and economy. The sonar devise could either be located in the environment, tracking the position of the robot from the emissions of a beacon signal. Or, it could be incorporated within the machine, showing its distance from various obstacles in its periphery in one or several directions (as depicted in FIG.6.3).

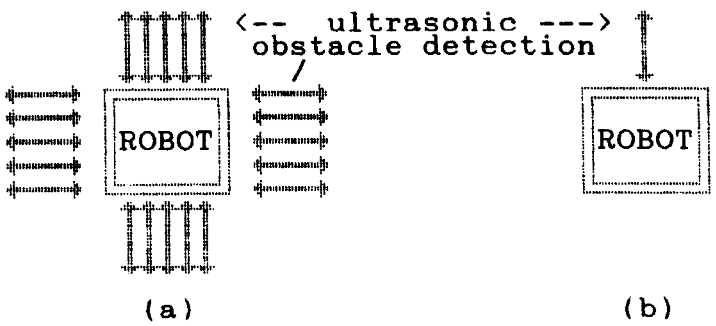


FIGURE 6.3. Two possible methods for ultrasonic detection of the environment: (a) - Multidirectional; (b) - Unidirectional.

It was contended that for the level of analysis of the process of learning we were interested in, and for the realization of primitive criteria of learning, a single narrow beam sonar located at the front of the robot would be more appropriate (and sufficient). Any fundamental issue explainable by the omni-directional sensing of the environment could also be explained, although in simpler terms, by the unidirectional sensing. After all, an identical compound map of the environment could be obtained by the simpler sensing device if only the robot made a 360° rotational scan of its environment around its central axis (which was possible); only the process would be a lot slower.

Of cou·se, it must be appreciated that this type of one-dimensional perception of the environment will give a far less precise indication of the relative position of the robot than a two or three dimensional sensing process. The environment will be seen in a much more immediate and changeable fashion, and, hence, decisions for actions have to be taken on such bases. Examples of this type of environment sensing is abundant in nature, and many simple animals manage to deal with various significant changes in their surroundings adequately using a single monocular input (chemical, electrical, tactile, olfactory, sonar, visual). The differences from the higher order sensing methods are only highlighted when a comparison is made by an external observer.

It can be said that a kind of hierarchy of 'environment perception' exists as far as the actual physical inputs to a robot within an environment (as below) is concerned - this is indicated in the ordering of FIG.6.4.

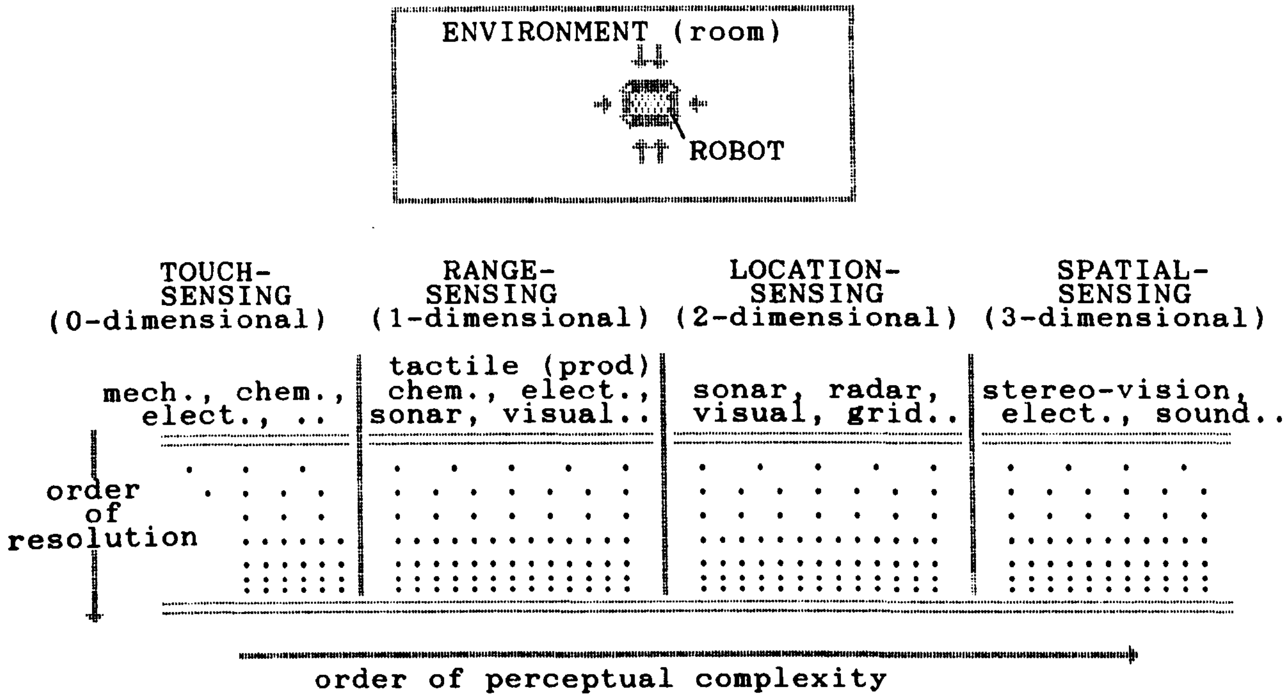


FIGURE 6.4. A representation of the hierarchy of possible methods for the physical sensing of the environment and its changes.

An important premise of this unidirectional design feature was the need to define a 'front' for the robot. This non-homogeneity implies that higher significances should be attributed to the inputs/outputs at a specific orientation of the robot's body. In other words, if the machine only 'sees' its forward direction then environmental changes occurring at its front should be much more 'significant' than one at its rear; similarly, 'forward' movements should be more 'consequential' than movements in other directions.

This issue may be considered as trivial, nevertheless, in our desire to incorporate minimal machine-specific features in the design it has to be further pondered upon. A solution would have been to build the robot completely symmetrically, using a single wide-angle sonar detector (or several detectors around the periphery) which could give a nondirectional sense of the nearest object to the robot; and also incorporating the capability to move in all directions from any point without the rotation of the 'body'. Yet, the ambiguity involved in such a diffused evenness of percept (and action) impelled designing the model in a 'forward-biased' fashion. The observation of an almost universal asymmetry in the organization of sense organs in all species, even in the very basic unicellular organisms, was a strong corroborating evidence towards the inclusion of such a non-uniformity in the design.

The sonar range sensor designed could indicate linearly the distance of the nearest object to the front of the robot at approximately its own floor height up to 5-meters quite accurately (±1-cm error) with a 7-digit binary code.

Additionally, eight touch sensors uniformly located around the perimeter of the chassis, and also a single centrally positioned shock-sensor, represented further 9-digits of binary input to the system.

Outputs were, in turn, defined by the direction and speed instructions to the single steering/drive wheel of the robot. Eight possible speeds (3-binary digits), sixteen possible angles for steering (4-binary digits), and one spare binary digit comprised the 8-bit output of the model. The robot could now be regarded as a system with following input/output configuration:

$$\overbrace{\boxed{T_1}\boxed{T_2}\boxed{T_3}\boxed{T_4}\boxed{T_5}\boxed{T_6}\boxed{T_7}\boxed{T_8}}^{\text{8-bit touch sensors}}$$

$$\overbrace{\boxed{R_1}\boxed{R_2}\boxed{R_3}\boxed{R_4}\boxed{R_5}\boxed{R_6}\boxed{R_7}}^{\text{7-bit range}}\overset{\text{shock-switch}}{\boxed{ss}}$$

INPUTS

$$\overbrace{\boxed{D_1}\boxed{D_2}\boxed{D_3}\boxed{D_4}}^{\text{4-bit direction}}\overbrace{\boxed{S_1}\boxed{S_2}\boxed{S_3}}^{\text{3-bit speed}}\overset{\text{spare}}{\boxed{xx}}$$

OUTPUTS

The information perceived by the mobile robot should now be sent to and instructions received from a computer. The choice was between a cable link or some form of wireless transmission of data (radio waves, ultrasonic, infrared, etc.). The cable connection could involve using special multi-way swivel connectors; or multiplexing the signal, and hence using a pair of wires for transmission. Yet, since an on-board power source was used the entanglements of the cable could pose problems in a long term free operation of the model. Therefore, a two-way radio link with appropriate error checking circuitry was designed for computer control and communication. Nevertheless, a multiplexed cable link option and use of remote a power source was also incorporated.

A question which is frequently asked by outside observers of such developments is that why not simulate the same robot on the computer screen rather than constructing it in hardware, specially in view of many advanced graphic capabilities of modern computers. Although, such simulations have indeed been undertaken by various workers (Webster, 1978), there are numerous shortcomings - of course, unless the objective is to devise a simple idealised robot/environment visual simulation.

Firstly, no present day computer, or for that matter in a foreseeable future, can faithfully simulate the rich complexity and unpredictability of the "natural" environment. Whereby, the senses are in a constant bombardment by a variety of parallel stimulations and natural forces; the depth, colour, contrasts, texture and all other physical influences of the real world have to be calculated. Even then many novel or unexpected events happen in the physical surroundings which cannot be predetermined in the simulation.

Secondly, the interaction of the robot itself with the environment seldom follows an exact predetermined path, and hence not only it would be difficult to simulate all dynamic trajectories or movements of the robot but almost impossible to predict all unexpected eventualities of its interactions.

6.1.2 SPECIFIC DESIGN FEATURES OF THE MOBILE ROBOT

In the following we will briefly outline the specific mechanical, electrical, electronic, and computer-interface features of our hardware model. Pointing out any particulars of design which might be of interest to the wider aspects of our discussions.

(i) - MECHANICAL FEATURES OF THE MOBILE ROBOT

The steering and drive mechanisms of the mobile robot were assembled on a circular chassis, as illustrated in FIG.6.5. The front wheel was used for both steering and drive, and appropriate step-down gearing mechanisms incorporated. The two rear wheels were free running castors allowing a high degree of manoeuvreability.

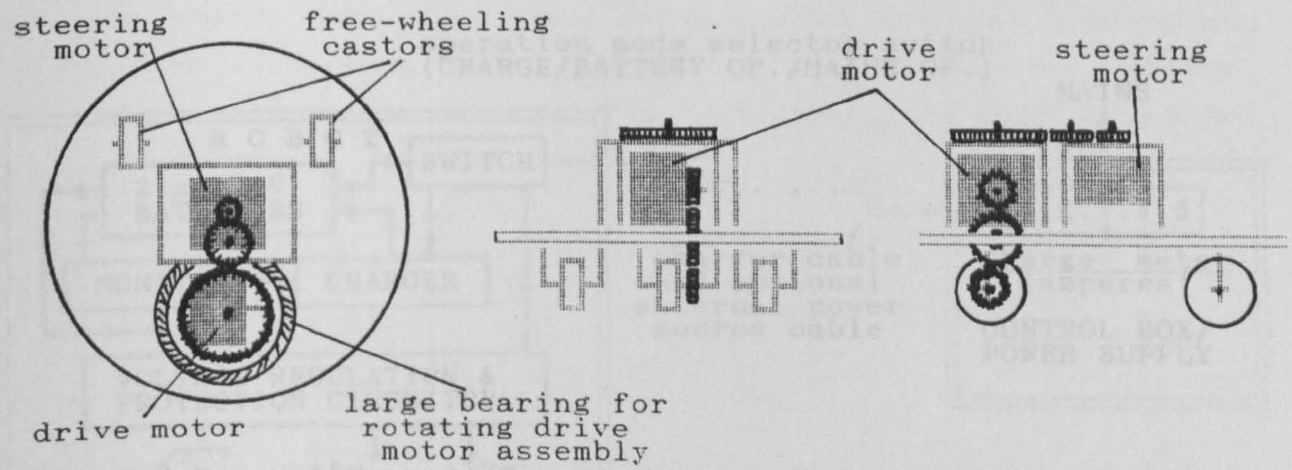


FIGURE 6.5. Three perspectives of drive and steering mechanisms of the robot.

Additionally, eight mechanical touch switches were designed to detect any physical contact made by the robot with its surrounding objects. These were spring loaded metal contacts spaced equally around the periphery of the main chassis. Also, a pendulum type shock switch (sensitivity adjustable) was constructed and placed in the centre of the robot's body, whose activation would signify a vigorous collision or tilting. These inputs are illustrated in FIG.6.6.

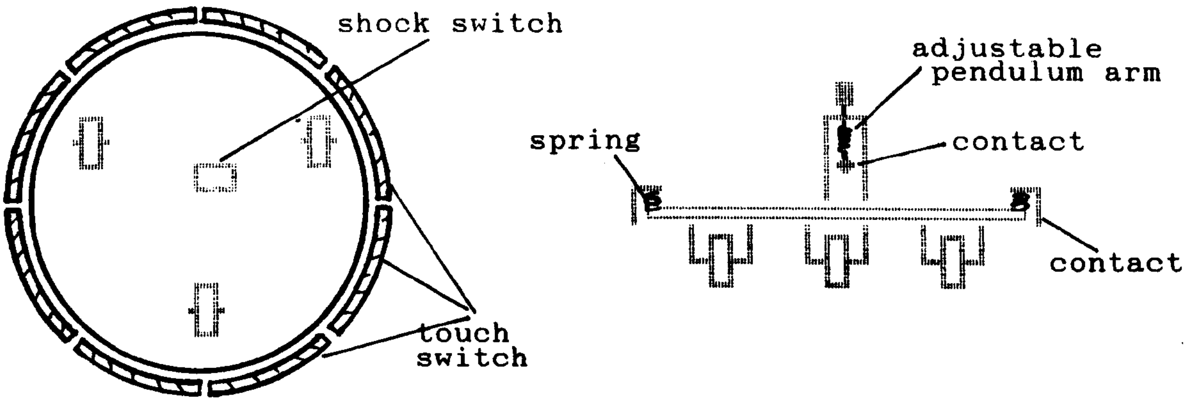


FIGURE 6.6. The position of eight touch switches and the shock switch.

(ii) - ELECTRICAL FEATURES OF THE MOBILE ROBOT

The normal power source for the robot was two 12-volts rechargeable lead-acid batteries connected in parallel, supplying power for approximately four hours of continuous operation when fully charged. Appropriate charging, regulating, protection and monitoring circuitry were devised and built. In addition, an on-board power regulator could supply the necessary +5 and +12 smooth voltage levels. This regulator circuitry was also able to use an external power source (rather than the batteries) as input - connected to the robot via an umbilical cable. The general electrical features of the robot are outlined in FIG.6.7.

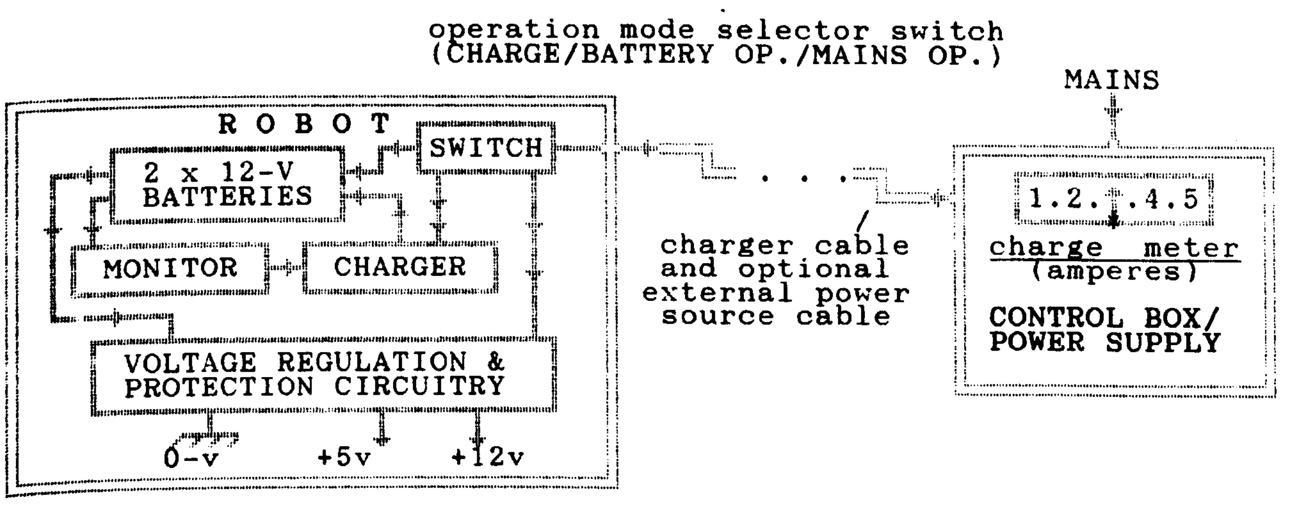


FIGURE 6.7. Schematic diagram of main electrical design features.

(iii) - ELECTRONIC FEATURES OF THE MOBILE ROBOT

There were three main areas of design involving electronic control of the hardware of the mobile robot, and electronic communication of information: (a) - Motor Control; (b) - Ultrasonic Range Detection; and (c) - Data Organization, Checking, and Transmission. The principal components of the mobile robot and its control system are illustrated in FIG.6.8.

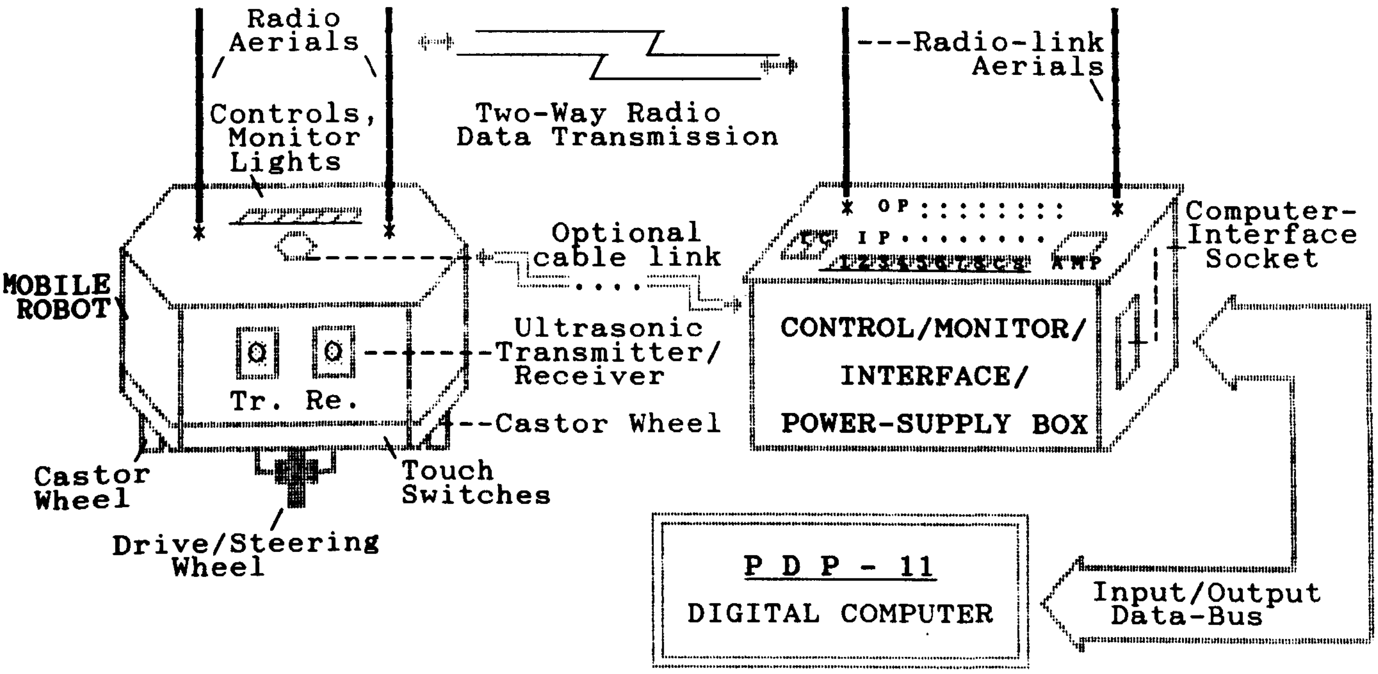
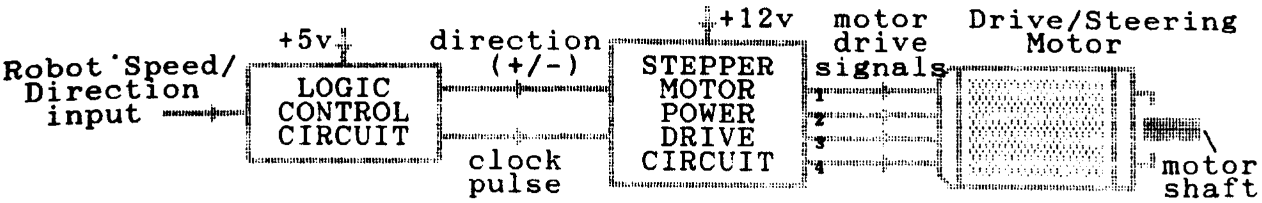


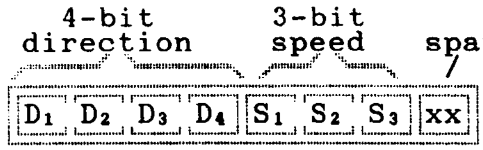
FIGURE 6.8. The basic components of the mobile robot system.

(a) - Motor Control

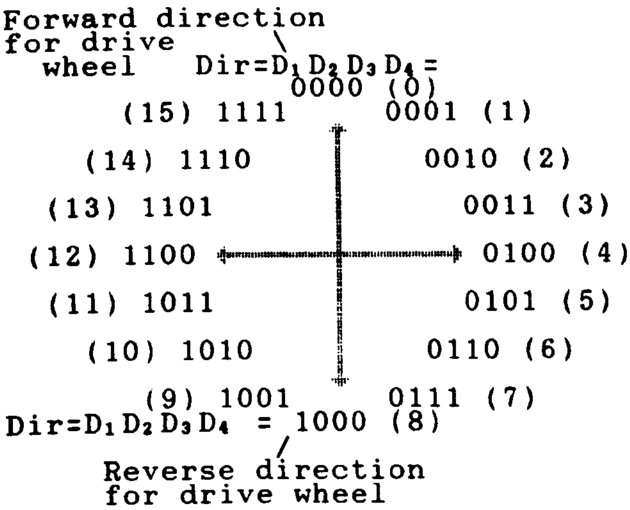
Two identical precision stepper motors were used for both steering and driving the robot. The use of stepper motors would obviate the need for devising feedback mechanisms to monitor speed or direction. Each stepper motor would be driven by a power circuitry which received a 'clocking' and also a 'direction' signal from its appropriate logic circuit. For each clock pulse the shaft of the motor rotates by 7.5 degrees in the direction specified by the polarity of 'direction' input.



Once the 3-bits of speed and 4-bits of direction instructions are received and verified by the robot, then the logic control circuits translate these data to appropriate number of clock pulses and direction of shaft rotation for the stepper motors. The rules for translations of the 7-bit output command, indicated below, are outlined in FIG.6.9.



(a)



(b)

No	S_1	S_2	S_3	S P E E D
0	0	0	0	STOP
1	0	0	1	clock ÷ 8
2	0	1	0	clock ÷ 7
3	0	1	1	clock ÷ 6
4	1	0	0	clock ÷ 5
5	1	0	1	clock ÷ 4
6	1	1	0	clock ÷ 3
7	1	1	1	SET 0-DIR.

(c)

FIGURE 6.9. (a) - The format of an output instruction to the mobile robot.
(b) - The translation codes for various direction commands.
(c) - The translation code for seven different speed commands, and one zero direction (forward) calibration command.

The direction instructions received (from the computer or control box) could be interpreted by the robot in two different modes, which were selectable by a switch. Either, the steering wheel would rotate in relative terms to its present setting. For example, rotating 90° to the right when a steering signal 0100 (4) was received, and taking no action when 0000 signal received. Alternately, the change of direction could be based on absolute terms in relation to the robot's body. Here, it was necessary to calibrate the steering mechanism for forward (zero) direction. Hence, using a small magnet and a reed-relay a feedback system was devised which upon the receipt of speed-signal 7 ($S_1S_2S_3$ = 111) would halt the robot and start a routine of zero-direction-calibration. This procedure could be repeated during the robot operation every 10-15 minutes to ensure that the selected directions remain accurate, and errors would not be accumulated. Another hardware design feature was the ability to change the steering wheel to a new direction using the shortest arc of rotation.

The drive circuitry could enable the selection of six different speeds (including zero-speed) in a single direction, although, a variable speed facility was not deemed to be as crucial as a high directional variability feature. Additionally, the speed control clock pulse frequency was also adjustable.

(b) - Ultrasonic Range Detector

The ultrasonic range detector is outlined in FIG.6.10. It could measure the distance of nearest object from the front of the robot at approximately floor level by detecting the shift between a transmitted pulse of ultrasonic wave and its returned reflection. Various pulse widths, cycle widths, frequencies, and controls were hardware adjustable to suit specific operational conditions.

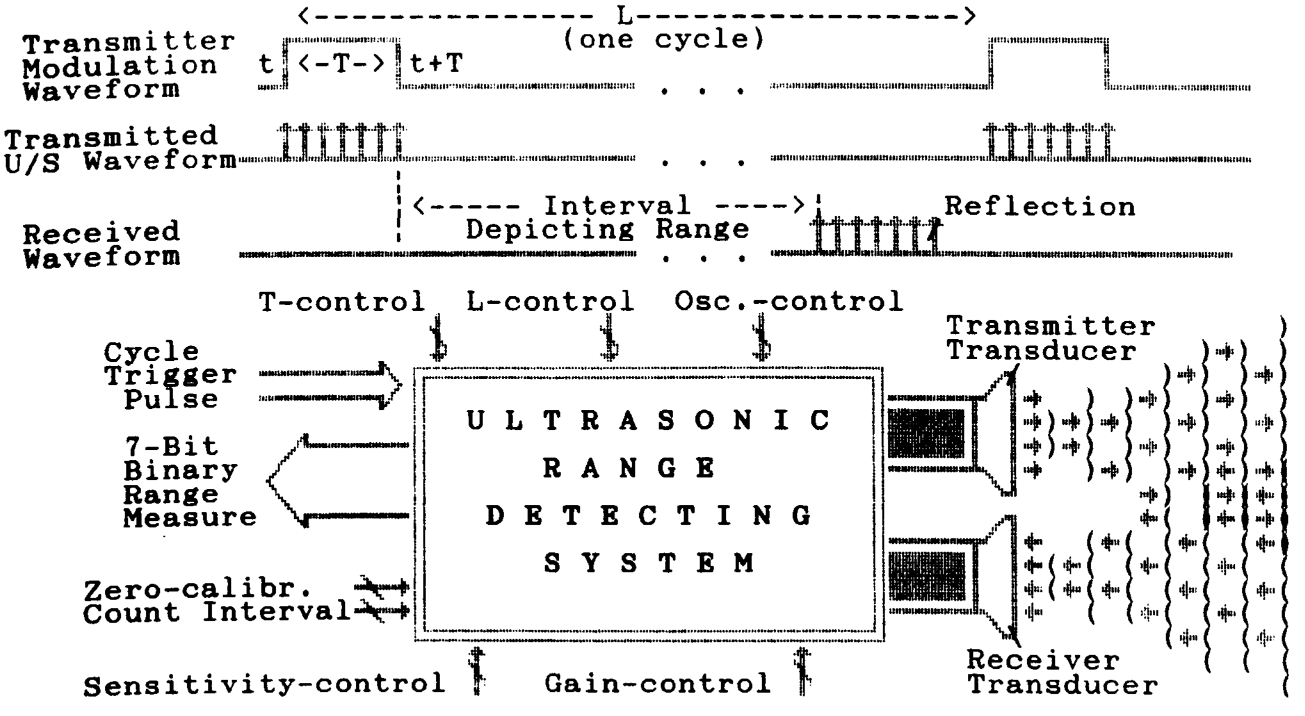


FIGURE 6.10. A diagram of the Range Detector system and its main controls, showing a representation of transmitted/received waveforms.

Each detection cycle, typically of order of 150 milli-seconds, could be either initiated independently by a clock, or governed by the main communications cycle of robot's data transmission system (details in following section).

(c) - Data Organization, Checking and Transmission

Information collected by the mobile robot about its environment (also the movement instructions received) had to be serially transmitted (and received) before interfacing to the computer. For this purpose a control box was constructed which as well as relaying and processing the data housed the power supply, the manual control panel, and also the monitoring displays. The 8-bit output and the 16-bit input data were transferred between the control box and the mobile robot using a pair of identical 27-MHz radio transceivers (same data could also be transferred by cable).

After an output instruction was received by the robot, first the number of binary bits and their parity would be checked, and if verified the data accepted as correct. Then, the robot input data would be transmitted to the control box, going through similar verification or rejection processes. An example of a typical transmission cycle approximately lasting 350-ms is depicted in FIG.6.11.

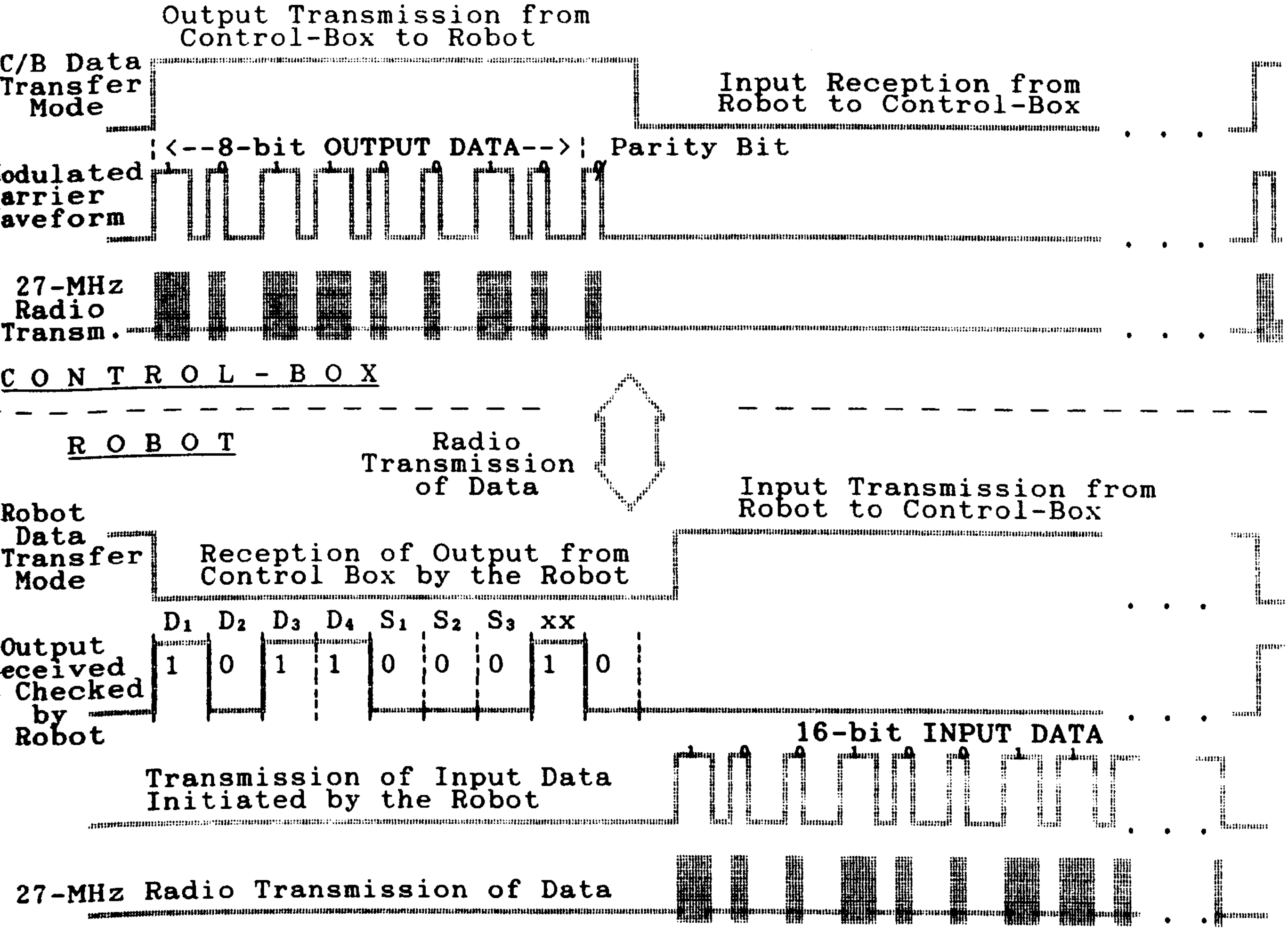


FIGURE 6.11. An illustration of phases of a typical data transmission cycle. Output <u>10110010</u> is transmitted from the control box, and input <u>10010011...</u> received, using Pulse-Width Modulation (PWM).

(iv) - THE COMPUTER INTERFACE

A high density 64-bit interface expansion card was used to connect the PDP-11 digital computer to the control box. This interface allowed 64 lines of input (or output data) be read (or written) in parallel into the computer. It also had the interrupt facility, whereby information could be read only when required, hence saving on computer processing times.

The 16-bit robot-input and 8-bit robot-output lines of the control box were connected to the above interface ports. Additionally, a hardware 16-bit binary clock was designed and connected to one of the ports. This clock could be reset and initialized at the start of an experiment, hence keeping track of the lapsed time, up to approximately six hours - a software designed clock could, more or less, perform the same task, but less economically.

6.1.3 SOFTWARE DEVELOPMENTS OF THE MOBILE ROBOT

The initial software programs developed after the completion and testing of the hardware design were mainly for the purposes of demonstration and investigation of the potentialities of the model.

At the lowest level, various utility routines were devised for: processing the output/input signals and checking their validity; setting operating conditions and initializing the system; or visually monitoring the states of the mobile robot on the computer screen.

At the next higher level, which can be thought of as the reflexive-action domain of the robot, certain deterministic procedures were formulated to enable the machine display specific behavioural patterns. These procedures could either be used as components of a more elaborate 'learning' program, or, alternately, used as an integral part of a learning scheme itself (e.g., in simulating reflex conditioning).

The inclusion of such "innate" behavioural repertoire not only posed challenging questions in trying to enable our perceptually simple model to display some fairly complex actions; but, also, provided interesting insights into the specific input/output characteristics of the model, and the limitations which the scope of percepts or actions (universe of discourse) can impose on a design. In addition, these reflexive processes could ensure the smooth operation of the robot over long durations, in particular, if trapped or entangled in its surrounding irregularities.

Although, as mentioned previously, we are principally interested in fundamental 'learning' issues in machines which incorporate minimal directiveness of behaviour at the outset. Nevertheless, the analysis of these algorithmic emissions of the machine should greatly assist us in devising 'learning' schemes, later. Whereby, behavioural patterns similar in nature to these algorithmic behaviours could be "acquired" - using the model's own experiences and inference systems, rather than the intelligence of its designer. Hence, these rule based sub-programs, which will be briefly outlined in the following, should provide good yardsticks for the level of complexity of 'learned' behaviours we may expect from our model.

(a) - Sub-Program to "TRAVERSE PARTICULAR TRAJECTORIES"

This was a simple program written to demonstrate intricate and interesting behavioural emissions from the robot, as its speed and direction output settings were changed according to certain rules. In one example, the value of various parameters of two time dependant polynomials would determine the type of function/trajectory the output variations would follow. The general form of these polynomials were:

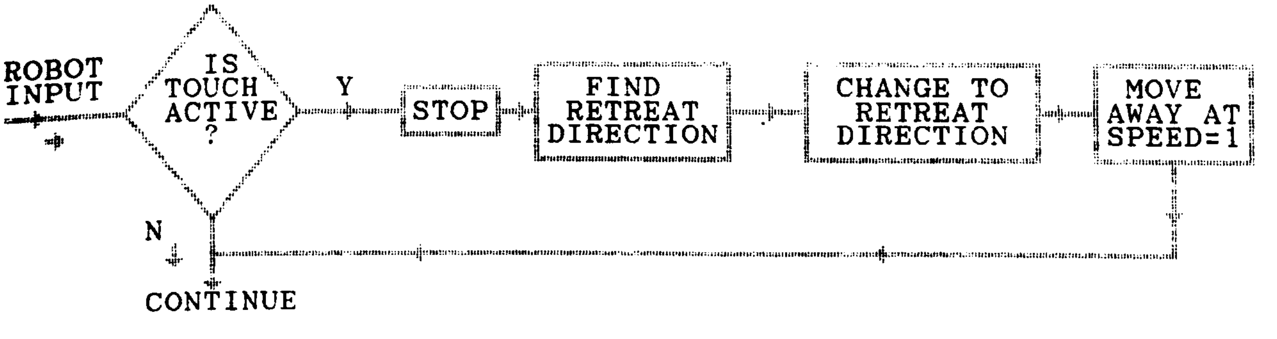
$\text{SPEED}(t) = A + Bxt + Cxt^2 + Dxt^3 \ldots$ (A, B, C, ... are constants)

$\text{DIRECTION}(t) = K + Lxt + Mxt^2 + Nxt^3 \ldots$ (K, L, M, ... are constants)

Because of the limited speed/direction values possible (speed = {0,...,6}, direction = {0,...,15}), a modular measure of the SPEED/DIRECTION (in their appropriate ranges) values were used for steering and driving the robot (i.e., SPEED(16) =-SPEED(0), SPEED(17) =-SPEED(1), ...).

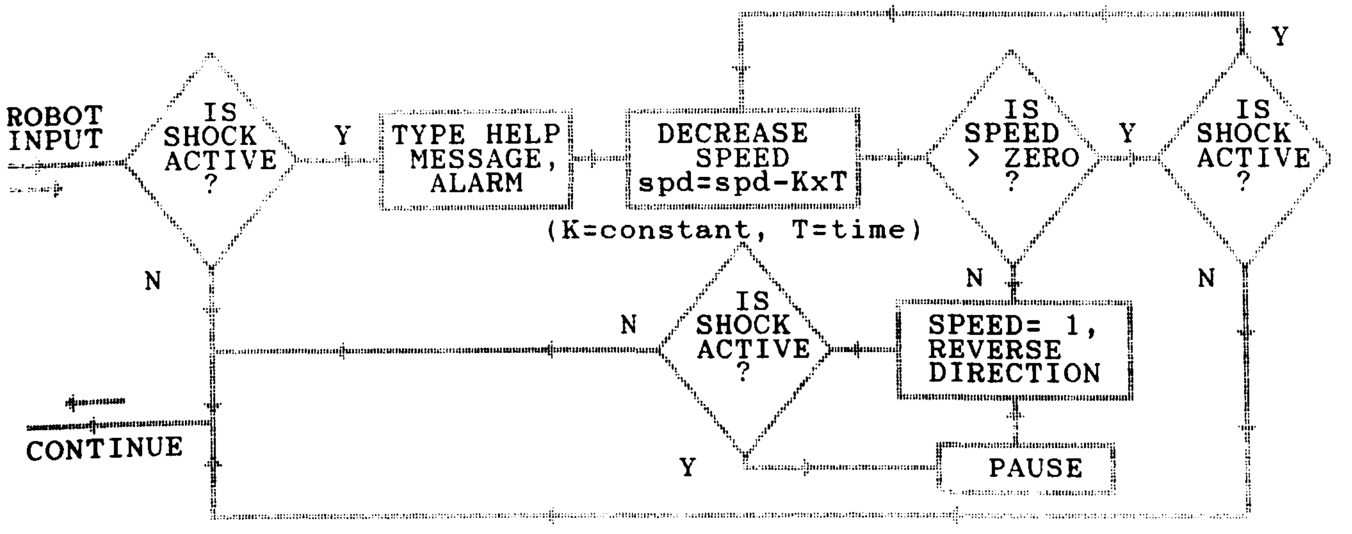
(b) - Sub-Program to "RETREAT AFTER COLLISION"

If one or more of the touch switches around the periphery of the mobile robot were activated, then the 'retreat' procedures of this program would steer the machine in a direction away from the activated contact switches. When only one switch is closed, a simple mathematical relation could give the retreat direction. But, for two or more switch contacts the extremes of the contacted points on the circumference had to be identified, then the retreat angle calculated accordingly. A general flow-chart of the processes involved is outlined in the following:



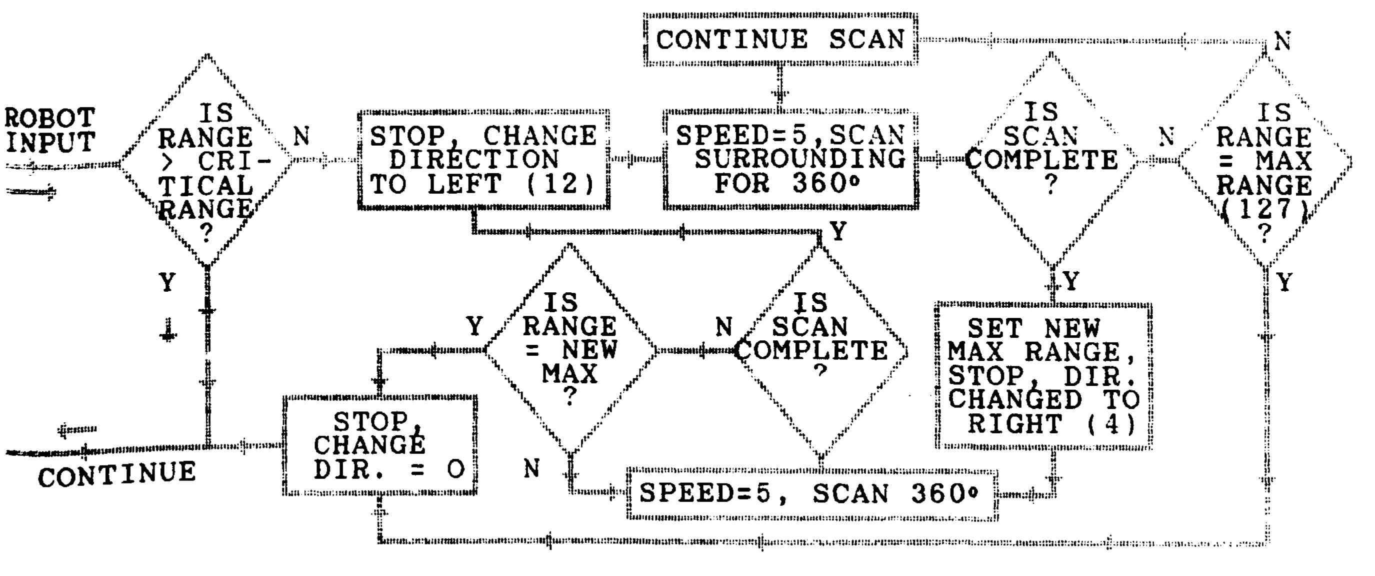
(c) - Sub-Program to "MANOEUVER AFTER SHOCK-SWITCH ACTIVATION"

A similar sub-routine to the 'touch reaction' was devised to deal with the activation of the 'shock' switch, which indicated that either the robot was on an uneven/slanted surface or it was being agitated vigorously. In either case a set of evasive actions were formulated, illustrated by the flow-chart below, to deal with such eventualities.



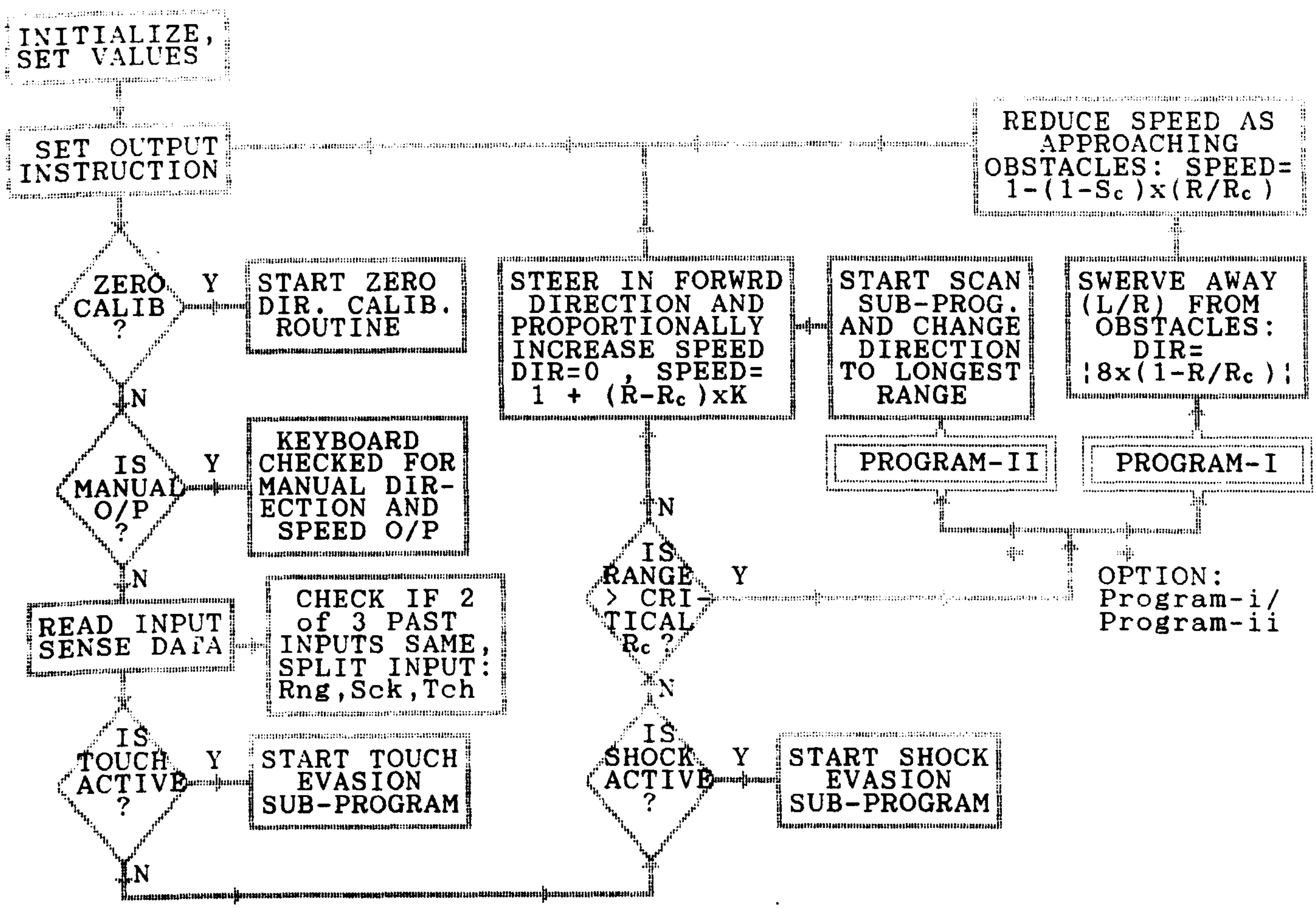
(d) - Sub-Program to "SCAN SURROUNDINGS AND FIND LONGEST RANGE"

This program allowed the robot upon nearing an obstacle (distance less than a critical preset range, R_c) go into a routine of 360° scanning of its surroundings, and once the direction of maximum distance was established would steer in that particular direction. In operation, this sub-program alone was capable of navigating the mobile robot around the room without collisions indefinitely. A radar-type display of the scanning process was also incorporated within the program, showing a simplified map of the room on the computer terminal. The scheme is illustrated in the following flow-chart.



(e) - Integrated Programs for Robot Operation

The above four sub-programs were combined with other subroutines and operational procedures in various integrated complex programs. Two examples are outlined in the flow-chart illustrations below. The program-(i) would steer the machine around the room in straight direction as long as the sonar distance (R) was larger than a critical value (R_c). Once within critical range, the direction and speed (S_c is the speed at R_c) would gradually start to change according to the criteria specified in the flow-chart boxes, steering the robot away from obstacles and avoiding collisions. Similarly, the program-(ii) would steer the robot in the same manner outside the critical zone, however, once inside the critical range it would go into a scanning mode and find the longest sonar direction. The normal operation could be interrupted by shock or touch activation, the zero-direction calibration, or keyboard override.



6.2 DESIGNING A 'LEARNING' PROGRAM FOR THE MOBILE ROBOT

Having investigated in the previous section some possible schemes which enabled our model to display interesting behavioural characteristics, now, we can attempt to devise 'learning' programs which will try to aim for a similar level of operational complexity.

The essence of our algorithmic programs could be summed up by few lines of mathematical expressions which determined speed/direction (outputs) as functions of range/touch/shock (inputs) and time. Hence, a learning program aiming towards same ends should somehow manage to extract similar relationships on its own accord after a period of experimentation. For example, elementary objectives of the 'learning' model could be to "AVOID COLLISIONS" or to "ROVE IN A ROOM KEEPING CERTAIN DISTANCE FROM OBSTACLES".

In the following first we will try to highlight and discuss some of the principal issues involved in designing such 'learning' programs, then we will attempt to outline the details of one scheme for our specific hardware model.

6.2.1 "TIME" AS AN INPUT TO THE MODEL

An important question should be addressed here, and that is the function of "TIME" variable in the schemes we have devised so far, and by inference, its significance to our later discussions - although, a deep analysis of the role of "time" in such modelling problems is outside the scope of our enquiry.

In our specific model a measure of time is not expressly perceived by the robot, but time is regarded as an input which affects the environment and the model in a parallel fashion. The independent clocking mechanism manages the coordination of the whole system, facilitating ease of analysis and control of various aspects of the operation.

It is contended that the inclusion of 'time' as a primary input is not an essentiality for such basic 'learning' systems. Firstly, 'learning' can be based on a simple spatial (rather than temporal) contiguity of events.

Secondly, it can be envisaged that some sort of temporal consequentiality of events may be used for 'learning' without explicitly referring to the element of external time as a clocking mechanism. Here, 'changes' or 'differences' of percepts may be used to represent an ordering of significant stimuli during the process of learning. However, in this case, it is assumed that stimuli perceived follow each other within reasonable periods of time. In other words, the detected environmental changes should be meaningful for the system, and time scales of external changes should relate to internal time scale of changes.

Thirdly, a different physical parameter of the model itself could represent an independent sequential scaling of events, either in step with external time (although out of phase), or with its own gradation.

6.2.2 TELEOLOGICAL CONSIDERATIONS

Another important aspect of our design is the teleological (directiveness) considerations. In Chapter-2 the psychological and physiological contexts of related concepts were examined, and also in Chapter-3 we discussed such topics in their general systems context. But, here, we will focus on the teleological issues involved in designing the more specific class of cybernetic simple 'learning' models.

(i) - GOALS and GOAL-DIRECTED BEHAVIOUR

Many of human and animal activities, and almost all their learned or adaptive behaviours, are unquestionably goal-directed in nature; and descriptions of learning processes in their biological domains invariably contain some directive concepts such as goals, intentions, motives or purpose. Yet, the successful translation of these concepts to the artificial domain of machines or systems, where no real sense of 'purposiveness' or 'intentionality' can be expressed, requires some careful consideration.

In biological systems it has been shown that specific neural mechanisms govern the motivational and drive aspects of basic patterns of behaviour, and also indications point to the involvement of physiological factors (hormones, chemicals, etc.) in the determination of higher purposive behaviours. On the other hand, in artificially constructed systems all such teleological notions are defined on subjective basis by external designers, and are generally seen as properties of the 'behaviour' of the system (not its construct). However, in spite of this fundamental difference, many workers have attempted to devise global definitions for goal-directed behaviour, applicable to both natural and artificial systems.

In the following diagram (FIG.6.12) the various levels of teleological activity as a general system property, related examples, and also schematic representations of their processes indicating types of goals and behavioural pathways are outlined in a tentative classification. Although, it must be mentioned that there is no consensus about the categorization or definitions of teleological behaviours amongst scientific workers from different disciplines (e.g., cybernetics, philosophy, A.I., psychology, etc.). For example, the first category of our classification in FIG.6.12 (the equilibrium-seeking/reflexive behaviour) is sometimes not considered as a goal-directed type activity. However, they are included in our taxonomy in an attempt to construct a broad framework, and also to demonstrate a developmental or evolutionary continuum from the simplest to the most complex modes of behaviour.

TYPES OF TELEOLOGICAL BEHAVIOUR	SIMPLE SCHEMATIC REPRESENTATIONS → algorithmic pathways ⇒ heuristic pathways	EXAMPLES- NATURAL SYSTEMS	EXAMPLES- ARTIFICIAL SYSTEMS
equilibrium-seeking or reflexive behaviour	Initial State (unique-pathway) Inevitable Final Goal	dynamic vascular flow-rate variations or knee-jerk reflex	movement of a pendulum or action of a light switch
simple error-reducing behaviour involving feed-back and a range of action choice	(choice of pathways) Initial State → Specific Final Goal	placing or moving an object to a desired position	action of a thermostat
complex but reducible algorithmic-ally determined pattern of behaviour	I. S. → Well-Defined Sub-Goal .. → W. D. S. G. → Well-Def. Final Goal	going to a known destina-tion by following a specific route	assembly process of an automated plant
complex but reducible to algorithmic & heuristically determined patterns of behaviour	I. S. → W. D. S. G. .. → W. D. S. G. → W. D. F. G.	acquiring a physical skill such as cycling or writing	some chess or checkers-playing programs
complex goal-directed behaviour seeking to optimise or maintain some general desired goal states or goal variables	I. S. → Tent-ative but WD SG .. → T. W. D. SG → Ill-Def. Final Goal; Final Goal Modifier	trying to maintain social status or family harmony, seeking a state of well-being	variations of the economic system of a country
complex goal-directed behaviour in pursuit of essentially unattainable or unsustain-able goals in the form of ideals or paragons	I. S. → T. W. D. SG . → T. W. D. SG → I. D. F. G. Well Def. Target Goal; Final Goal Modifier	pursuit of happiness, trying to establish a new world record in sports	? possibly some models in A.I. or cognitive-psychology

FIGURE 6.12. A possible taxonomy for the various levels of teleological behaviours and systems, together with appropriate examples.

The examples of the first three levels of goal-directedness as defined in the hierarchy of FIG.6.12 are quite abundant in nature, their teleology being implicit in the simple consequence of their physical actions, but the higher types of goal-directedness are only seen in the more advanced animals capable of some mental processing or abstraction. In humans it has become clear that the lower levels of hierarchy of goals serve the higher levels in determining goal-directedness.

Aside from the first category of our teleological taxonomy the process of learning could be involved in the formation of all other types of goal-directed behaviour. In addition, the more complex modes of teleology, in particular those involving heuristic concepts, could entail some anticipatory aspects, such as 'expectations' and 'predictions'. These notions inevitably require a higher processing and conceptual level which can somehow internally model the consequence of actions, overseeing the development of goal-directed behaviour.

An example of different characterizations of teleological processes is proposed by Ackoff and Emery (1972) in defining 'Purposeful Systems'; which are defined as systems that are able to change or select their goals under constant environmental conditions, and are also able to adopt different functional means to achieve the same goal. Their 'ends', 'goals', 'objectives', and 'ideals' being equivalent to the 'specific-goals', 'well-defined-goals', 'ill-defined-goals', and 'target-goals' of our categorization of FIG.6.12.

(ii) - DEFINING GOALS

The task of defining and setting goals in cybernetic systems that are characterised by an input/output/internal-processing configuration, such as our model, could be carried out in four principal ways:

(a) - At the most basic level, the 'goals' of the machine are unknown at the outset, and the model itself gradually starts to develope its own goals following a process of interaction with its environment. Either, procedures designed into the system will enable the machine to extract goals from of its interactions. Or, the system can start with near random or very primitive behavioural characteristics and approach a more complex form of behaviour. However, in both cases some initial organization is deemed to be necessary, since it is inconceivable that without any a priori physical biases or imposed general operational constraints, such as aiming towards the attainment of instability/stability/change/equilibrium, the model would be able to randomly extract useful goals from its interactions with environment. The various entropic analysis of systems (e.g., the law of requisite variety) attest to this need for some initial organizational order.

(b) - More specific, yet all encompassing, goals could be incorporated within the model at the outset. These goals have to be maintained, achieved, aimed, or kept within certain limits during the operation of the system. 'Hedonic' principals (pain-pleasure, reward-punishment, etc.) can be considered as simple expressions of this type of goals. In our particular model an example of this method of goal setting is:

"WHEN TOUCH SWITCHES ARE ACTIVE" ⟶ PAIN, and
"WHEN TOUCH SWITCHES ARE NON-ACTIVE" ⟶ PLEASURE

"WHEN RANGE LARGER THAN CRITICAL RANGE" ⟶ FAVOURABLE

The above simple goal criteria alone have the potential to direct the development of the activities of the model, following some 'learning' phase, towards 'obstacle-avoidance' type behaviour.

(c) - A compound pronouncement of the desirable target events, as extrapolated by the designer, could be used as goals. Such goals could either be a sequence of the more fundamental type-(b) goals; or, a logical/mathematical combination of type-(b) goals - whose pursuit should "algorithmically" result in the desired end pattern of behaviour. For example, in the case of our mobile robot we may have goal events defined as:

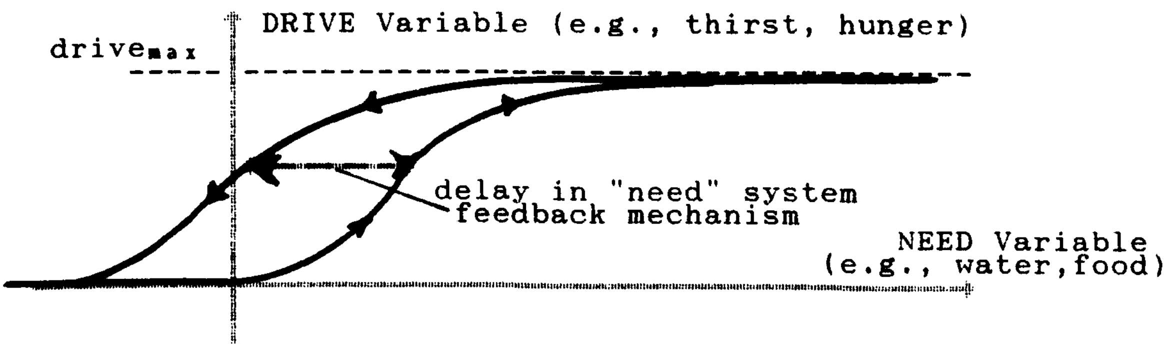
"WHEN ACTIVATION OF THE SHOCK SENSOR IS FOLLOWED BY TOUCH SENSORS"; or
"WHEN SONAR RANGE IS ZERO AND FRONT TWO TOUCH SENSORS ARE ACTIVE".

(d) - The designer of the system, using introspection, could determine that a certain sequential/mathematical/logical relationships between inputs and outputs would be conducive to bringing about the desired behavioural pattern, without actually knowing the exact steps for attaining the objective. In our particular model examples of this "heuristic" type goal events could be:

"WHEN THE PRODUCT OF RANGE AND SPEED IS LESS THAN A CONSTANT K"; or
"WHEN ALL TOUCH SWITCHES ARE INACTIVE FOR TIME T AND SPEED $\neq$ 0".

(iii) - DRIVE AND MOTIVATION

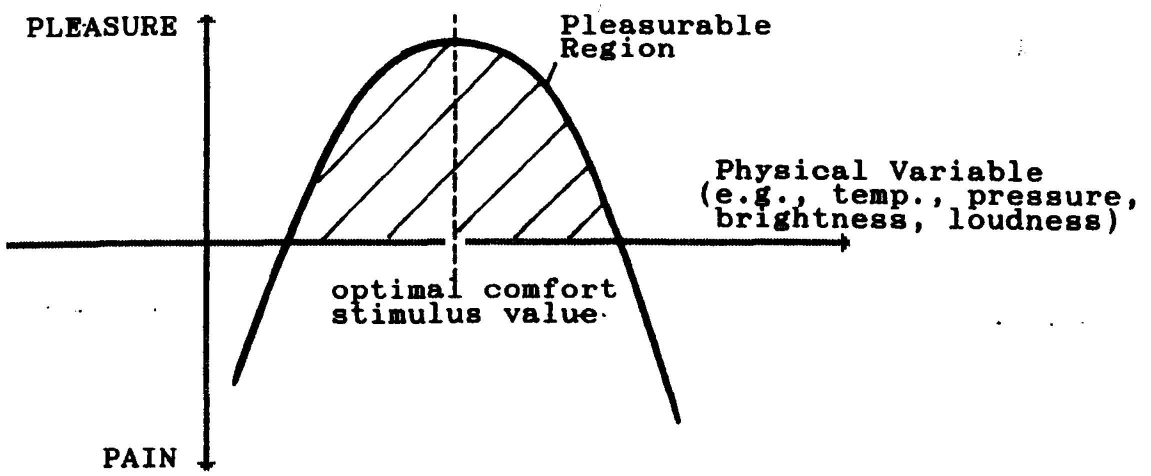
Most of animals' simple drive mechanisms have action which typically can be represented by the following generalised response curve:



Additionally, various factors such as fatigue, sensitization, boredom, attention, distraction, time, etc. could affect an animal's motivational aspects, and reduce or increase the drive accordingly (this is equivalent to changing the skewness of the curve in the previous diagram).

There is, also, a complex interaction between the various drive and need mechanisms of an animal and some higher hedonic centers, which indicate the severity of other more global variables. These hedonic mechanisms could be considered as 'reward-punishment' centers, or alternately, more fundamentally, as 'pain-pleasure' centers.

In the context of elementary 'survival', the 'pain' mechanisms are those which signal undesirable inputs or outputs, and stop the organism from exceeding various critical limits of its physical capabilities. Conversely, 'pleasure' mechanisms consolidate and enforce the most utilitarian, efficient, or desirable modes of action; or try to sustain pleasurable experiences. For simple physically quantifiable stimuli (also for some complex or abstract stimuli) the function of a pain-pleasure mechanism can be typically depicted by following illustration:



Here, 'motivation' can be loosely defined as an aversion from painful or attraction towards pleasurable stimuli; and 'drive' can be considered to be proportional to the degree of such aversive or attracting forces.

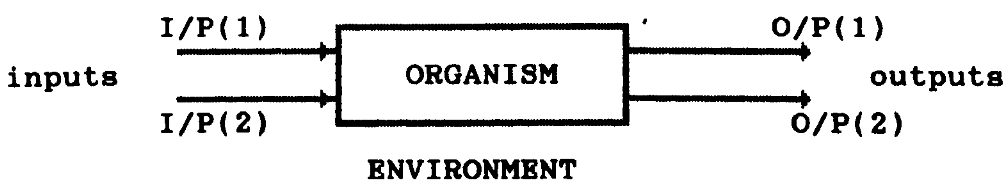
Of course, as before, different interference factors could change the shape of a pain-pleasure curve, or cause a shift in its x-y axis. Similarly, compound, conflicting, or competing pain-pleasure mechanisms can result in a much more intricate curve shapes, which have various nodes and troughs, or have sharp transitions from pain to pleasure areas.

In some observations of animal behaviour in nature (e.g., sea-gulls) it has been established that coincident activation of various conflicting innate drive mechanisms (maternal vs. escape) results in an intermittent non-stable response from the animal. The normal solution to this problem provided by evolution is to develope a hierarchical order for various motivational variables.

(iv) - CHOOSING HEDONIC CENTERS FOR SIMPLE LEARNING MODELS

Now, we shall discuss how to go about choosing hedonic principals for simple 'learning' models such as our experimental mobile robot.

In our argument we shall consider a simple hypothetical organism in an environment which exerts physical or information-based inputs upon it, as indicated below:



Now, lets suppose that the organism starts from a non-goal-directed random behaviour. Then, we (as external observers) notice that one of the inputs, say I/P(1), which can represent "hunger" should be followed by a specific output, say O/P(2), which could represent "intake of food". In our judgement this sequence of events should increase the chances of survival for the organism. In other words, it is desirable that an association should be established between I/P(1) and O/P(2).

Of course, not every pairing of associations will make sense, and both external causalities of events and, also, internal connectivities of the structure of the organism have to be taken into account. For instance, if O/P(1) represented "sleep" then the association of I/P(1) and O/P(1) (hunger and sleep) would not be a very positive one (normally).

Hence, it is clear that unless the organism goes through the whole gamut of evolutionary development, involving genetic mutations and natural selection, then such useful associations cannot come about from within a random system that has no way of assessing the relevance of inputs to outputs or vice versa. Therefore, for any autonomously 'learning' system some internal measure of utility of actions is considered as necessary, even in systems that need to 'learn' simple associations.

The designer of a machine equivalent to the above hypothetical organism, which is required to 'learn' a simple rule of association, can either explicitly include the desired connections between inputs and outputs (e.g., I/P(1) and O/P(2)) within the design; or, alternately, can incorporate means by which the machine could discover such associations by chance, by search, or by following specific procedures. However, since we are interested in "acquiring" these associations, then we would like to include the capability of evaluating useful input-output sequences within the model. For example, the activation of some 'favourable internal states' could indicate that the machine should try to reinforce/repeat/reexperience the immediately preceding events.

Dedicated 'pain-pleasure' centers are probably the simplest hedonic mechanisms for directing learning. Yet, the dilemma exists that the very notions of "pain" or "pleasure" themselves have to be learned by a particular artificial system, on the basis of the contributions they make to its 'survival' (in a similar fashion to the natural organisms), and therefore an even more fundamental measure of hedony should be chosen. However, this regression in manifesting ever more simple hedonic principals cannot be resolved, unless we revert back to phylogenetically determined pain-pleasure centers.

A search for the most trivial hedonic system possible for machines could, therefore, on the one hand, result in a pair of very trivial internal indicators; and, on the other hand, at even a more fundamental level, yield hedonic indicators that are simply expressed in terms of the system's inputs and outputs. In our particular model the simplest pain-pleasure mechanism envisaged is to define a specific digit of the input code as the indicator of pain/pleasure (i.e, PAIN=0, PLEASURE=1).

The above discussions point to the fact that hedonic principles themselves can be developed during the process of learning and interactive experience. Hence, in design they should be modifiable in nature, and possibly involve some sort of hierarchy.

Now, having decided to incorporate a simple hedonic system into the model, then the various 'goals' of a self-goal-determining 'learning' system can be defined on the basis of the activations of its 'pleasure' centers and suppressions of its 'pain' centers. In a sense, the teleological behaviour 'learned' by the use of pain-pleasure mechanisms will not be so much "directed" by 'goals', as in the case of externally set goals, but will involve "searching" for goal events which the model can subsequently aim for. In

FIG.6.13 these two opposing methods of setting goals and the role of 'pain-pleasure' centers is illustrated.

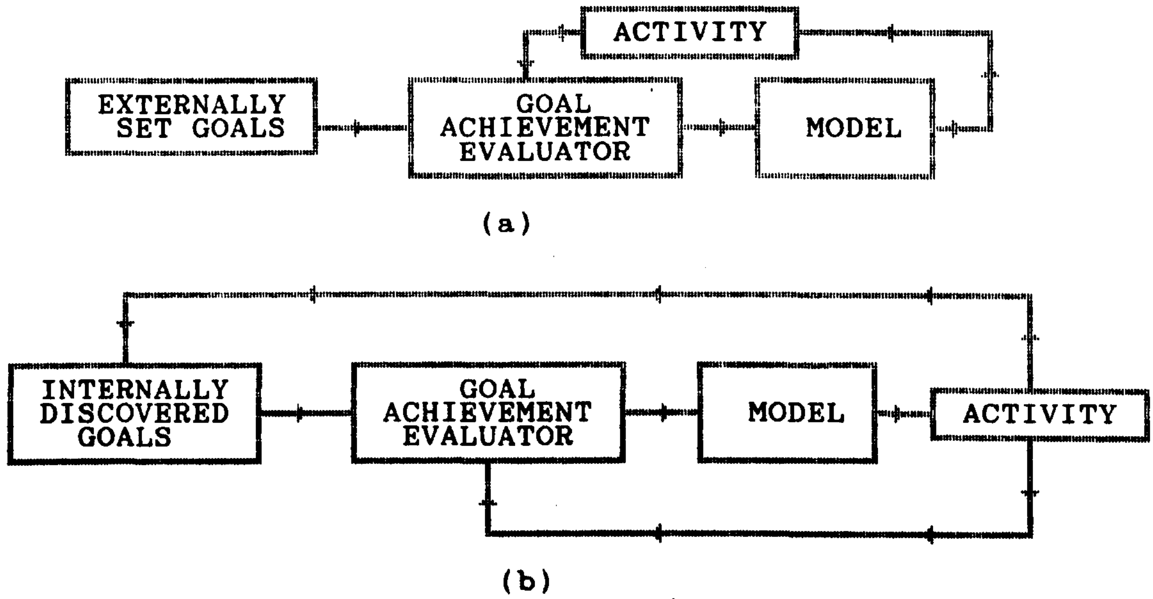


FIGURE 6.13. (a) - Externally set goals, (b) - Internally set goals.

For our particular mobile robot it was decided that rather than designing specific goal/drive centers, whose inputs were to be maintained within appropriate limits, the goal events would simply be defined in terms of conditions set upon one or a combination of inputs and/or outputs.

In some similar experimental exercises, to convey an artificial sense of a 'need' mechanism in the robot, special inputs such as battery level indicators are incorporated which is deemed to 'motivate' the machine towards searching for a battery-charger outlet. Although, drive and need mechanisms are well investigated attributes of animal behaviour, seemingly involved in every aspect of their learning, nevertheless, this superficial inclusion of an artificial 'sense' of a 'need' in the design, by express reference to motivational inputs, was not found to be conducive to our elementary approach to the modelling of the learning process.

Hence, no attempt is made to explicitly manifest 'survival' or other animal-specific aspects into the model. However, concepts such as drive and motivation can be later defined as composite higher order functions of the more basic hedonic parameters. For example, if "utility" (pleasure) in the model was defined as:-

"KEEPING RANGE AT MAXIMUM AND SPEED AT 1",

then, starting from a random pattern of behaviour, the robot's behaviour should converge towards movements in the centre of the room after a period of appropriate 'learning' - this being equivalent to explicitly stating a 'need' for the robot to stay in the middle of the room. Although, the exact consequence of such goal-seeking operations will not be obvious for many combinations of inputs and outputs at the outset; neither a 'convergence' towards a final 'learned' pattern of behaviour can be guaranteed.

6.2.3 EPISTEMOLOGICAL CONSIDERATIONS

Previously, we have referred to epistemological or knowledge related aspects of the learning process as one of its most important integral considerations, specially, when investigating the acquisitional aspects of learning. Knowledge and learning are implicitly connected, since when an organism improves/changes some facet of its behaviour through learning, it clearly has gained additional knowledge about itself or its environment.

Firstly, it must be emphasised that the scope of learning and knowledge acquisition even for an entity as simple as our model is almost limitless. To illustrate this point, the eight touch-switches, the shock-switch, and the seven-bit range input can provide $2^8 x 2 x 2^7 = 2^{16}$ different patterns of input; and the three-bit speed and four-bit direction outputs provide $2^3 x 2^4 = 2^7$ different output combinations. If we were simply to consider direct associations between single inputs and outputs, then there are $2^7 x 2^{16} = 2^{23}$ possibilities. Added to this are the temporal or combinatorial possibilities of associations between sets of inputs and outputs. Hence, it is conceivable that a great deal of knowledge, only bounded by the size of storage, can be represented and accumulated by our simple model.

The scanning program devised earlier was an example of the high level knowledge representation possible by the mobile robot, whereby a 2-dimensional map of the room was composed by appropriate programming of the robot. Many other conceptualizations are indeed possible; of course, the level of such conceptualizations, which should also be compatible with the sensory complexity of the model, need to be described at the beginning of a modelling exercise, and objectives pinpointed.

The analysis of the nature and organization of knowledge is one of the important aspects of epistemology of learning, which in humans concerns issues such as the content of thoughts, concepts, mental-images, discriminatory factors, imaginations, etc. These analyses also involve the investigation of rules, operations and laws which underline such phenomena. The above issues form the core of many philosophical deliberations, and have instigated various diametric standpoints amongst workers of learning related fields (e.g., empiricism vs. rationalism). In earlier chapters of this thesis such issues were discussed to some extent, particularly in sections dealing with cognitive studies of learning.

If we suppose that an organism after undergoing a specific experience endures a change in its knowledge state - as a result of some external input/information, or as a consequence of observations made on the results of its responses, or following some internal processes, or some other method. Then, there are various important considerations that have to be taken into account.

Firstly, what was the level of the knowledge at the outset of learning. The organism can start from zero or very little knowledge and gradually build up (learn) a composite picture of its operational domain. It can also start with a priming of certain parameters, and upon confronting a particular experience develope specific patterns of knowledge rules. Alternately, at the most knowledge-intensive stratum, all applicable knowledge can be incorporated within the system at the start, and only refinements and omissions made to its organization after a period of operation.

Secondly, how can we know that knowledge has been acquired or changed within the organism. Here, the answer lies in evaluating a certain facet of the organism's behaviour or its "performance", before and after a specific experience. By judging certain responses, an "inference" can be made that learning caused a certain change in knowledge. In a way, knowledge could be thought of as an entity lying dormant within an animal, and is only activated when the attainment of some goal is desired. The organism may 'know' how to do many tasks, but, unless it is properly motivated it may not perform the task. Hence, the clear distinction between "learning" and "performance" should be appreciated.

Thirdly, what is the nature of the knowledge that an organism has, and how can it be represented. This could involve a variety of different forms. In its simplest manifestation, the implicit physical interconnection between inputs and outputs could represent knowledge. Yet, from our previous discussions we may speculate that to convey any interesting learning features a second representational meta-level is desirable. In this higher level, probably, the simplest form of knowledge representation is the mere recording of events, which can be a representation of combinations of input and output sequences, or a sort of 'copy' of the 'image' of temporal connectivities of inputs and/or outputs. Many other more complex conceptualizations are also possible for representing knowledge and its rules. Examples are: 'parameters in algebraic expressions', 'decision trees', 'graphs and networks', 'formal grammars', 'production rules', 'frames and schemas', or other linguistic/mathematical/logical/conceptual expressions.

The work in the discipline of A.I., specially in some of its branches such as 'expert-systems', is highly pivotal on the issue of representing knowledge; and there have been impressive developments in these areas within the past few years, particularly in commercial knowledge-based systems. Most of the effort has been directed towards the efficient collecting, storing, and utilizing of information by computers. For example, decision trees, classifiers or various rule-based systems can manifest the knowledge involved in a task domain; and programming languages (e.g., LISP, PROLOG) have been developed to manipulate these knowledge bases. However, learning or acquisition of knowledge is not, generally, emphasised in these fields. In general, in these "top-down" approaches to knowledge formation, only 'inferences' are made upon the knowledge-structure - although in some instances new rules may be 'extracted'.

On the other hand, in the "bottom-up" disciplines, such as connectivism or pattern-recognition, the interest is focussed on the more trivial representations of knowledge. Rather than describe objects and relations in terms of their underlying rules or higher level linguistic characteristics, their elements or basic features are identified. Therefore, more fundamental criteria are used to accumulate knowledge about the experiences of an organism - by compiling descriptions of objects and events in terms of subsets of more simpler entities.

Here, in our model, due to our particular cybernetic inclination to the modelling of learning, we will not include the higher strata of conceptualization and representation of knowledge. The level of the complexity of the model, its perceptive powers, and its behavioural 'learning' potentialities indicates that we should try to devise knowledge representation schemes which rely on simple consequences of inputs and outputs. Hence, the major concern should be the organization and restructuring of this knowledge base stored in the 'memory' of the system. We should also try to incorporate almost no, or minimal, prior knowledge into the system, enabling the model to build up a perception of the world around it in a progressive manner - even managing to develope its own goals.

6.2.4 MEMORY and GENERALIZATION

The faculty of 'memory', which can be defined as the capability to retain and recall past experiences, has been discussed in previous chapters of this thesis. In Chapter 2 the cognitive aspects of memory were scrutinized by outlining its characteristics and hypothesising about its various facets of formation, retention, recall, consolidation, decline, and utilization. Additionally, in the various 'learning' models discussed a memory-element of some sort was either implicitly or explicitly present. Here, we shall attempt to discuss such features in relation to our simple cybernetic model.

The most trivial form of recording a 'memory' of an external event is to make a permanent or temporary physical impression of that event. For example, a finger-print mark or the oscillations of a tuning-fork both indicate that a certain event occurred in the past. Although, certain animal cognitive systems (e.g., imprinting, iconic-memory) apparently exhibit this type of memory formation/retention characteristics. The real interest for the modelers of the learning process, actually, lies in the more complex forms of memory, which convey information in a more accessible and modifiable form.

Yet, it must be pointed out that if the non-trivial view of memory is adopted, then the presence of a memory element is not a necessity for all 'learning' models. For example, simple 'learning' systems can be envisaged which forge connections or associations between inputs and outputs, by modifying connection weights or creating new connections on basis of current percepts. Criteria used can be the instantaneous contiguity of events, or modifications of some mathematical function (e.g., conditional probabilities) for each recurrence of inputs.

For an inquisitive cybernetic modeler of 'learning' systems, clearly, looking simply at the external behaviour of the system or the processing of internal signals alone will not yield any results of much interest. Hence, some level of interaction of a model and its environment in the form of internal abstraction of external world is desirable; using which the model can make inferences, solve problems, and even hypothesise about its interactions with the environment. The knowledge-base which will provide the necessary framework for this abstraction can be contained within the memory of the system; and the simplest method envisaged for representing this type of knowledge in memory is to merely record that output O occurred following the occurrence of input I.

Now, to investigate a simple memory function further, lets consider a hypothetical system that has input $I=\{I_1,I_2,...\}$ and output $O=\{O_1,O_2,...\}$, and starts from a completely random I——>O transformations. The system has a simple goal mechanism in the form of pain-pleasure centers, and also has unlimited storage capacity.

The system, illustrated in FIG.6.14, commences operation by first reading an input I_i and choosing randomly an output O_j from the table of I——>O transformations (initially equally distributed). The transformation tables indicate the probability distributions for various outputs for any given input. Next, the goal evaluator will indicate whether the 'pain' or 'pleasure' centre has been activated, and a value $G_k=\{$ 1=pleasure , 0=pain $\}$ will be attributed to that particular I_i——>O_j pairing. If this pairing is already recorded in the infinite memory of the system (in one or several instances), then its previous goal values are looked up and modifications made to the selection probabilities (i.e., their distribution) of the I——>O transformation table - according to some criteria based on current and past goal values. However, if the sequence of the I——>O is novel, then the pairing together with its goal value is simply stored in the memory.

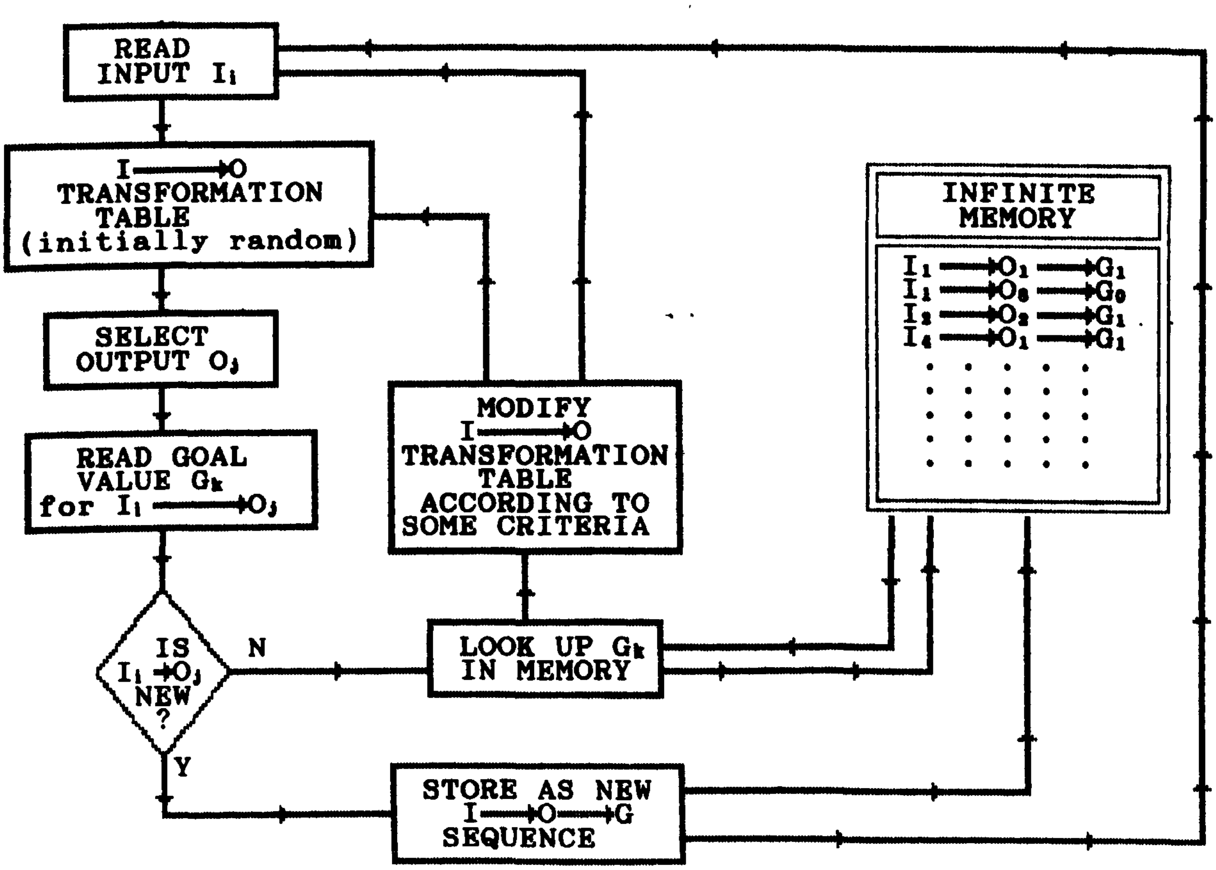


FIGURE 6.14. A simple idealised hypothetical 'learning' system with unlimited memory.

This simple scheme will in theory enable the model to 'learn' any existing desirable associations between inputs and outputs after an arbitrary period of operation. Similarly, using equivalent procedures, desirable relationships of output--->input pairings or sequences of inputs and/or outputs could also be discovered. Of course, all such learning is conditional upon the very existence of favourable underlying causality of inputs and outputs, and in many instances no convergence (or useful outcomes) may result - even after very long periods of operation.

The principal limitations and disadvantages of such idealised learning models are: firstly, the immense amount of storage capacity which might be required; and secondly, the time consuming nature of the processes involved in recalling, scanning and comparing events with stored past experiences; particularly, for compound sequences of inputs and outputs. Today's computers (digital, analogue, or parallel) are not really equipped to deal with such complexities for any non trivial 'learning' system of interest which operates on these basis. It is also not envisaged that future technological progress in computing machinery will promote this type of exhaustive approach to the modelling of learning.

Therefore, measures have to be taken to limit the amount of stored information in the memory. Various criteria can be used to do this. For example, we can record only I——>O relations which bring about a positive goal event (pleasure). But, perhaps, the best indications can be found in the discoveries made as to how the brain's equivalent limiting functions are performed in nature.

Seemingly, memory mechanisms in animals and humans retain the trace of a particular event for varying periods of time, depending on various underlying significance factors (e.g., motivation, attention). Two principal components, the 'short-term' and the 'long-term', are identified in natural memory mechanisms; enabling unused or irrelevant information to be discarded or replaced. In artificial models of 'memory' systems, also, similar principles can be employed. Normally, three principal criteria are used to manifest 'forgetting' in artificial systems, which are 'utility', 'age' and 'relevance'. Extreme care should, however, be taken so that earlier information is not overwritten by new information which somehow might be dependant on the earlier information.

Additionally, groupings of information are apparently made in the brain, by forming 'generalizations' of primary percepts; and, at a higher level, inputs are classified and heuristics formed to obviate the need for looking at every permutation of events and their significances.

The generalization of stimuli (inputs) is a characteristic of all learning processes in animals, and also a prominent feature of many of their perceptual mechanisms. The generalization process ensures that a response can be evoked by a broad range of stimuli which are appropriately similar to the stimulus already encountered by the organism. Yet, the critical levels of similarity at which two stimuli evoke the same response may vary for different animals, and also are dependant on the specific type of learning; in many cases, this range of similarity itself is refined by the animal during its learning process. In behavioural psychology the two types of generalization (non-learned and learned) are distinguished as 'primary' and 'secondary' generalization.

In our hypothetical learning model of FIG:6.14 a generalization could mean redefining groups of inputs as a new 'generalized' input. For example, the three sequences I_1-->O_1-->G_1, I_2-->O_1-->G_1, I_3-->O_1-->G_1 could be represented by a new sequence $I_{(123)}$-->O_1-->G_1. Alternately, some sort of logical or mathematical relationship could be discovered or hypothesised, which would enable a more parsimonious expression of events which elicit the positive goals.

6.2.5 'LEARNING' SCHEMES FOR SIMPLE CYBERNETIC MODELS

Having looked at some periphery issues and considerations deemed important to our level of investigation of the learning process and its modelling. We can now elaborate more on the way a working 'learning' system can be devised.

But, perhaps it is worth reiterating that our intentions in the scrutiny of learning are all directed towards its global and fundamental features. Hence, no particular pattern of behaviour (skill) is of special interest. Neither, shall we attempt to engage in discussions of higher 'mental' aspects of learning, such as consciousness, self-awareness, meaning, understanding, etc.

The idea is to try to define a limited but concise model, with limited scope of activities and task potentials, engaged within an environment that can exert a set of defined inputs upon it. The model should start with almost no external knowledge, aside from general directiveness information, and gradually 'learn' some interesting (non-trivial) behavioural patterns from its interactions with the environment. The question is posed, whether a consistent closed system could be built which exhibits the fundamental characteristics of the learning process, such as 'unpredictability', 'repeatability', 'self-improvement', 'permanence', 'generalization', 'recognition', 'recall', 'forgetting', 'relearning', etc.; and also whether a hierarchy of learning could be developed. Although, not necessarily copying the empirics of equivalent natural processes.

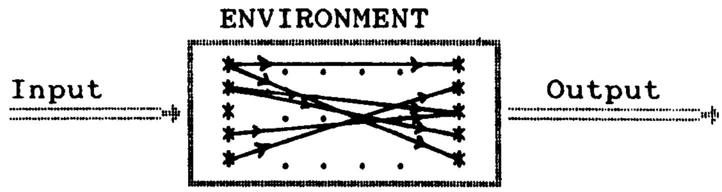
A look at other generalized attempts at the field of the modelling of the learning process reveals that, in the main, either their designers quickly abandon the global aspects of their supposedly all-encompassing endeavour, and focus on a particular task oriented domain; or, the vagueness of their abstractions, and the intricacies of mathematics involved, makes no accommodation for any coherent practical realization - of course, there are some exceptions (e.g., Andreae's 'learning' systems).

We have already discussed some of the fundamental features common to most learning which we would like to incorporate in our model, such as 'elementary goals', 'pain-pleasure reinforcement', 'short-term' and 'long-term' memory, 'generalization', and a 'knowledge-structure'. Next, we have to decide whether our model will be a kind of 'child-model' with no (or very little) prior knowledge; whether, it should start from a rigid knowledge structure and become more general purpose, or vice-versa; and whether, associations should be 'created' during learning, or simply 'refined'. Also, the extent of randomness incorporated into the system at the outset, to enable the model discover novel or unpredictable patterns of behaviour should be contemplated.

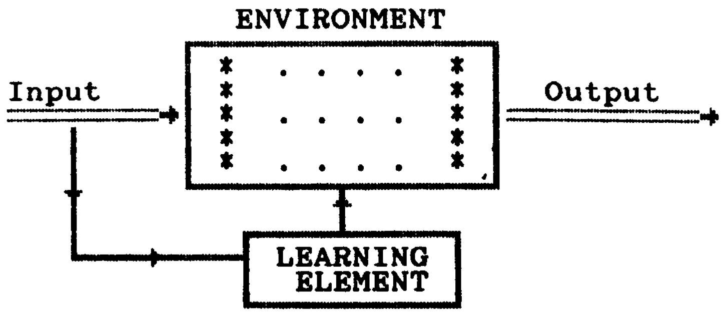
Some other important considerations are whether the environment or the domain of learning should be 'fixed' or 'changeable', in the changeable case a more critical evaluation mechanism should be designed. Similarly, the extent of external contributions made by a 'supervisor' or 'teacher' during the process of learning should be considered

Before we describe the details of one such 'learning' program, in the following, a simple ascending order of various types of simple learning will be proposed and discussed. The hypothetical "system" illustrated is considered to be a conglomeration of an array of input and an array of output cells, whereupon, associations could be established following a 'learning' phase. Here, the object of exercise is to identify and highlight the principal pathways of information exchange in a progressive manner:-

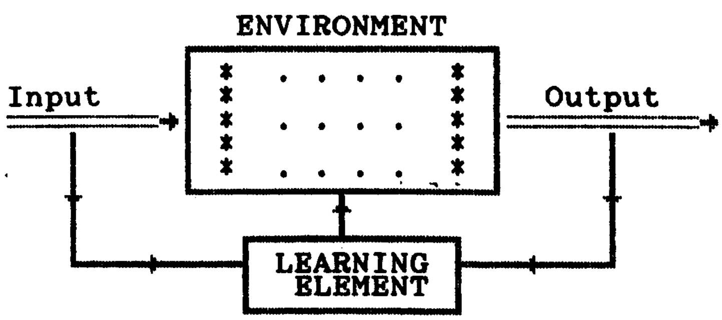
(a) - In the most trivial case, the connections are prewired, and there is no change in the associative connections between inputs and outputs during the operations of the system. As in the case of innate mechanisms or reflexive behaviours of animals.



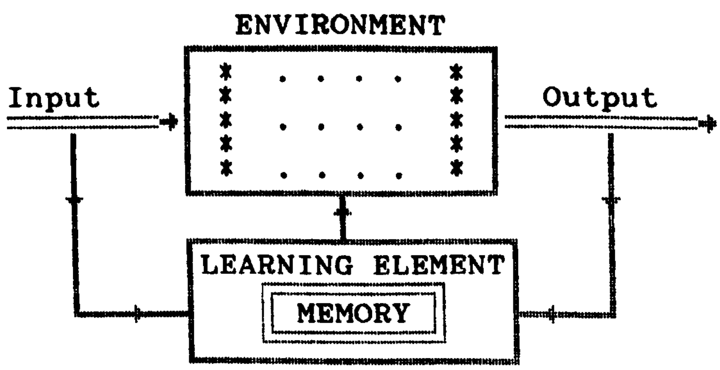
(b) - Initially there are no connections (or there are only partial connections), but after a period of time and going through a 'learning' phase some deterministic or stochastic associations are established, by a simple process of perceiving certain cue inputs. An example of this type of learning in nature is "imprinting" in birds.



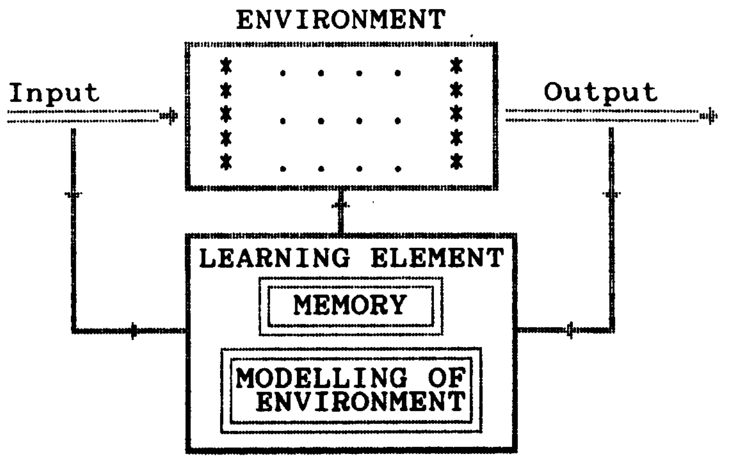
(c) - Initially there are no connections between the inputs and outputs, but now the system has a feedback loop from its outputs, and can evaluate its actions. Hence, after a 'learning' phase, the system learns associations based on some measure of utility (or hedony).



(d) - Internal mechanisms enable the system to store a memory of previous input–output sequences, and initiate actions on basis of present and past similar experiences.



(e) - Certain internal faculties allows the system to make abstractions about its external world, and hence make predictions or establish expectations about events and their consequences at some time in future. This is essentially the capability of conceptualizing a model of the external world.



6.2.6 AN EXAMPLE OF A 'LEARNING' PROGRAM FOR THE MOBILE ROBOT

In Chapter-3 (section 3.4) we outlined and discussed the elements of a generalised 'learning' system. A representation of the principal components of such a system is illustrated in FIG.6.15.

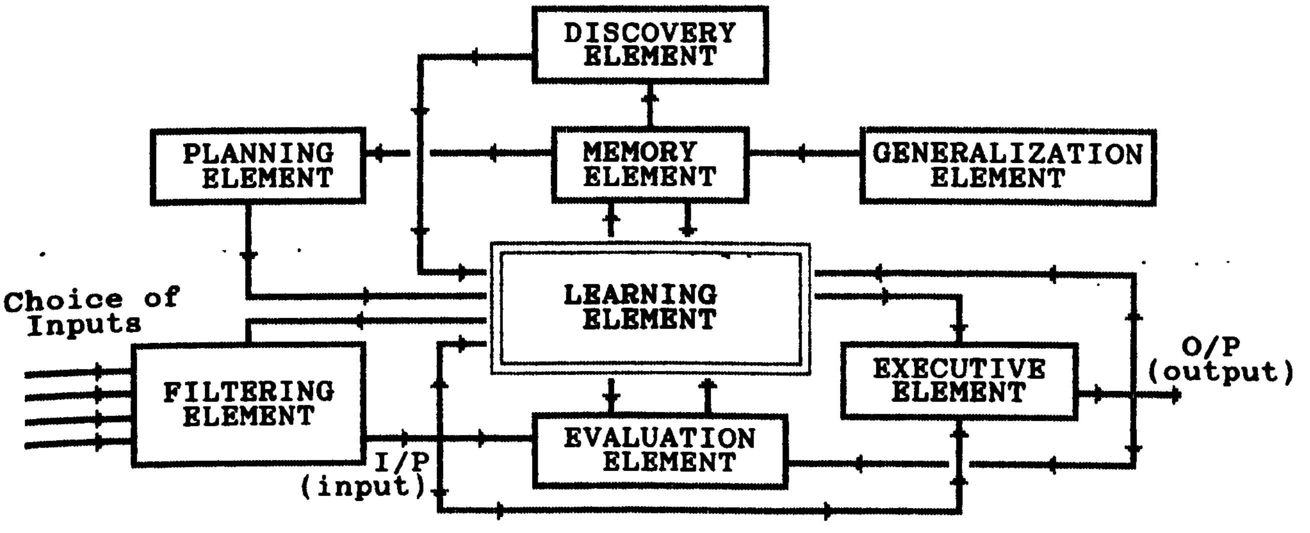


FIGURE 6.15. The eight major components of a generalized 'learning' system.

In the remainder of this chapter we will outline a specific example of a 'learning' program which can be used in conjunction with our hardware model. Although, the majority of the elements of the generalised 'learning' system of FIG.6.15 will be implemented in our proposed system, nevertheless, due to our level of approach some higher order features, such as 'planning', are omitted.

When we come to decide what to include in a 'learning' program choices are, indeed, enormous. We can incorporate a variety of concepts developed in behavioural and cognitive studies of learning; and also we can employ numerous techniques put forward by worker in learning related areas of disciplines such as A.I., P.R., Control Systems, Cybernetics, etc.

If the objective is to devise a program to simulate the simpler forms of adaptive behaviour in animals, for example, "habituation" or "conditioning", then the task is fairly straight forward. Firstly, a list of reflexive behaviours can be defined, specifying how a particular stimulus elicits a response. Secondly, criteria can be formulated for modifying these reflexive associations, or forging new associations, according to specific mathematical/logical rules. For instance, temporal contiguity of events could form the basis for such modifications in an exercise to simulate first order conditioning. Additionally, a simple evaluation unit may be desirable for keeping track of previous successful incidence of events. Other refinements can also lead to the realization of more elaborate models, which simulate actual laboratory observations of lower forms of learning behaviour much more realistically, and in more detail.

In our following proposed 'learning' program we will attempt to devise a scheme that will try to synthesise higher modalities of learning behaviour. The target of exercise is not the instinctive or basic reflexive type adaptive modifications of behaviour, but those categories of learning as defined in psychology by 'Trial & Error Learning' or 'Associative Learning'.

Our starting point in design will be to externally assign goals of the system. Previously, we have discussed various aspects of teleological behaviour and goal-selection in this chapter, however, in the absence of underlying instigating forces, similar to those found in nature (e.g., evolution, biological need mechanisms), we will not argue the case for 'utility' of actions from the point of view of the model itself. In other words, we will not reason why a certain goal seeking behaviour is "good" or "necessary" for the machine, or why the system "needs" to take a certain course of action. The goals will be set by an external observer, they are of the primary type of the hierarchy of goals discussed earlier (section 6.2.2(ii)), and do not immediately convey the type of behaviour (if any) they hope to induce.

Some general conditions set by the experimenter will be the goals which the model will try to maintain (or avoid). Although, even this type of goals can be randomly selected by the machine itself. But, again another externally set higher order evaluation system is necessary to judge the relative merits of the internally set goals - and recurssion of argument will continue.

Thus, for a simple manifestation of teleology, the goals of the model can be defined by setting conditions upon one or any combination of inputs and/or outputs. Even for our simple mobile robot the permutation of the choices based on the 16-bit input and the 8-bit output is enormous. Hence, for further simplification we can make the goal states independent of outputs of the robot - action (output) dependant goals are in principal similar to our simpler input only dependant goals, and can be incorporated within a later elaboration of the 'learning' program.

For example, if a 16-bit input was represented as:-

	8-bit touch sensors								7-bit range							shock-switch
I/P =	I_0	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}

Then, a goal G for the system can be defined as: "the sustaining of an input with active bits (i.e., "1's") at specified digits" - in the case of "pleasurable" or "attractive" goals. In other words:-

	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}	g_{12}	g_{13}	g_{14}	g_{15}
G =	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0

the significant goal bits ("1's")

In the case of "painful" or "aversive" goals having "0's" at specified digits of input will be the desirable objective of the system. Yet, the goal will be said to be activated only when the undesirable (painful) events occur - when "1's" occur at specified digits of input.

Every time all of the significant bits of the goal are at state "1", we can say that the goal is totally satisfied. Accordingly, partial activation of such bits can signify a lower degree of goal attainment.

Hence, a measure of 'goal attainment' can be defined by:

$$\text{Degree of Goal Attainment} = A = \sum_{j=0}^{15} g_j I_j , \quad \text{where} \quad 0 \leq A \leq 15.$$

Next, we have to consider how experiences (inputs) of the model are registered and stored in its memory. Initially, the mobile robot is assumed to be randomly roving in its environment, receiving inputs from its surroundings at a regular clock interval T. Similarly, outputs of the machine are also updated at the same clocking intervals.

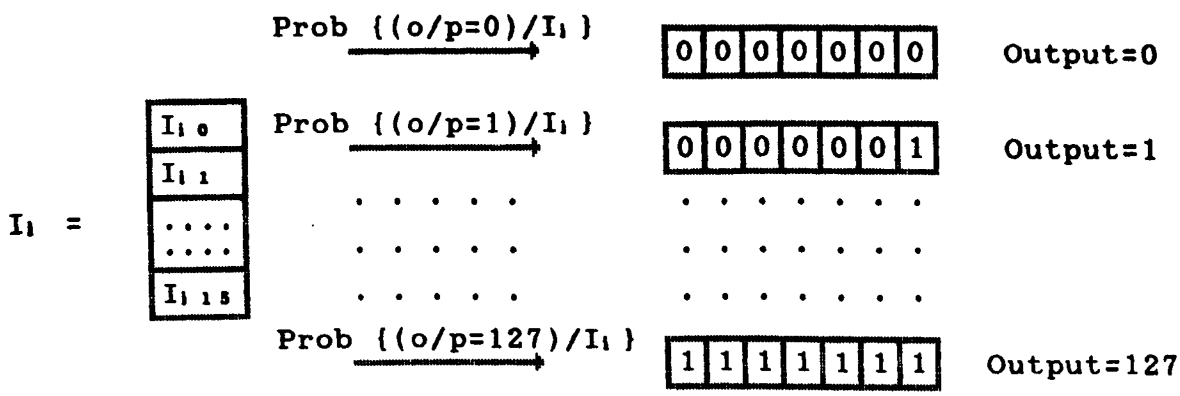
The interdependence of outputs and inputs is the issue of prime concern here, since, if a causality of response and stimulus is established, then we could simply arrange to select appropriate actions for attaining maximum goals. Such a dependance could take one of three forms. Either inputs and outputs are found to be wholly dependant deterministically; or partially/stochastically dependant; or completely independent upon each other.

The discovery of any deterministic or stochastic associations between inputs and outputs will, indeed, mean that the control which is incidental with the process of learning has been established. However, in real situations, the problem is much more complex, and such direct dependencies are very rare or difficult to pinpoint or define. Normally, it is the temporal or spatial sequencing or contiguity of a series of inputs and outputs which is the crucial factor in signifying causality.

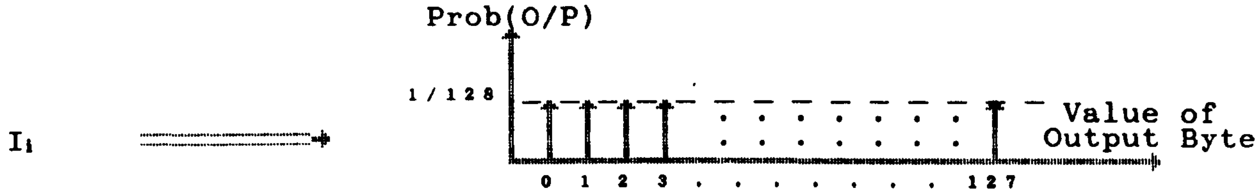
The robot starting from a totally random behaviour should, upon the incidence of goal events, be able to extract from interactions with the environment whether there is any dependance between its actions and percepts; hence, modify its behaviour towards sustaining the goal events; and also specify and record such dependencies for future utilization. The 'learning' scheme described here will try to achieve these objectives.

At the beginning of the exercise, each input state I_i can lead to any one of $2^7 = 128$ outputs ($2^3 = 8$ speed choices and $2^4 = 16$ direction choices) with equal probability:-

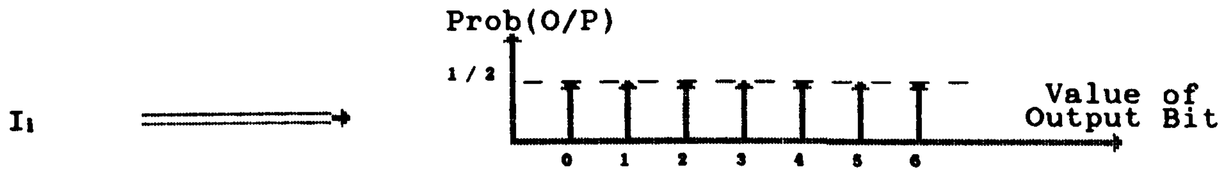
(initially all Probabilities = 1/128)



The initial random distribution can be interpreted in two ways. We can either say that for a given input the 128 outputs have an equal probability distribution:-



Or, we can say that for input I_1 each of the seven bits of output has probability 1/2 of being "1":-



Now, either method of depicting the probability distributions of outputs for a given input can be used, depending on various design considerations, and also on the computer storage limitations. But, the important issue here is the criterion which should be used for updating and modifying probabilities as the experiment progresses.

After each attainment of the desired level of goal event, the probability distribution of the immediately preceding occurrences is adjusted. Increasing the probability of the most recent output and decreasing the probability of all other outputs. This process of successive modification of probabilities should increase the expectation of goal events, provided there is an underlying causality between outputs and inputs.

However, if no improvement is observed in the model's performance, and a convergence cannot be detected in the output probability distribution curves, then, we can conclude that a single layer analysis of input-output relationships is not adequate for our problem. Hence, a deeper level of scrutiny is necessary. In other words, longer sequences of inputs-outputs (three or higher) should be looked at, and same probability adjustment procedures repeated for the more elaborate combined inputs.

Additionally, the critical values of goal attainment could also be adjusted to improve the performance of the system - a lower level of goal (pleasurable or painful) attainment (A) can signify the activation of probability modification phase of the program.

Both the variation of depth of look-back and the change of critical goal attainment level will be dynamically adjustable. Whereby, if after a number of input cycles no pronounced improvement is seen in performance, or no goal states are confronted, then appropriate modifications is made to the depth of analysis or the degree of goal attainment. Conversely, if very fast fluctuations of performance are observed, or goal states are constantly activated, then the inverse changes to above can be made.

Various practical considerations impels us to set certain constraints on the amount of information stored for analysis in the memory of the system. Firstly, not every single occurrence of input-output sequence will be stored, but only those which precede the instances of goal (painful/pleasurable) activation. Nevertheless, a 'Short-Term-Memory' (S.T.M) will hold a record of a prescribed number (k) of input-output sequences.

Secondly, once significant events are stored in the 'Long-Term-Memory' (L.T.M.) of the system, then the L.T.M. memory structure can go through various processes of reorganization:-

(a) - If an input (or a sequence of input-output) is not repeated over a number (l) of clock intervals, then it is 'deleted' from the L.T.M. as being an isolated case, or for being irrelevant.

(b) - If groups of different inputs converge towards the same output distribution pattern, then more 'generalized' new groupings of inputs can be defined.

(c) - Inverse of the above process, whereby, a compound input (or group of inputs) is 'divided' into individual inputs (or smaller sub-groups) - as the need arises to focus on a particular input.

The process of searching through the contents of the L.T.M., and comparing the incoming information with previously stored data, is another potentially time consuming phase of the 'learning' program - which could create a processing bottleneck. A solution considered is to translate the binary data of the L.T.M. into 'analogue' form, and hence use the much faster analogue-comparator hardware available to search through the contents of the memory. This solution is particularly useful when looking up 'sequences' of input-outputs rather than single inputs.

Once a previously stored output probability distribution is discovered in the L.T.M., then the modification of such probabilities can be carried out in a number of ways. The specific procedure used is, normally, considered as the core feature of a 'learning' program. An example of these modification rules will be outlined here.

Suppose input I_i is initially assigned a random output probability distribution function $P_B(O/P)$. Each output will have an equal probability of 1/128 to be selected:-

$$P\{(o/p=0)/I_i\} = 1/128,\ P\{(o/p=1)/I_i\} = 1/128\ ,\ .\ .\ .\ .\ ,\ P\{(o/p=127)/I_i\} = 1/128$$

and also, $$\sum_{j=0}^{127} P\{(o/p=j)/I_i\} = 1\ .$$

Now, using a random generator, suppose o/p=74 was selected, and at next incidence of clock pulse it is noticed that a goal state with attainment value a ($0 \leq a \leq 15$) is perceived. We look back at preceding input, which was I_i, and reinforce the probability of occurrence of o/p=74; at the same time reducing the probabilities of all other outputs occurring, given input I_i.

If the probability of occurrence of each output for a given input is represented by a 7-bit binary number (representing 128 different values), and the output itself identified by another 7-bit binary number. Then, one way of digitally storing the output distribution functions $P_B(O/P)$ in the L.T.M. is to only assign probability values to those outputs which have been reinforced in the past; since all other outputs will have equal and complementary probabilities. Hence, the entry in L.T.M. could simply entail:-

$$I_i \longrightarrow P\{(o/p=74)/I_i\} = 1/128 + F(a)\ ,$$

where, F(a) is an incremental function depending on goal attainment value a. Similarly, all other output (o/p≠74) probabilities, given input I_i, will be:-

$$P\{(o/p=m)/I_i\} = \{1 - (1/128 + F(a))\}/127\ ,$$

where, m is any output other than 74. However, these probabilities need not be stored in the L.T.M. of the system, and can simply be worked out at the output selection stage of the program when required.

Compound output probabilities (i.e., two or more outputs with reinforced probability values) can also be dealt with in a similar fashion. Obviously, in such cases the mathematics involved will be more complex. Yet, to simplify the probability adjustment procedures we can choose to only reduce the probabilities of outputs that have not contributed to goal attainment up to that point, leaving other previously reinforced output probabilities unaffected.

In the following diagram (FIG.6.16) the 'learning' program we have outlined here is illustrated in chart form, with brief explanations of its different functional elements. We have also indicated that the manifestation of some higher aspects of learning, such as 'expectancy', 'generalization', 'heuristic formation', 'theorem proving', or 'model building' can all be directed towards the data which will be gradually built up in the L.T.M. of the system.

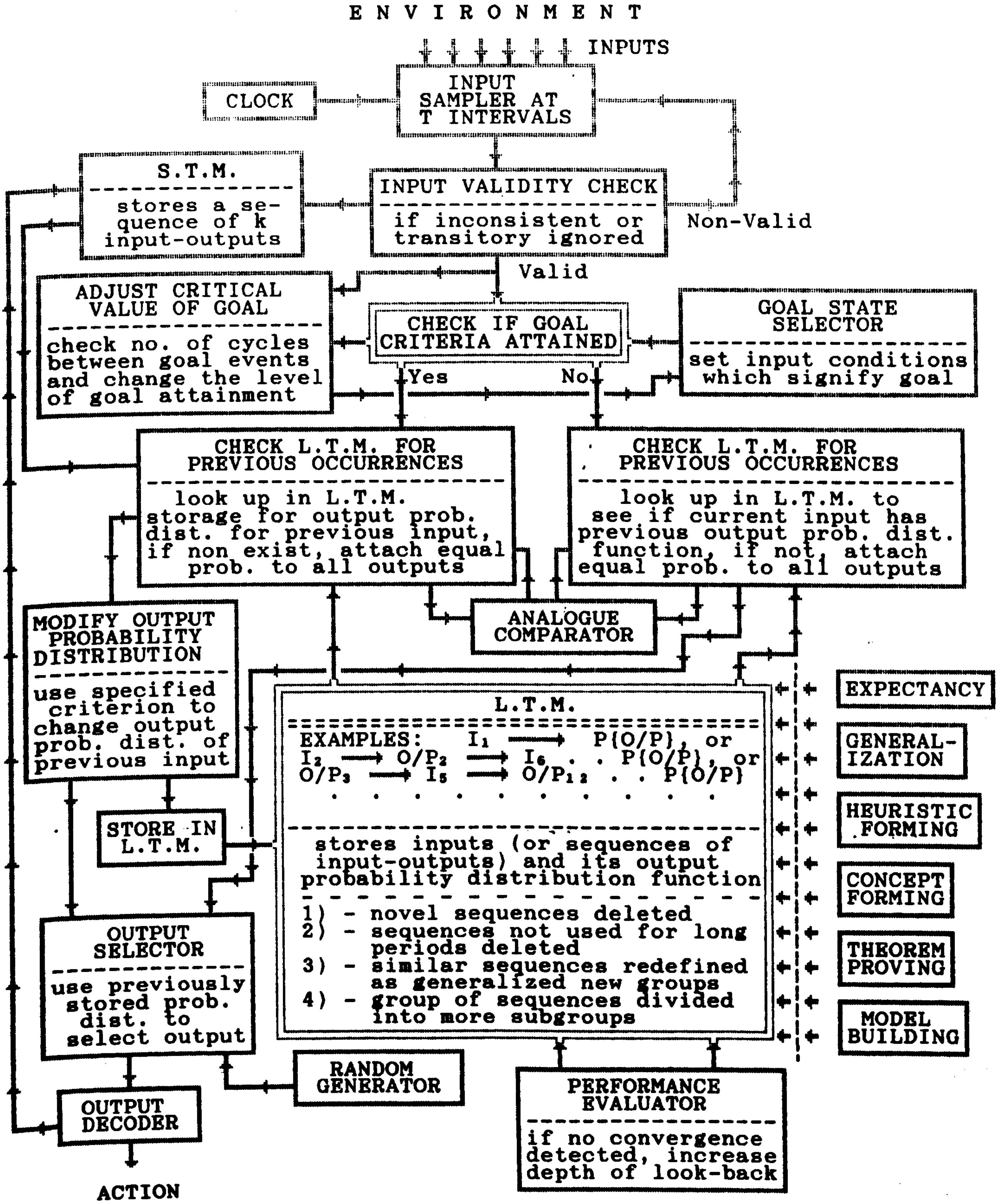


FIGURE 6.16. A block diagram of the tentative 'learning' program proposed. The program can enable the mobile robot, which initially starts with random activity, discover ('learn') input-output associations using externally set goal attainment and output modification criteria. "Input" in the above diagram refers to both single inputs or compound sequences of inputs and outputs. P{O/P}'s are output probability distributions.

6.3 SUMMARY CONCLUSIONS OF CHAPTER AND SOME POSSIBILITIES FOR FURTHER PURSUIT

In this chapter we discussed the many issues and considerations that can arise in the design of simple 'learning' models, in particular those that we have categorized as 'cybernetic'. We also embarked on the analysis and description of one such exercise in more detail, starting from the design stage to full implementation, concluding with a proposal for the implementation of a 'learning' program within the 'body' of the model. In following concluding notes of the chapter we will first enumerate the system elements which were highlighted as the principal components of all cybernetic 'learning' models.

As a more speculative postscript, we will also briefly consider the kind of hypothetical formalism which could provide a possible vehicle for the universal manifestation of learning in models. Indicating whether some of the well established mathematical methodologies such as the 'group theory' can be used to develope this elusive formalism. Finally, some areas of mathematics, will be advocated as promising avenues for further pursuit.

(i) - One of the central goals of this thesis was the identification of the fundamentals of 'learning' models. In our coverage of the various approaches to the task of modelling of learning we were confronted by many classifications, terminologies, depths of analysis, and formalism. In many instances the same concept or process was referred to by different labels. At other times the levels of analysis were so disparate that no direct correlates or relevances could be established between their descriptive languages.

Nevertheless, the recurrence of some principal elements has been evident within all 'learning' models discussed, in one form or other. So, in conclusion we can identify six central components as the very basic requirement of all (non-trivial) 'learning' systems. While reminding ourselves of the intricate and varied manner these components are manifested or are interrelated, the six elements are as follows:-

- THE INPUT ELEMENT: Incorporates all sensory input processes, filtering, decoding of information, and generally making sense of the external world.

- THE ASSOCIATION ELEMENT: Incorporates processes which determine how connections are established between inputs and outputs, entropy (organization) increasing processes, or the processes which bring about the changes or improvements in the system. 'Learning' itself can be considered as an implicit facet of this element's function.

- THE EVALUATION ELEMENT: Incorporate all teleological processes, the setting of goals and assessment of their attainment.

- THE MEMORY ELEMENT: Incorporates processes related to the storing of external and internal events; and their subsequent recall, distortion, decay, or interference. This element has generally been characterised by a duality of 'short-term' and 'long-term' memory.

- THE GENERALIZATION ELEMENT: Is closely related to the 'association' and the 'memory' elements; and incorporates processes which enhance the efficiency of the system, or extend the capabilities of the system into higher levels. Conceptualization or theorizing can be considered as some of the advanced manifestations of this element.

- THE OUTPUT ELEMENT: Incorporates the final decision processes involved in selecting appropriate responses for the 'learning' system. Including all stochastic or deterministic executive processes.

(ii) - One of the prevailing characteristics of cybernetic models has been their reliance on some form of abstract tool, normally borrowed from well-established branches of mathematics (e.g., control-theory, conditional-probability, stability, etc.). It is also true to say that in some instances original mathematical elaborations have been made, or even completely new abstract language and framework devised. Hence, we can say that cybernetics does not have a distinct methodology as other traditional sciences; and the decline of cybernetics within the past decade could be partly attributed to the lack of a formal language for an adequate expression of intuitive ideas that so many workers from diverse fields of science were able to articulate within its boundaries.

Some pioneers of cybernetics, such as Ashby, indeed tried to formulate abstract frameworks for the "cybernetic" analyses of problems. But, by tradition cybernetic is a subject involved directly with real world problems, and two main shortcomings became evident in such efforts. Either, the techniques were too rigid and hence narrow in application; or, they were extremely vague conceptualizations and inapplicable in reality. Paradoxically, the most successful techniques were branched away from the mainstream of cybernetics, and evolved into new disciplines which seldom acknowledge their cybernetic heritage.

Here, we shall attempt to speculate on what kind of unifying cybernetic axiomatic system would be needed for the realization of a true cybernetic language. In particular, for our purpose of designing cybernetic 'learning' models.

At the core of mathematics lies the concept of natural number system, and the linear progression of the magnitude of a number (real or integer) as we move along the single dimensional axis from minus to plus infinity. Over the centuries many powerful classical theoretical frameworks, such as euclidian geometry or probability theory, have been developed which rely on this implicit relationship of magnitude of numbers (i.e., the number 6 has an implicit magnitudal relationship to the number 2, represented by a number 3).

Some branches of mathematics such as 'set theory' or 'logic' also incorporate other fundamental measures within their building blocks, namely, the notions of 'set' and 'truth'. Although, the natural number systems still remains an inseparable component of their infrastructure.

Similarly, in dealing with other more specific domains, higher mathematical abstractions have been devised which do not directly refer to the concept of natural numbers at their micro-level. The numerous mathematical methodologies used in A.I. and cognitive sciences are of this group.

The question to be posed from the cybernetic perspective is whether another elemental basis can be abstracted on par with the 'natural number system' which could be more appropriate to the task of emulating mental processes. Now, instead of numbers relating implicitly to each other by "magnitude" they can have "associative connections".

The comprehension and the realization of such a system at the higher semantic level is a fairly straightforward task. For example, in a semantic network, we can give labels to various items such as "chairs", "tables", "desks"; and simply from a definition "items of furniture" we can see an inherent relationship amongst these distinct labels which does not change according to which combination of labels are considered together. Furthermore, other inferences can be made about these labels, such as what material they are made from. Mathematically these conceptualizations can be represented by different means. For example, a multi-dimensional vector whose elements represent different 'features' could be defined.

However, for the systems which are based on the more rudimentary criteria (i.e., pattern-recognition, neural-nets, connectionist, or other cellular systems) the normal method of abstracting associative connections is to either formulate complicated equations, or to devise ad hoc coding systems. Moreover, in many realizations of such models the associations of elements are expressed in terms of physical/structural properties and other hardware specifics of the system which are invariably very difficult to analyze mathematically.

If an associative axiomatic basis was available, then indeed the problem which was addressed to in the P.R. section of the Chapter-5, the 'preservation of invariances', would be a trivial one. Since now every element of the system would implicitly carry information about its neigbouring elements. But, as to what the specifics of this so called 'associative number system' should be, we can only speculate in a vague manner.

An area of mathematics whose study could provide interesting insights into how a concise framework of theories can be developed using rudimentary non-numerical basis is the 'group-theory'. In group-theory operations and theories are defined in terms of elements of 'sets' (collections of items) rather than numbers. Although, in practice the majority of problems tackled by group-theory are those which use sets of numbers. In any case, it would be an interesting challenge to investigate the possibilities of utilization of 'group theory' as a framework for manifesting simple cybernetic 'learning' systems.

As mentioned previously, the majority of formalisms developed in current areas of A.I. are ad hoc problem-oriented techniques. In other rudimentary 'parallel 'approaches of today such as P.R., neural-nets or connectionist although the level of investigation is more generalized and fundamental, nevertheless, abstractions devised are fairly conventional and use techniques which are by and large domain-dependant.

Recently a new and non-conventional approach to the problem of mathematical expression of thought processes has been proposed by some workers in Oxford University (Deutsch, 1985; Penrose, 1986,1987,1988). Principally, their ideas originate from the conceptualizations of 'quantum-theory'; also, many of their arguments are developed on the basis of the findings of quantum-theory, by drawing parallels with quantum concepts. Their basic contention is that using such concepts alternate non-algorithmic computing procedures can be devised for the analysis of mental events. These procedures are neither stochastic nor heuristically based. Although most of the work in this field is highly theoretical and deals with aspects of computability of such procedures (as in the context of Turing machines), some of its adherents (e.g., Deutsch,1986) have proposed hypothetical designs for 'quantum computers' and 'Universal quantum computers', which are deemed to be much more efficient computational devices than the conventional processors in tackling 'parallel' problems - such as the intellectual functions.

This fresh outlook is also a promising avenue for further research into the problem of design of a possible mathematical abstraction to be globally utilised in 'learning' systems, and one which is more akin to the underlying mechanisms of the brain.

CHAPTER 7

CONCLUSIONS

7.0 INTRODUCTION

In this final concluding chapter of our thesis we will begin by a brief review of the areas covered in previous chapters, followed by a generalized discussion of some of the principal issues that have come to the forefront of our enquiry. We will also argue the case for the broad perspective of this work; and promote the rudimentary cybernetic approach to the problem of modelling of learning, while attempting to speculate on some fruitful course of future pursuits in this field.

7.1 A BRIEF REVIEW OF THESIS

Firstly, we set the historical and the philosophical background which led to today's diverse variety of learning-related subjects. Next, we outlined the dominant issues, disciplines, problems, dichotomies, definitions, and approaches involved in the investigations of the learning process; focussing on the main topic of our interest, namely, the modelling of learning and the design of simple cybernetic hardware 'learning' models.

Later, a more detailed study of the 'pure' empirical aspects of learning was embarked. The three major approaches were 'behavioural', 'cognitive', and 'brain studies'. The principal methods of analyses and domains of research were discussed; and various definitions, discoveries, and theories formulated within each approach scrutinized. Additionally, some other periphery approaches to the study of learning, such as 'evolutionary', 'social', 'educational', and 'developmental' were identified; and various categorizations and taxonomies of the learning process (and behaviour) were outlined.

Next, the attention was turned on the modelling of learning. A 'natural' vs. 'artificial' distinction was pursued for the way that such problems are approached - which basically reflects the difference between 'simulation' and 'synthesis'. The tools and techniques used for representation and analysis of learning models were surveyed; some of the abstractions (mathematical or other) used for realization of these models were also considered; and some notions propagated from explorations of such abstractions were discussed.

Workers in many secondary disciplines have, also, used the phenomenon of learning in devising working models, utilizing techniques and formalism particular to their discipline. A broad study of the distinct approaches to such synthesis problems was made, following loosely the historical precedence of subjects involved. Typical examples of work in each field were cited, the principal exponents identified, and central points of contention highlighted.

Finally, the specific hardware/software project undertaken as an exercise in construction of an autonomous cybernetic 'learning' model was described, and the main considerations involved in the design of the whole class of similar endeavours discussed.

7.2 "LEARNING": A MULTI-FACETED PHENOMENON

Learning is a very broad and multi-faceted concept, which if taken in its wide adaptive context embraces a diverse range of observations in the living world. To illustrate this point, some form of 'learning' or 'adaptive' process is thought to be involved in the following accounts:-

- The young of an animal is unable to carry out a task. Yet, after a period of development it is able to.
- A species of animals cannot, seemingly, do a task. Yet, after few generations they attain the capability of doing that task.
- An animal is reacting in a certain way to a stimulus. Yet, later on it reacts differently.
- An adult animal cannot perform a task. Yet, after a while it carries out the task quite proficiently.
- A person does not recognize a pattern. Yet, at next confrontation it does.
- A mental connection does not exist between concepts. Yet, after some time it is created.
- An object cannot be labelled with a name. Yet, later a linguistic label is established.

Further, To appreciate the role that "learning" plays in biological life, in the following we will attempt to tentatively delineate its attribution to the development of an organism's neural structure, while considering some other influencing and governing factors. The hypothetical organism (illustrated in FIG.7.1) begins life with part of its nervous system organised into specific structures which allows the organism to display its range of innate and reflexive behaviours. During its lifetime, the organism is affected by numerous 'external' influences, which, in turn, result in the reorganization of its neural structure. 'Internal' (temporal) processes, also, lead to other neural reorganizations. Moreover, interactions between externally and internally induced order brings about even more changes in neural structure. Now, the learning process is seen to be acting at all three levels of reorganizations of nerve structure. Which further demonstrates the ubiquity of this phenomenon, and the reason behind the complexity and the diversity of its explanations.

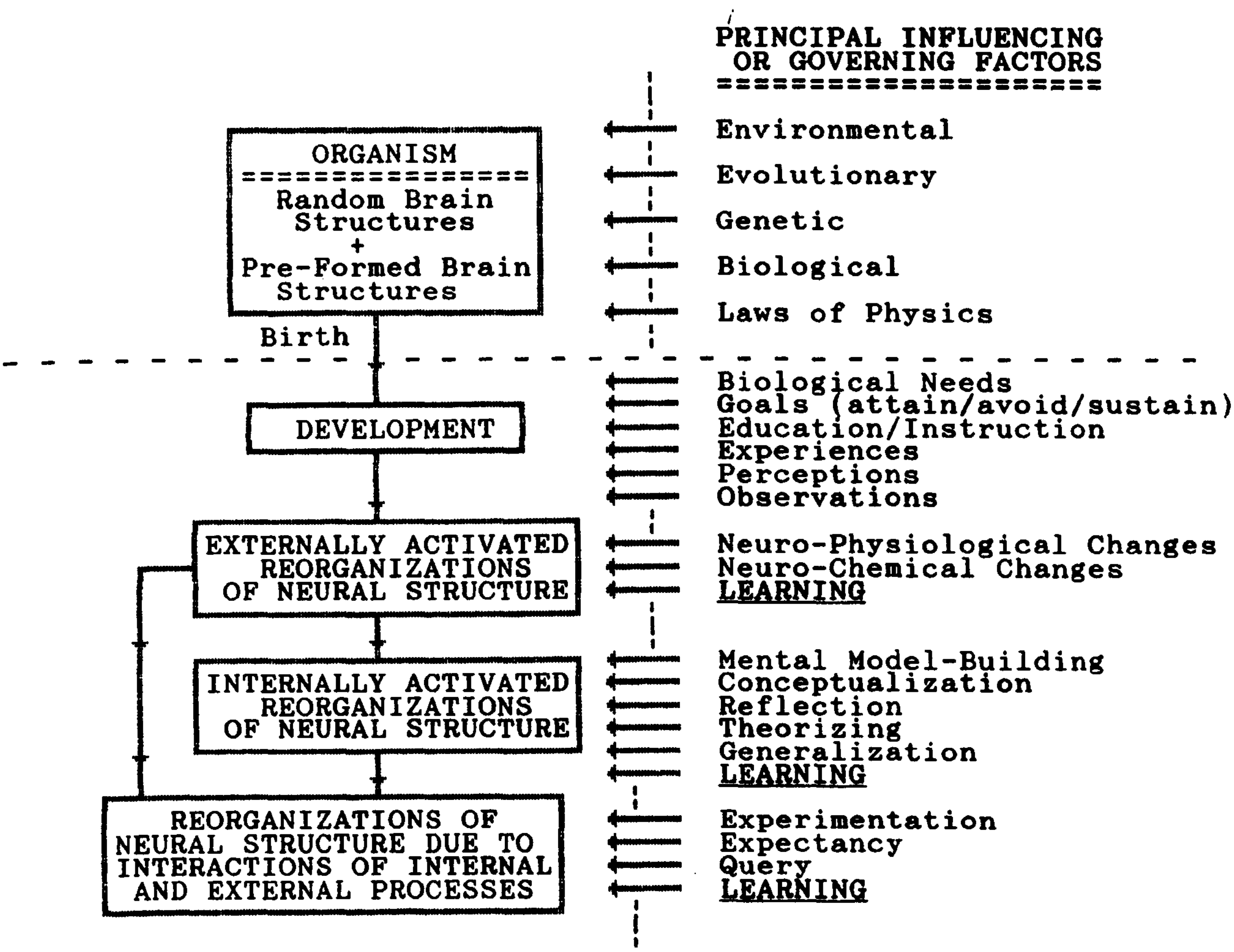


FIGURE 7.1. A simplified representation of stages of development of an organism's neural structure to highlight the role of "learning", with an indication of principal influencing/governing factors.

7.2.1 LEVELS AND METHODS OF INVESTIGATION OF THE LEARNING PHENOMENON

The question, foremost in our mind at the start of our enquiry, was whether we had good working knowledge of natural learning processes, behaviours, and mechanisms. So that, equipped with such knowledge, we could pass judgement on the various modelling attempts from different scientific corners in trying to simulate or synthesise the learning process within the boundaries of their own paradigm.

To acknowledge the multiplicity of layers of investigations of learning, let us consider a simple learning situation. For example, a human subject engaged in a cognitive experimentation on learning, such as learning to associate pictures and names, can be looked at. Although, it is now the undisputed belief that all changes occurring during a learning experience is registered and, hence, can only originate from the nervous systems. Nevertheless, the effects of such changes can be observed in a variety of manners:-

We can look at the molecular or sub-molecular changes occurring in the subject's nervous system. We can investigate the chemical, physiological, or electrical changes which result amongst the neurons. We can holistically study the chemical, physiological, or electrical changes coming about in groups of neurons, regions of the nervous system or its totality. We can theorize about the organizational cognitive aspects of the brain, and hence try to explain the changes in terms of our theories. We can construct various abstract or physical models to synthesise or simulate the appropriate neural processes. We can examine the external behavioural changes of the subject. We can devise hypothesis based on these external behavioural observations. We can investigate that particular observation in the context of several other identical or similar observations. We can attempt to design hardware or software models which can simulate the behaviour of the subject. We can ask the subject to linguistically describe the changes he experiences. Finally, we can investigate the broader phylogenetic, onthogenetic, or philosophical aspects of our learning experiment.

Therefore, the scope of investigations of the phenomenon of learning is astonishingly intricate and multi-layered. Furthermore, the domains of its application cover a very wide spectrum; encompassing humans, animals, machines, abstract systems, computer programs, etc.

Similarly, the theories proposed in the various explanatory planes have been different in character; and have involved assorted modes of manifestation, such as, mathematical, descriptive, physical, computer programming, etc. The degrees of 'applicability', 'generality', 'precision', 'complexity', 'predictability', and 'objectivity' of these theories have, also, been diverse. Some compromising precision to enable a wider scope of application; yet, normally, a resulting vagueness makes the validation of the theory or its disproof very difficult. Others by narrowing down the definitions, and increasing the objectivity, usually by use of abstract formalisms, are able to devise theories that are precise, but only applicable to certain problem domains. Here, the ever increasing elaborations of abstractions is such that after a while the initial objectives are, seemingly, lost by an engrossment in the theoretical details of analysis; and the whole field becomes very inaccessible from outside.

If we regard the sciences of learning as a whole, then there are many areas of vagueness, contradiction, controversy, cross-definition, and inconsistency. However, this is hardly surprising due to the intricacy of the phenomenon under observation; and the diversity of backgrounds, methodologies, descriptive languages, prejudices, and preferences of the workers involved in its investigation.

Previously, in Chapter-1, we had outlined the connectivity of the numerous disciplines directly or indirectly involved with the study of learning in a "tree" format. But, for a person entering into this subject, perhaps a better metaphor would be a "maze" of corridors, some interconnected; but, mostly having a singular pathway, with only occasional glimpses into other avenues - through windows stained with a particular bias. The workers within each field not only pursue the objectives of their own paradigm, but spend effort on trying to trivialize or discredit alternate views. Once, a unilateral blinkered view of the subject is adopted, then progress within its course of development, normally, distances the subject away from other related topics; leading to either further specialization, or extreme complexity of abstraction.

Indeed, for a prospective researcher on the topic of learning and its modelling who intends to survey the subject there are no exact maps for the above described maze of corridors. Only indications of its nodes, and partial portrayal of interconnections or proximities. Hence, it was deemed vital, as one of the objectives of this thesis, to outline a broad perspective of the whole topic without a particular prejudice; so that, possibilities of collaborative research are not compromised. Next, we will summarize our conclusions about the principal modes of study of the learning phenomenon.

(i) - BRAIN-SCIENCES & NEUROLOGICAL STUDIES

The Brain Sciences approach to the studies of learning involves the investigation of the electrochemical activity and the physiology of the brain. Many important contributions have been made to the understanding of the nature of information transfer in nervous systems, the effects of localised brain damage on learning, and also the discovery of various functional contributions of brain regions to the process of learning. Yet, Brain Studies have, generally, steered away from holistic conceptualizations about the process of learning. An analogy which can be made, to stress the shortcomings of this approach to the problem, is to compare the task with an attempt to discover the workings of a motor car engine by either listening to sounds emitted from it, or analyzing individual components or small areas of the engine in isolation.

There are two principal underlying criteria in the studies of the nervous system. The 'reductionist' view which contends that isolated studies of components of the brain could help to build up a complete picture; hence, its proponents engage in the study of individual or groups of neurons, or try to characterize specific regions of the brain. While, the 'holistic' view refutes reductionism, and sees the brain as a kind of receptor (similar to a television set) of external information and knowledge; also, a great proportion of our percepts are considered to be dependant on external 'order' and 'causal relationships', hence, the 'knowledge' which is conveyed cannot be inferred by any sort of neurological investigations of the brain.

However, it is a truism that workers in brain sciences unanimously accept that there is, so far, no clear understanding of neuronal processes at work during the learning process. Hence, no solid global explanations have been put forward; but, only isolated analytical principals proposed, as glimpses into the complex overlapping systems at work.

(ii) - PSYCHOLOGICAL & BEHAVIOURAL STUDIES

The disciplines of behavioural and physiological psychology hold the heritage of the scientific scrutinies of learning; and have contributed a great deal to the main body of knowledge about the phenomenon of learning. The data collected, based mainly on experiments on animals, where controllability and repeatability were the principal concerns, and theories formulated on such data, have been the primary source of reference for the modelers of the learning process. Yet, by its very nature, psychology has not managed to sever itself from empirical results. Generally, different types or aspects of learning are investigated independently, without adopting a holistic view of the subject and trying to interconnect or relate topics.

Similarly, in devising learning theories only particular areas of learning are targeted, and no unifying proposals are made which can describe learning as a functional entity that can progressively manifest itself in different forms. The weaknesses of such discontinuities have been particularly highlighted when various fundamental notions are applied to 'human learning', which traditionally has not been the principal domain of investigation, due to practical or ethical reasons.

The way that learning is, typically, investigated in these sciences is to define certain quantifiable aspects of a behavioural pattern and use it to monitor performance; and, hence, hypothesise or organise all observations around such definitions. This 'external' view of the learning process has meant that there is very little hypothesising about the underlying mechanisms involved.

Although, some learning-related psychological concepts, such as 'drive', 'motivation', 'attention', 'memory', etc., show some direct correlations with certain neural substrate; nevertheless, most generalised theories proposed have many shortcomings and inconsistencies, and are constantly being refined on basis of new evidence - the principal obstacle being the task of integration of various concepts.

(iii) - COGNITIVE STUDIES

The workers in cognitive psychology have been concerned with the study of learning primarily in the domain of humans. The investigations of the cognitive scientists (and later on the A.I. scientists) have involved modelling and simulation of higher aspects of human learning, such as 'problem solving' or 'theorem proving'. This "top-down" paradigm approaches the problem by trying to breakdown the totality of the solution, which is normally taken to be known at the beginning (intuitively), into smaller components. A technique employed at many instances is to divide the goals into sub-goals. However, this methodology has been inadequate in dealing with simpler learning situations where elementary goals direct the learning process using 'experiences', and where knowledge is progressively 'accumulated' from simple principles - here, the contrast between the direction of approach of two diametric 'top-down' and 'bottom-up' trends is again emphasised.

The cognitive study of human learning, with the predominant involvement of 'language' and 'subjective thought', has reintroduced the philosophical dimension as a central facet of enquiry. While, the philosophical concerns, having instigated the whole chain of scientific scrutinies of the subject of learning, had been demoted to a much less significant position in the behavioural and neurological studies of learning.

7.2.2 THE CONTINUITIES AND THE HIERARCHIES OF LEARNING

The categorizations of the learning process has assisted the scientific scrutiny of the subject. Nevertheless, some workers have recognised various continuities amongst the different modes of the learning process, and have tried to define learning in terms of a hierarchy; while, others have attempted to explain its processes from a unitary point of view. Even, the commonality of the origin of 'instincts' and learning behaviour have been argued, despite having clear differences in underlying neural mechanisms (i.e., instincts imbedded in DNA genetic coding, learning contained in neural organization).

If a continuum was established across the diverse range of learning phenomena observed in nature, whereby, complex modalities of learning could be explained from the simpler ones; implying that the learning differences seen amongst animals are only "degrees" of their underlying neural endowments, and not fundamental "jumps". Then, it is conceivable that rules or principals could be described, analogous to the evolutionary laws of 'natural selection', for building up the hierarchy of learning processes.

An important issue raised in the above scenario is whether neurons or their equivalent abstractions are the only medium for the manifestation of learning in the true sense. Since, given rules to elaborate the whole gamut of learning from basic elements, the task of 'externalization' of 'learning' from its natural domain would be much simpler. The next step would be to devise a global 'learning operator', which could also be effectively applied to the artificial domains of machines or computers - for realizing any of the layers of the learning hierarchy, ranging from simple learning seen in unicellular organisms to human learning, and perhaps beyond.

The phylogenetic (or evolutionary) studies of the leaning process is one of the subjects which does try to establish a continuity between different strata of its hierarchy, and narrow down the distinctions. Certain laws of 'perception' (e.g., significance of high pitched sound as danger) can be observed amongst a diverse range of species, even those which are not phylogenetically closely related. Although, these similarities can be explained in terms of evolutionary selection processes, nevertheless, it does not obscure the fact that there is an underlying external commonality involved, governed by the causalities of the physical surroundings.

The commonality of the external world, genetic kinship of species, and various empirical observations have prompted workers in this field to speculate that the progression of learning from its simple modalities to its higher complex forms goes through specific stages in nature. Implying that the functional hierarchy of learning observed within an animal corresponds to the order of the evolutionary development of learning within species; the order of complexity of learning processes seen amongst different species; and the onthogenetic sequence of learning seen during the growth of an animal.

In our earlier analysis of learning we also drew some generalized comparisons between the processes of "learning" and "evolution", in a very fundamental way - equating some of the underlying aspects of each process in a broad sense. It was, indeed, observations of this nature which instigated the wide perspective of this thesis.

A tentative evolutionary ordering of adaptive behaviour is schematically outlined in FIG.7.2. The final outcome of the development of different types of reaction mechanism is indicated by examples in the final column. Each hierarchical level has a particular characteristic, novel in essence, but, not reducible to or deducible from those preceding it.

EVOLUTIONARY PHASES OF ADAPTIVE ACTIVITY	SCHEMATIC FUNCTIONAL DIAGRAM OF THE REACTION MECHANISM	EXAMPLES OF EVOLVED BEHAVIOUR
PHASE ONE =========== Random irritability of simplest types of living matter	Organism / Reaction-System Input from Envir. → RANDOM PAIRING → Output to Envir.	NON-DIRECTIVE BEHAVIOUR
PHASE TWO =========== Innate directive reactions of animals with simple sensory mechanisms	I/P → DIRECT/INDIRECT PAIRINGS → O/P	TAXIS, REFLEXES
PHASE THREE ============= More complex innate behavioural patterns based on sequences of inputs-outputs	I/P → SERIAL PAIRINGS (memory needed) → O/P	INSTINCTS, INNATE BEHAVIOURS
PHASE FOUR ============ Simple modifications of performance and behaviour patterns according to the 'utility' of actions	I/P → ASSOCIATIVE LINK INHIBITED/ACTIVATED → O/P	HABITUATION, SENSITIZATION
PHASE FIVE ============ Complex modification of behaviour pattern according to some measure of success or failure of output	I/P → OUTPUT EVALUATION (reward/punishment) → O/P	ASSOCIATIVE-LEARNING, CONDITIONING
PHASE SIX =========== Anticipation of inputs and outputs by reflection or processing of past experiences	I/P → INPUT EXPECTATION & INPUT PREDICTION → O/P	REASONING, THINKING, HIGHER-LEARNING

FIGURE 7.2. A possible schematic model for the evolutionary hierarchy involved in the development of learning in organisms. Each stage compliments previous stages, and contains all the earlier types of adaptations. The final column outlines some examples of behaviour developed at each level.

7.3 THE MODELLING OF LEARNING

As we have seen, although the volume of knowledge accumulated on the topic of learning is indeed massive, nevertheless, our real understanding of its underlying processes is fairly limited; and, hence, the theories and hypothesis proposed are speculative, fragmented, and at best isolated to a particular aspect or definition of the learning process.

The path of modelling of learning was adopted by many researchers as an aid to its better understanding, a tool for experimentation, or a method for verification of theories. The initial endeavours were mainly attempts at simulating some 'natural' aspect of the learning phenomenon, but, the development of this methodology has seen a gradual permeation of its notions into 'artificial' domains, where synthesis of learning is the prime objective.

Of course, an overriding and pivotal question has always been present, and that is whether the learning process is an implicit quality of "life", and, hence, should not really be applied to non-biological form. It is, indeed, sometimes implausible or impractical to translate certain features of learning which are clearly biologically oriented to the artificial domain. However, there are other characterizations of learning which could be manifested artificially in a different, or even a more efficient, form, only if the biological constraints were severed.

An inherent problem of designing 'artificial learning systems' also lies in the tools which are used in such exercises. Computers, machines and mathematical abstractions are different from brains in fundamental ways, yet, we regularly attribute them with human and animal values and expect them to behave in 'natural' ways. If we compare a simple process of learning, say in a child, to that in a computer we can see that, first and foremost, there is structural 'growth' and 'development' in the organization of the brain of the child, while the only changes occurring in a computer are in its program. Although, it must be pointed out that this issue is subject to interpretation, and various researchers either see an isomorphic relationship between neuro-physiological changes and programming changes, or the learning in a child is analyzed at the more abstract 'knowledge' level where similarities can be drawn between 'mental states' and 'states of a computer program'.

A characteristic observation on some learning models in fields such as brain-studies, psychology, cybernetics, A.I., P.R. is that their designers, although initially start off by a generalized study or theoretical analysis of a particular mode of learning, are too willing to hurriedly apply concepts defined for their machines, abstractions, or computer programs to human learning, without considering the underlying implications carefully. Another problem, also, arises when mathematical abstractions are exploded into areas of complexity and vagueness. Both these hasty elaborations are adverse distractions from the original objectives.

The task of designing a 'learning' model is, generally, approached from a previously fixed view point. A worker, competent in the formalism of his particular field, attempts to incorporate some adaptive or learning criterion in the design of a model. The elaborations of the model could involve mathematical (or other) techniques which are normally employed in that discipline (e.g., use of statistical techniques in psychology). Furthermore, it is intended that the immediate applications and usage of such 'learning' models should be found within the discipline itself. Hence, very few modelling exercises on learning approach the problem from a neutral 'pure' stand point, which would allow the study of the subject independently, without reliance on a specific domain of application or methodology.

Examples of 'learning' models surveyed in this work have been very varied. They have ranged from 'black box' type system models; descriptive models; abstract or machine simulations of learning processes; to computer programs and mathematical formalisms for synthesising learning, using deterministic, stochastic, heuristic, or algorithmic techniques.

In 'learning control systems' the 'behaviour' of a system was either totally or partially known, but the 'structure' was, normally, less known or unknown. After identifying significant inputs and outputs of the system, a 'learning' technique was used to discover a structure which would give rise to an appropriate system behaviour. Two principal criteria were used. Either, internal connections were deemed to be present, and simply their strength was 'adjusted' (increased/decreased); or, the connections were 'created' on basis of some associations.

In state-space, automata theoretical, or some other mathematical realizations of 'learning' systems the behaviour was normally specified by functions or rules which were not inherently obvious from the particular structure of its representation or form – the behaviour was, normally, selected (intuitively/scientifically/arbitrarily) on basis of some utility consideration. This aspect sharply contrasts the natural learning mechanisms, which singularly determine behaviour.

The neural-nets, the self-organizing systems, the pattern-recognition, and the connectionist approach to the modelling of learning were all closely related in terms of their underlying objectives, and also methodologies. The central point of contention in these reductionist views of the subject was that by designing networks of, mainly, identical elements, and using simple principles, it was possible to devise complete systems that behaved in a complex manner. More or less, imitating the way nature has, through the process on evolution, managed to manifest learning. An important question which was explicitly raised by these modes of investigations of learning was that if we have a particular system structure, displaying a pattern of external behaviour, and we intend to modify its behaviour, then, is it better to physically restructure the underlying organization of its elements, or try to modify the functional behaviour of each element - and what are the relative advantages.

The A.I. approach to the modelling of learning did not involve designing systems with inherent learning capabilities; but, principally involved devising computer programs which could 'synthesise' learning in some form - and not necessarily including 'natural' correlates. The common methodology was to use a 'knowledge base' of well developed ideas, and try to 'learn' (or discover) some features or implicit rules of the knowledge base by certain reorganizations of its structure. Also, a prominent feature of such models was their reliance upon an external 'teacher' or 'guide' for evaluating and directing their actions for maximum utility.

7.3.1 THE RUDIMENTARY CYBERNETIC APPROACH TO THE MODELLING OF LEARNING

There have been many attempts in the past 3-4 decades to devise cybernetic experimental 'learning' models. However, with the exception of very few, mostly have either concentrated on a narrow path of application or have involved some non-qualified speculation. Hence, it was one of the principal aims of this thesis to provide a broad framework for a systematic analysis of the problem - pinpointing the controversial areas, and highlighting the main issues involved.

A common feature of most cybernetic 'learning' models is their reliance on some form of mathematical formalism for representation and analysis of the behaviour of the model. Yet, to be able to display interesting learning capabilities, such as problem solving or concept learning, some designers have also incorporated the facility of abstracting the environment as well. Whereupon, the model is able to manipulate the 'knowledge' gained through its experiences, and make deductions by experimenting on the "model" of the real world rather than going through the actual steps themselves. The need for such a higher 'mental modelling' level becomes evident when we consider various learning situations in humans (or higher primates). For example, in trying to learn to solve a problem, different possibilities are examined within the mental framework of the problem and appropriate inductions made.

The approach adopted in the design of our own 'learning' hardware/software model, and by implication promoted as a fruitful course of investigation, was the rudimentary approach. The main advantages of such an approach to the modelling of learning are the simplicity and relative objectivity of definitions; limitations of output-input interactions; economy on the processing resources; and the possibility of implementing fundamental notions observed in the natural domain of learning. Conversely, the disadvantages are that their achievements are normally uninteresting, and such models are incapable of performing any complex tasks.

The first step of the design was to define the particular type or the class of problems that the model was expected to solve. For example, was it suppose to learn to 'play games', 'solve equations', 'analyze graphs', 'perform tasks', 'refine skills', or in the extreme be a generalised 'learning machine'. Our learning model could be categorized in the final 'generalized' grouping. However, the model could also be used for various classes of problem-solving experiments, or for the verification of simple learning theories.

The model designed was not intended for simulation of human learning. However, since it is the understanding and hopefully duplication of such higher forms of learning which is the final objective of all such research endeavours, some accommodation was made for the incorporation of certain important higher cognitive notions, such as S.T.M., L.T.M., recall, forgetting, generalization, etc. into the primitive model. Although, these elaborations were only made after considering the fundamental features of the basic model first.

The main challenge was to design a closed analytical model that can exhibit some learning within a limited universe of percepts and actions. It should be able to fully exploit all possibilities of model-environment interaction. Also, without, actually, replicating in detail the characteristics and peculiarities of natural learning processes, it should be able to demonstrate simple rules of knowledge reorganization which resemble patterns of 'recall', 'recognition', 'forgetting', etc. Additionally, the model should have a potential for elaboration into higher levels, depicting the hierarchy observed amongst natural learning processes.

The theoretical analyses of computing machinery have shown that, in principal, all Turing machines are capable of representing all patterns of behaviour expressible mathematically. Our proposed simple 'learning' model could, indeed, be also regarded as a Turing machine (with the addition of an unlimited memory); and, hence, theoretically could depict any pattern of behaviour. Yet, in practice, we must determine the appropriate level of complexity that the structure of a machine or system demands. In the same way that the human brain is not only utilised for dealing with trivial variations of percepts.

Further testaments for the appropriateness of behaviour and structure come from the evolutionary studies of the brain and behaviour. Fossil (and other) evidence indicate that neural potentialities were actually present before a species began displaying a certain pattern of behaviour, or exhibit some skills. For example, the acquisition of flight by birds was preceded by the development of the required brain regions in their cortex. Similarly, the manipulative and linguistic capabilities of man only followed the development of appropriate neural substrate. The apparent immense overcapacity and redundancy of the brain mechanisms is another clue into the way the process of evolution has been tackling this matter.

Therefore, we concluded that a model designed to 'learn' a task should have the potentiality for learning that type of task in the first place. In the case of our simple mobile robot, these potentialities were initially investigated. The "behaviour", the "goals", and the "tasks" of the model were all simply defined in terms of its inputs and outputs; however, from the sizable range of possibilities only those were chosen which signified an interesting feature, or conveyed a meaning, to us as human observers. Indeed, it would have been an interesting exercise to investigate the type of behaviour that such a simple model would itself develope, if it starts from completely random beginnings, and also chooses its teleological criterion randomly, without interference from external programmer/instructor.

Another choice had to be made as to what level of 'intelligence' should be implanted into the model at the beginning. Whether, the model was to start from a totally random behaviour and approach more directive behaviour; or, a lot of information was to be pre-programmed into the model, enabling a proficient behaviour from the start. In many similar exercises an intermix of these two extremes is evident. Yet, sometimes falsely the system is attributed as starting from random behaviour, while it is only 'programed' to behave as such. At other times, a model's directiveness is implicit from its physical construct, and care should be taken in attributing concepts such as memory, etc. to elements of these systems. In our particular 'learning' model the minimal initial level of knowledge was aimed for, and the directiveness of action was determined by simple hedonic criteria (set externally).

Other important considerations were the level of external guidance provided by a 'teacher' (instructor); and the choice of search techniques used for finding solutions - whether 'exhaustive'/'random'/'parallel'/'serial' or other techniques were to be employed for comparing information and discovering associations.

7.4 AN OVERVIEW

Today, we see that workers from numerous disciplines engage in research topics which involve "learning". The prevailing opinions dominant within a discipline are, in fact, the culmination of years of complex development and filtering of ideas within that discipline, and other branches of related scientific fields. Yet, the individual worker embarking on research, generally, does not consider the lineage of the subject, and simply occupies himself with specific problems posed from a particular perspective. Therefore, it is common that at some point of the development of his work due to the narrowness of approach, and a lack of appreciation of the width and the depth of the subject and its heritage, he is confronted with ambiguities, holes of knowledge, or inconsistencies.

The special nature of the subject of our scrutiny has meant that, unlike other classical fields of science, the avenues chosen for its investigation are not fully exploited before a new, and fundamentally different, line of enquiry becomes dominant. Progress is being made in most areas involved with the simulation and synthesis of learning in a parallel fashion, the pace of development being governed by trends (and fashions) as much as by other underlying technological factors or empirical discoveries. Hence, it is important that an awareness of the 'genesis' of the subject of learning should be maintained; and from time to time the previously tried (and seemingly dismissed) avenues revisited in the light of newly acquired knowledge or technological tools.

The cybernetic approach to the modelling of learning is an example of this type of partially traversed pathway of investigation, whereby, its high point of popularity was, perhaps, reached more than two decades ago. This approach, both, as an intuitive mode of enquiry, and also as a precise formal methodology providing an objective framework of analysis, promises future prospects which should be rigorously pursued. Nevertheless, it is believed that until an accompanying mathematical formalism which can depict the concept of 'association' is not formulated (more or less the way real numbers express the notion of 'linear relation'), true cybernetic 'learning' model cannot be fully realized.

In final analysis, although, our very broad approach to the problem of the modelling of learning, covering a multitude of subjects, has not furnished us with precise indications of the most fruitful lines of pursuit; or, yielded accurate 'global' judgements, without being too general; or, in fact, served to strengthen any convictions. Nevertheless, this wide perspective has immensely increased our knowledge, understanding, and the objectivity of the subject; and has been instrumental in loosening some preconceptions and prejudices about the potentialities of various approaches, by exposing their shortcomings and clarifying many underlying issues. Perhaps, the only definitive conclusion we can make is the reaffirmation of the complexity of the task, and the acknowledgment of the inadequacy of our understanding of the details of natural learning processes and their mechanisms.

BIBLIOGRAPHY & REFERENCES

Ackoff, R.L. and Emery, F.E. (1972):
On Purposeful Systems. Travistock.

Agrawala, A.K. and Kanal, L.N. (1983):
Perceptron. Encyclopedia of Computer Science and Engineering, Ralston, A. and Reilly, E.D. (Eds.).

Akama, K. and Ichikawa, A. (1979):
A Basic Model for Learning Systems. Proceedings of the 6th IJCAI, 1979.

Albus, J. (1979):
A Model of the Brain for Robot Control. BYTE, June-September 1979, (4-Parts).

Alcock, J. (1975):
Animal Behaviour: An Evolutionary Approach. Sinauer Associates.

Aleksander, I. (1970):
Some Psychological Properties of Digital Learning Nets. International Journal of Man-Machine Studies, volume 2.

Aleksander, I. (1971):
Microcircuit Learning Computers. Mills and Boon.

Aleksander, I. and Hanna, F.K. (1976):
Automata Theory: An Engineering Approach. Edward Arnold.

Aleksander, I and Stonham, T.J. (1978):
A Guide to Pattern Recognition Using Random Access Memories. Technical Memorandum, N/R/064, Brunel University, Department of Electrical Engineering.

Aleksander, I. (1978):
The Human Machine. Georgi Publishing.

Aleksander, I. (1980):
Whatever Happened to Cybernetics? Technical Memorandum, N/S/103, Brunel University, Department of Electrical Engineering.

Aleksander, I. and Burnett, P. (1983):
Reinventing Man. Penguin.

Aleksander, I. (1984):
Designing Intelligent Systems: An Introduction. Kogan Page.

Allen, S.A. and Rossetti, T. (1978):
On Building a Light-Seeking Robot Mechanism. BYTE, August 1978.

Allen, S. and Allen, S.A. (1979):
Simple Maze Traversal Algorithms. BYTE, June 1979.

Amarel, S. (1970):
Problem Solving and Decision Making by Computer: An Overview. Cognition: a Multiple View, Garvin, P.L. (Ed.), 1970, Spartan Books.

Amari, S. (1972):
Learning Patterns and Pattern Sequences by Self-Organising Nets of Threshold Elements. IEEE Transactions on Computers, November 1972.

Amari, S. (1977):
Neural Theory of Association and Concept-Formation. Biological Cybernetics, Volume-26.

Anderson, J.R. (1984):
Cognitive Psychology. Artificial Intelligence, Volume-23.

Andreae, J.H. (1964):
STELLA: A Scheme for a Learning Machine. Proceedings of the 2nd IFAC Congress, 1964.

Andreae, J.H. (1969):
Learning Machines: A Unified View. Encyclopedia of Linguistics, Information and Control, Meetham, A.R. (Ed.), 1969.

Andreae, J.H. and Cashin, P.M. (1969):
A Learning Machine with Monologue. International Journal of Man-Machine Studies, Volume-1.

Andreae, J.H. (1972):
The STELLA Learning Machine. Man-Machine Studies Report UC-DSE/1, University of Canterbury, Christchurch, New Zealand.

Andreae, J.H. and Cleary, J.G. (1976):
A New Mechanism for a Brain. International Journal of Man-Machine Studies, Volume-8.

Andreae, J.H. (1977):
Thinking with the Teachable Machine. Academic.

Andreae, P.M. (1977):
A Teachable Machine in the Real World. Man-Machine Studies Report UC-DSE/11, University of Canterbury, Christchurch, New Zealand.

Andreae, P.M. and Andreae, J.H. (1978):
A Teachable Machine in the Real World. International Journal of Man-Machine Studies, Volume-10.

Andreae, J.H. (1979):
Muscles for Thought. Man-Machine Studies Report UC-DSE/15, University of Canterbury, Christchurch, New Zealand.

Andreae, J.H. (1983):
Design of a Conscious Robot. Man-Machine Studies Report UC-DSE/22, University of Canterbury, Christchurch, New Zealand.

Andrew, A.M. (1958):
Machines Which Learn. The New Scientist, November 1958.

Andrew, A.M. (1959):
The Conditional Probability Computer. Proceedings of the Symposium on Mechanization of Thought Processes (II).

Andrew, A.M. (1967):
Learning Systems. Automaton Theory and Learning Systems, Stewart, D.J. (Ed.), Academic Press, 1967.

Andrew, A.M. (1970):
Statistical Theory and Self-Organization. Progress of Cybernetics, Volume-1, Rose, J. (Ed.).

Andrew, A.M. (1972):
Significance Feedback and Redundancy Reduction in Self-Organizing Networks. Advances in Cybernetics and Systems Research, Volume-1, Pichler, F. and Trappl, R. (Eds.).

Andrew, A.M. (1975):
Why Robotics? Kybernetes, Volume-4, 1975.

Andrew, A.M. (1978):
Learning in Networks with Minimal Disruption. Progress in Cybernetics and Systems Research (III), Trappl, R. and Klir, G.J. (Eds.).

Andrew, A.M. (1980):
Autopoiesis and Self-Organization. Progress in Cybernetics and Systems Research, Volume-(VII), Pichler, F.R. and Hanika, F.de.P. (Eds.).

Andrew, A.M. (1983):
Artificial Intelligence. Abacus Press.

Angyan, A.J. (1959):
MACHINA REPRODUCATRIX: An Analogue Model to Demonstrate some Aspects of Neural Adaptation. Proceedings of the Symposium on Mechanization of Thought Processes (II).

Anzai, Y. and Simon, H.A. (1979):
The Theory of Learning by Doing. The American Psychological Association, Inc.

Arbib, M.A. (1966):
Simple Self-Reproducing Universal Automata. Information and Control, Volume-9.

Arbib, M.A. (1969):
Theories of Abstract Automata. Prentice-Hall.

Arbib, M.A. (1970):
Cognition - A Cybernetic Approach. Cognition: a Multiple View, Garvin, P.L. (Ed.), 1970, Spartan Books.

Arbib, M.A. (1972):
The Metaphorical Brain: An Introduction to Cybernetics as Artificial Intelligence and Brain Theory. Wiley.

Armer, P. (1960):
Attitudes Toward Intelligent Machines. Computers and Thought, Feigenbaum, E.A. and Feldman, J. (Eds.), 1963, McGraw-Hill.

Ashby, W.R. (1952):
Design for a Brain. Wiley.

Ashby, W.R. (1956):
Design for an Intelligence-Amplifier. Automata Studies, Shannon, C.E. and McCarthy, J. (Eds.), Princeton University Press.

Ashby, W.R. (1956):
An Introduction to Cybernetics. Wiley.

Ashby, W.R. (1956):
Self-Regulation and Requisite Variety. Systems Thinking, Emery, F.E. (Ed.), Penguin, 1969.

Ashby, W.R. (1959):
A Simple Computer for Demonstrating Behaviour. Proceedings of the Symposium on Mechanization of Thought Processes (II).

Ashby, W.R. (1960):
Adaptation in the Multistable System. Systems Thinking, Emery, F.E. (Ed.), Penguin, 1969.

Ashby, W.R. (1962):
Principles of the Self-Organizing System. Principles of Self-Organization, Von Forester, H. and Zopf, W. (Eds.), 1962, Pergamon Press.

Ashby, W.R. (1967):
The Set Theory of Mechanism and Homeostasis. Automaton Theory and Learning Systems, Stewart, D.J. (Ed.), 1967, Academic Press.

Atkinson, R.C., Bower, G.H. and Crothers, E.J. (1965):
Introduction to Mathematical Learning Theory. Wiley.

Bagley, J.D. (1970):
A Paradigm for Artificial Intelligence. Progress of Cybernetics, Volume-2, Rose, J. (Ed.).

Banerji, R. and Mesarovic, M.D. (Eds.) (1970):
Theoretical Approaches to Non-Numerical Problem Solving. Lecture Notes in Operations Research and Mathematical Systems, Volume-28, Springer-Verlag, 1970.

Banerji, R.B. (1976):
A Data Structure Which Can Learn Simple Programs from Examples of Input-Output. Pattern Recognition and Artificial Intelligence, Chen, C.H. (Ed.), 1976, Academic Press.

Banerji, R.B. (1977):
A Comparison of Three Problem-Solving Methods. Proceedings of the 5th IJCAI, 1977.

Barto, A.G., Sutton, R.S. and Brouwer, P.S. (1981):
Associative Search network: A Reinforcement Learning Associative Memory. Biological Cybernetics, Volume-40.

Barto, A.G. (1984):
Simulation Experiments with Goal-Seeking Adaptive Elements. University of Amherst, Massachusetts, Department of Computer and Information Science.

Beck, A.H.W. (1967):
Words and Waves. World University Library.

Beletskii, V.V., Kirsanova, T.S. and Tchudinov, P.S. (1975):
Walking Control and Dynamics of Systems with Two Legs. Proceedings of the 4th IJCAI, 1975.

Bellman, R. (1964):
Dynamic Programming, Learning, and Adaptive Process. Computer and Information Sciences (1st), Tou, J. and Wilcox, R.H. (Eds.).

Bellman, R. (1978):
Artificial Intelligence: Can Computers Think? Boyd and Fraser.

Beurle, R.L. (1962):
Functional Organization in Random Networks. Principles of Self-Organization, Von Forester, H. & Zopf, W. (Eds.), Pergamon Press-62.

Biermann, A.W., Fairfield, J.R. and Beres, T.R. (1982):
Signature Table Systems and Learning. IEEE Transactions on Systems, Man, and Cybernetics, Volume-12.

Bobrowski, L. (1982):
Rules of Forming Receptive Fields of Formal Neurons During Unsupervised Learning Processes. Biological Cybernetics, Volume-43.

Bobrowski, L. (1984):
Supervised and Unsupervised Learning Algorithms in Formal Neurons. Cybernetic Systems: Recognition, Learning, Self-Organisation, Caianiello, E.R. and Musso, G. (Eds.), 1984, Wiley.

Bond, A.H. and Mott, D.H. (1978):
Learning in a Mobile Robot. Artificial Intelligence Group Report No.102, Queen Mary College, London University.

Bond, A.H. and Mott, D.H. (1981):
Learning of Sensory-Motor Schemas in a Mobile Robot. Proceedings of the 7th IJCAI, 1981.

Booth, A.D. (1970):
Cybernetics and Information: A Simple, Self-Organizing Automaton. Progress of Cybernetics, Volume-2, Rose, J. (Ed.), 1970.

Boulding, K.E. (1978):
Ecodynamics: A New Theory of Social Behavior. Sage Publications.

Bowden, M.A. (1977):
Artificial Intelligence and Natural Man. Harvester Press.

Bower, G.H. and Hilgard, E.R. (1981):
Theories of Learning. Prentice-Hall.

Bozinovski, S. (1982):
A Self-Learning System Using Secondary Reinforcement. Cybernetica and Systems Research, Trappl, R. (Ed.), 1982.

Broadbent, D.E. (1965):
Applications of Information Theory and Decision Theory to Human Perception and Reaction. Progress in Brain Research, Volume-17.

Brown, D.J.H. (1975):
An Application of Concept Formation Techniques to the Learning of a Linear Language. Progress of Cybernetics and Systems Research, Vol:2.

Buchanan, B.G., Mitchell, T.M., Smith, R.G. and Johnson, C.R. (1978):
Models of Learning Systems. Encyclopedia of Computer Science and Technology, V-11, Belzer, J., Holzman, A. & Kent, A. (Eds.), Marcel Dekker.

Budden, F.J., (1972):
The Fascination of Groups. Cambridge University Press, 1972.

Bundy, A. (1983):
The Computer Modelling of Mathematical Reasoning. Academic Press.

Bundy, A., Boulay, B.D., Howe, J. and Plotkin, G. (1984):
How to Get a Ph.D. in AI. Artificial Intelligence: Tools, Techniques, and Applications, O'Shea, T. and Eisenstadt, M. (Eds.), 1984, Harper and Row.

Bundy, A. (1985):
Discovery and Reasoning in Mathematics. Proceedings of the 9th IJCAI.

Burks, A.W. (Ed.) (1970):
Essays on Cellular Automata. University of Illinois Press, 1970.

Bush, R.R. and Mosteller, F. (1951):
A Mathematical Model for Simple Learning. The Psychological Review, V-58.

Bush, R.R. and Mosteller, F. (1955):
Stochastic Models for Learning. Wiley.

Butler, J.A.V. (1976):
Modern Biology and Its Human Implications. Hodder and Stoughton.

Caianiello, E.R. (Ed.) (1967):
Neural Networks. Springer-Verlag.

Caianiello, E.R., De-Luca, A. and Ricciardi (1967):
Neural Networks: Reverberations, Constants of Motion, General Behaviour. Neural Networks, Caianiello, E.R. (Ed.), 1967), Springer-Verlag.

Caianiello, E.R. and Musso, G. (Eds.) (1984):
Cybernetic Systems: Recognition, Learning, Self-Organisation. Wiley.

Capehart, B.L. and Terry, R.A. (1968):
Digital Simulation of the Homeostat Modified to Show Memory and Learning. IEEE Transactions on Systems Science & Cybernetics July 1968.

Chan, D.F. and Phillips, S.R. (1975):
ROBANOV: A Range-Based Robot Navigation and Obstacle Avoidance Algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Sept-1975.

Chapman, B.L.M. (1959):
A Self-Organizing Classification System. Cybernetica, Volume-2.

Chen, C.H. (1976):
On Statistical and Structural Feature Extraction. Pattern Recognition and Artificial Intelligence, Chen, C.H. (Ed.), 1976, Academic Press.

Chen, C.H. (Ed.) (1976):
Pattern Recognition and Artificial Intelligence. Academic Press.

Cin, M.D. (1975):
Modifiable Automata and Learning. Prog. in Cyb. & Systems Research, V-2.

Cohen, P.R. and Feigenbaum, E.A. (Eds.) (1982):
The Handbook of Artificial Intelligence (Volume-III). Kaufmann.

Coles, L.S., Robb, A.M., Sinclair, P.L. Smith, M.H. & Sobek, R.R. (1975):
Decision Analysis for an Experimental Robot With Unreliable Sensors. Proceedings of the 4th IJCAI, 1975.

Conant, R.C. (1969):
The Information Transfer Required in Regulatory Process. IEEE Transactions on Systems Science and Cybernetics, Volume-SSC-5.

Connel, J.H. and Brady, M. (1985):
Learning Shape Descriptions. Proceedings of the 9th IJCAI, 1985.

Corcoran, D.W.J. (1971):
Pattern Recognition. Penguin.

Cudhea, D. (1979):
Artificial Intelligence. Recreational Computing, Volume-7, Numbers 4 & 5.

Culbertson, J.T. (1956):
Some Uneconomical Robots. Automata Studies, Shannon, C.E. and McCarthy, J. (Eds.), Princeton University Press.

Culbertson, J.T. (1963):
The Minds of Robots. University of Illinois Press.

Davis, L (1985):
Applying Adaptive Algorithms to Epistatic Domains. Proc. of the 9th IJCAI.

Dalenoort, G.J. (1982):
Visual Pattern Recognition by Neural Networks, Theoretical Considerations and a Simulation Experiment. Progress in Cybernetics and Systems Research, Volume-(VI), Pichler, F.R. and Trappl, R. (Eds.).

Dethier, V.G. and Stellar, E. (1970):
Animal Behavior. Prentice-Hall.

Deutsch, D. (1985):
Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. Proceedings of Royal Society London, A 400, 1985.

Deutsch, J.A. (1960):
The Structural Basis of Behaviour. Cambridge University Press.

Deutsch, S. (1967):
Models of the Nervous Systems. Wiley.

Dickinson, A. and Boakes, R.A. (Ed.) (1979):
Mechanisms of Learning & Motivation: A Memorial to J.Konorski. Erlbaum.

Distefano, J.J. (1967):
Feedback and Control Systems. McGraw-Hill.

Dixon, J.K. and Slagle, J.R. (1983):
Robotics. Encyclopedia of Computer Science and Engineering (2nd Ed.), Ralston, A. and Reilly, E.D. (Eds.).

Doberkat, E.E. (1978):
Some Investigations of Learning by Means of Dynamic Programming. Progress in Cybernetics and Systems Research, Volume-(IV), Trappl, R. and Pask, G. (Eds.).

Doran, J.E. (1968):
Experiments with a Pleasure-Seeking Automaton. Machine Intelligence, Volume-3, Meltzer, B. and Michie, D. (Eds.), Edinburgh University Press.

Doran, J.E. (1969):
Planning and Generalization in an Automaton/Environment System. Machine Intelligence, Volume-4, Meltzer, B. and Michie, D. (Eds.), Edinburgh University Press.

Doran, J.E. (1970):
Planning and Robotics. Machine Intelligence, Volume-5, Meltzer, B. and Michie, D. (Eds.), Edinburgh University Press.

Dreyfus, H.L. (1967):
Why Computers must have Bodies in order to be Intelligent. Review of Metaphysics, September 1967.

Dreyfus, H.L. (1972):
What Computers Can't Do: A Critique of Artificial Reason. Harper & Row.

Drozen, V. (1970):
A Mathematical Model of Associative Memory. Progress of Cybernetics, Volume-1, Rose, J. (Ed.).

Duda, R.O. (1970):
Elements of pattern Recognition. Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications, Mendel, J.M. and Fu, K.S. (Eds.), 1970, Academic Press.

El-Fattah, Y.M. and Foulard, C. (1978):
Cybernetics of Learning. Lecture Notes in Control and Information Sciences, Volume-9, Balakrishnan, A.V. and Thoma, M. (Eds.), 1978.

Elfes, A. and Talukdar, S.N. (1983):
A Distributed Control System for the CMU Rover. Proc. of the 8th IJCAI.

Elithorn, A. and Jones, D. (Eds.) (1973):
Artificial and Human Thinking. San Francisco.

Elstob, C.M. (1978):
A Cybernetic Analogue for Human Behavior. Progress in Cybernetics and Systems Research, Volume-(IV).

Elstob, C.M. (1980):
Information, Meaning and Knowledge. Progress in Cybernetics and Systems Research, Volume-(VII), Pichler, F.R. and Hanika, F.P. (Eds.).

Elstob, C.M. (1982):
The Formal and the Physical. Cyb. & Systems Research, Trappl, R. (Ed).

Emery, F.E. (Ed.) (1969):
Systems Thinking. Penguin.

Estes, W.K. (1959):
The Statistical Approach to Learning Theory. A Study of a Science, Koch, S. (Ed.), Volume-2, McGraw-Hill, 1959.

Estes, W.K. (1959):
Statistical Models for Recall and Recognition of Stimulus Patterns by Human Observers. Self Organising Systems: Proceedings of an Interdisciplinary Conference, 1959, Yovits, M.C. and Cameron, S. (Eds.), Pergamon Press.

Estes, W.K. and Neimark, E.D. (Eds.) (1967):
Stimulus Sampling Theory. Holden-Day.

Farley, B.G. and Clark, W.A. (1954):
Simulation of Self-Organizing Systems by Digital Computer. Transactions of IRE, PGIT 4, 1954.

Farley, B.G. (1959):
Self-Organizing Models for Learned Perception. Self Organising Systems: Proceedings of an Interdisciplinary Conference, 1959, Yovits, M.C. and Cameron, S. (Eds.), Pergamon Press.

Feigenbaum, E.A. and Feldman, J. (Eds.) (1963):
Computers and Thought. McGraw-Hill.

Feigenbaum, E.A. and McCorduck, P. (1984):
The Fifth Generation Artificial Intelligence and Japanese Computer Challenge to the World. Addison-Wesley.

Feldman, J.A. (1985):
Connections: Massive Parallelism in Natural & Artificial Intelligence. BYTE.

Ferguson, R. (1981):
PROLOG: A Step Toward the Ultimate Computer Language. BYTE, Nov-1981.

Fernald, L.D. and Fernald, P.S. (1978):
Introduction to Psychology. Houghton Mifflin.

Feurzeig, W. and Lukas, G. (1974):
A Programmable Robot for Teaching. Advances in Cybernetics and Systems, Volume-2, Rose, J. (Ed.), 1974.

Fikes, R.E., Hart, P.E. and Nilsson, N.J. (1972):
Some New Directions in Robot Problem Solving. Machine Intelligence, Volume-7, Meltzer, B. & Michie, D. (Eds.), Edinburgh University Press.

Fikes, R.E. and Nilsson, N.J. (1972):
STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. Artificial Intelligence, Volume-2, 1971.

Fikes, R.E., Hart, P.E. and Nilsson, N.J. (1972):
Learning & Executing Generalized Robot Plans. Artificial Intelligence, V-3.

Filo, A. (1979):
Designing a Robot from Nature. BYTE, February-March (2 Parts) 1979.

Findler, N.V. and McKinzie, W.R. (1969):
Computer Simulation of a Self-Preserving and Learning Organism. Bulletin of Mathematical Biophysics, Volume-31.

Findler, N.V. (1969):
Some New Approaches to Machine Learning. IEEE Transactions on Systems Science and Cybernetics, Volume-SSC-5, 1969.

Findler, N.V. (1983):
Artificial Intelligence. Encyclopedia of Computer Science and Engineering (2nd Ed.), Ralston, A. and Reilly, E.D. (Eds.).

Fischer, P.C. (1983):
Turing Machine. Encyclopedia of Computer Science and Engineering (2nd Ed.), Ralston, A. and Reilly, E.D. (Eds.).

Flynn, A.M. and Harris, J.G (1985):
Recognition Algorithms for the Connection Machine. Proc of the 9th IJCAI.

Fogel, L.J., Owens, A.J. and Walsh, M.J. (1966):
Artificial Intelligence Through Simulated Evolution. Wiley.

Ford, N. (1987):
How Machines Think. John Wiley & Sons.

Forsyth, R. (Ed.) (1984):
Expert Systems: Principles and Case Studies. Chapman and Hall.

Freitas, R.A., Healy, T.J. and Long, J.E. (1981):
Advanced Automation for Space Missions. Proceedings of the 7th IJCAI.

Friedberg, R.M. (1958):
A Learning Machine: Part I. IBM Journal, January 1958.

Friedberg, R.M., Dunham, B. and North, J.H. (1959):
A Learning Machine: Part II. IBM Journal, July 1959.

Friedman, L. (1967):
Instinctive Behaviour & its Computer Synthesis. Behavioural Science, V-12.

Friedman, L. (1969):
Robot Control Strategy. Proceedings of the 1st IJCAI, 1969.

Friedman, L. (1977):
Robot Learning and Error Correction. Proceedings of the 5th IJCAI-1977.

Frude, N. (1984):
The Robot Heritage. Century Publishing.

Fu, K.S. (1970):
Statistical Pattern Recognition. Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications, Mendel, J.M. and Fu, K.S. (Eds.), 1970, Academic Press.

Fu, K.S. (1970):
Stochastic Automata as Models of learning Systems. Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications, Mendel, J.M. and Fu, K.S. (Eds.), 1970, Academic Press.

Fu, K.S. (1970):
Learning Control Systems - Review and Outlook. IEEE Transactions on Automatic Control, April 1970.

Fu, K.S. (1971):
Learning Control Systems and Intelligent Control Systems: An Intersection of Artificial Intelligence and Automatic Control. IEEE Transactions on Automatic Control, February 1971.

Fu, K.S. and Tou, J.T. (Eds.) (1973):
Learning Systems and Intelligent Robots. Plenum Press.

Fukushima, K. and Miyake, S. (1978):
A Self-Organizing Neural Network with a Function of Associative Memory: Feedback-Type Cognitron. Biological Cybernetics, Volume-28.

Fu, L.M. and Buchanan, B.G. (1985):
Learning Intermediate Concepts in Constructing a Hierarchical Knowledge Base. Proceedings of the 9th IJCAI, 1985.

Garvin, P.L. (Ed.) (1970):
Cognition: a Multiple View. Spartan Books.

Gause, D.C. (1971):
New Considerations in Adaptive Programming Systems. Sixth International Congress on Cybernetics, 1971.

Gause, D.C. and Rogers, G. (1983):
Cybernetics and Artificial Intelligence. Cybernetics: Theory and Applications, Trappl, R. (Ed.), 1983, Hemisphere Publishing.

George, F.H. (1967):
Learning, Language and Computers. Automaton Theory and Learning Systems, Stewart, D.J. (Ed.), 1967, Academic Press.

George, F.H. and Stewart, D.J. (1967):
Computer Programming and Learning Theory. Automaton Theory and Learning Systems, Stewart, D.J. (Ed.), 1967, Academic Press.

George, F.H. (1971):
Cybernetics. Hodder and Stoughton.

George, F.H. (1972):
Cybernetics and Automata Theory. Internal Monograph, Cybernetics Department, Brunel University.

George, F.H. (1973,1961):
The Brain as a Computer. Pergamon Press.

George, F.H. (1977):
The Foundations of Cybernetics. Gordon and Breach.

George, F.H. (1986):
Artificial Intelligence - Its Philosophy & Neural Context. Gordon & Breach.

Gill, S. (1959):
Possibilities for the Practical Utilization of Learning Processes. Proceedings of the Symposium on Mechanization of Thought Processes-II.

Gilstrap, L.O. and Lee, R.J. (1960):
Learning machines. Bionics Symposium (1st): Living Prototypes - The Key to New Technology, Steele, E.J. (Ed.), 1960.

Giralt, G., Sobek, R. and Chatila, R. (1979):
A Multi-Level Planning and Navigation System for a Mobile Robot; A First Approach to HILARE. Proceedings of the 6th IJCAI, 1979.

Glorioso, R.M. (1975):
Engineering Cybernetics. Prentice-Hall.

Gregory, R.L. (1966):
Eye and Brain. World University Library.

Griffith, J.S. (1971):
Mathematical Neurobiology: An Introduction to the Mathematics of the nervous system. Academic Press.

Griffith, A.K. (1974):
A Comparison and Evaluation of Three Machine Learning Procedures as Applied to the Game of Checkers. Artificial Intelligence, Volume-5.

Guetzkow, H. (1962):
Inter-Nation Simulation: An Example of A Self-Organizing System. Self-Organizing Systems, Yovits, M.C., Jacobi, G.T. and Golostein, G.D. (Eds.), 1962, Spartan Books.

Gupton, J.A. (1979):
Talk to a Turtle. BYTE, June 1979.

Guzman, A. (1968):
Decomposition of a Visual Scene into Bodies. Proceedings of the Fall Joint Computer Conference, 1968.

Hand, D. (1978):
Pattern Recognition: The Ultimate Challenge. Personal Computer World, July 1978.

Hand, D.J. (1979):
Evolutionary Programming: An Intelligent Answer? Personal Computer World, April 1979.

Harmon, L.D. (1961):
Studies With Artificial Neurons: Properties and Functions of an Artificial Neuron. Kybernetik, Volume-1, 1961.

Harmon, L.D. (1962):
Natural and Artificial Synapses. Self-Organizing Systems, Yovits, M.C., Jacobi, G.T. and Golostein, G.D. (Eds.), 1962, Spartan Books.

Harth, E.M. (1966):
Brain Models and Thought Processes. Automata Theory, Caianiello, E.R. (Ed.), 1966.

Hatori, T. (1970):
On the Fundamental Notion of Cybernetics. Progress of Cybernetics, Volume-1, Rose, J. (Ed.), 1970.

Haugeland, J.H. (1985):
Artificial Intelligence: The Very Idea. MIT Press.

Hayes, P.J. (1970):
Robotologic. Machine Intlelligence, Volume-5, Meltzer, B. and Michie, D. (Eds.), Edinburgh University Press.

Hayes, P.J. (1975):
A Representation for Robot Plans. Proceedings of the 4th IJCAI, 1975.

Hayes, P. (1978):
Trends in Artificial Intelligence. International Journal of Man-Machine Studies, Volume-10.

Hebb, D.O. (1949):
The Organization of Behavior. Wiley.

Heiserman, D.L. (1976):
Build Your Own Working Robot. Foulsham-TAB.

Heiserman, D.L. (1981):
Robot Intelligence...with Experiments. TAB Books.

Hendel, R.J. (1979):
Mathematical Learning Theory: A Formalized, Axiomatic, Abstract Approach. Information and Control, Volume-41.

Hewitt, C. (1969):
PLANNER: A Language for Proving Theorems in Robots. Proceedings of the 1st IJCAI, 1969.

Hilgard, E.R. and Marquis, D.G. (1940):
Conditioning and Learning. Appleton-Century-Crofts.

Hinton, G.E. and Anderson, J.A. (Eds.) (1981):
Parallel Models of Associative Memory. Lawrence Erlbaum Associates.

Hinton, G.E. (1985):
Learning in Parallel Networks. BYTE, April 1985.

Hoffman, H.S. (1962):
The Analogue Lab: A New Kind of Teaching Device. American Psychologist, Volume-17.

Holland, J.H. (1970):
Outline for a Logical Theory of Adaptive Systems. Essays on Cellular Automata, Burks, A.W. (Ed.), University of Illinois Press, 1970.

Hollis, R. (1977):
NEWT: A Mobile, Cognitive Robot. BYTE, June 1977.

Hormann, A.M. (1962):
Programs for Machine Learning (Part-I). Information and Control, V-5.

Hormann, A. (1964):
How a Computer System can Learn. IEEE Spectrum, July 1964.

Hormann, A.M. (1964):
Programs for Machine Learning (Part-II). Information and Control, V-7.

Hull, C.L. (1943):
Principles of Behavior. Appleton-Century-Crofts.

Hunt, E.B. (1975):
Artificial Intelligence. Academic Press.

Hyden, H. (1970):
The Question of a Molecular Basis for the Memory Trace. Biology of Memory, Pribram, K.H. and Broadbent, D.E. (Eds.).

Iijima, J., Kanayama, Y. and Yuta, S. (1981):
A Locomotion Control System for Mobile Robots. Proceedings of the 7th IJCAI, 1981.

Inoue, H. (1985):
Building a Bridge Between AI and Robotics. Proceedings of the 9th IJCAI.

Jacker, C. (1964):
Man, Memory and Machines. MacMillan.

Jacobs, W. (1970):
Help Stamp Out Programming. Proceedings of the 4th Systems Symposium at Case Western Reserve University, Benerji, R.B. & Mesarovic, M.D. (Eds.).

Jacobs, W. and Kiefer, M. (1973):
Robot Decisions Based on Maximizing Utility. Proceedings of the 3rd IJCAI.

Justice, A. and Gervinski (1968):
"SOBLN". Mathematical Thinking in Behavioral Sciences: Rending from Scientific American, Messick, D.M. (Ed.), 1968, W.H. Freeman.

Kanayama, Y. (1983):
Concurrent Programming of Intelligent Robots. Proc. of the 8th IJCAI.

Kauffman, S.A. (1969):
Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets. Journal of Theoretical Biology, Volume-22.

Keckler, W.G. and Larson, R.E. (1970):
Control of a Robot in Partially Unknown Environment. Automatica, Vol-6.

Kent, E.W. (1978):
The Brains of Men and Machines. BYTE, January-April (4 Parts) 1978.

Kibler, D.F. (1985):
Parallelism in AI Programs. Proceedings of the 9th IJCAI, 1985.

Kiehn, R.M. (1979):
Artificial Intelligence and Entropy. BYTE, June 1979.

Kitagawa, T. (1973):
Biorobots for Simulation Studies of Learning and Intelligent Controls. Learning Systems and Intelligent Robots, Fu, K.S. and Tou, J.T. (Eds.), 1973, Plenum Press.

Kitajima, S. and Asai, K. (1973):
A Method of Learning Control Varying Search Domain by Fuzzy Automata. Learning Systems and Intelligent Robots, Fu, K.S. and Tou, J.T. (Eds.), 1973, Plenum Press.

Kleene, S.C. (1956):
Representation of Events in Nerve Nets and Finite Automata. Automata Studies, Shannon, C.E. and McCarthy, J. (Eds.), 1956, Princeton University Press.

Klir, J. and Valach, M. (1967):
Cybernetic Modelling. Iliffe Books.

Klix, F. (Ed.) (1979):
Human and Artificial Intelligence. North-Holland Publishing.

Kochen, M. (1974):
Representations and Algorithms for Cognitive Learning. Artificial Intelligence, Volume-5.

Kodratoff, Y. and Ganascia, J.G. (1984):
Learning as a Non-Deterministic but Exact Logical Process. In The Mind and the Machine: Philosophical Aspects of Artificial Intelligence, Torrance, S.B. (Ed.), 1984, Ellis Horwood.

Koffman, E.B. (1968):
Learning Games Through Pattern Recognition. IEEE Transactions on Systems Science and Cybernetics, Volume-SSC-4.

Kohonen, T. (1972):
Correlation Matrix Memories. IEEE Transactions on Computers, Vol:C-21.

Kohonen, T. (1977):
Associative Memory: A system-Theoretical Approach. Springer-Verlag.

Kohonen, T. (1984):
Self-Organized Formation of Feature Maps. Cybernetic Systems: Recognition, Learning, Self-Organisation, Caianiello, E.R. and Musso, G. (Eds.), 1984, Wiley.

Kolokouris, A.T. (1986):
Machine Learning. BYTE, November 1986.

Koplowitz, J. and Noton, D. (1973):
Motivation System for a Robot. IEEE Transactions on Systems, Man, and Cybernetics, July 1973.

Krushinsky, L.V. (1965):
Solution of Elementary Logical Problems by Animals on the Basis of Extrapolation. Progress in Brain Research, Volume-17.

Krylov, V.Yu. (1973):
Some Problems in Building Mathematical Models of Behaviour. Seventh International Congress on Cybernetics, 1973.

Kuzin, E.S., Pozdyak, G.E. and Fominykh (1975):
Planning the Activity of Robot with Artificial Intelligence. Proceedings of the 4th IJCAI, 1975.

Kvitashvili, A. (1970):
Cross-Connection Neuron Network Learning and Self-Learning Ability. Progress of Cybernetics, Volume-1, Rose, J. (Ed.), 1974.

Langley, P. and Carbonell, G. (1984):
Approaches to Machine Learning. Report No. CMU-CS-84-108, Carnegie-Mellon University, The Robotics Institute.

Lasker, G.E. (1974):
Threshold Automata. Advances in Cybernetics and Systems, Volume-2, Rose, J. (Ed.), 1974.

Lasker, G.E. (1974):
Theory of Mobile Automata. Advances in Cybernetics and Systems, Volume-2, Rose, J. (Ed.), 1974.

Laumond, J.P. (1983):
Model Structuring and Concept Recognition: Two Aspects of Learning for a Mobile Robot. Proceedings of the 8th IJCAI, 1983.

Lee, W. (1971):
Decision Theory and Human Behavior. Wiley.

Leiman, A.L. and Christian, C.N. (1973):
Electrophysiological Analysis of Learning and Memory. The Physiological Basis of Memory, Deutsch, J.A. (Ed.).

Lenat, B. (1977):
The Ubiquity of Discovery. Proceedings of the 5th IJCAI, 1977.

Lerner, A.Ya. (1972):
Fundamentals of Cybernetics. Chapman and Hall.

Levine, G. and Burke, C.J. (1972):
Mathematical Model Techniques for Learning Theories. Academic Press.

Lewis, R.A. and Bejczy, A.K. (1973):
Planning Considerations for a Roving Robot with Arm. Proceedings of the 3rd IJCAI, 1973.

Lewis, R.A. and Johnson, A.R. (1977):
A Scanning Laser Rangefinder for a Robotic Vehicle. Proceedings of the 5th IJCAI, 1977.

Lewis, T.G. (1980):
What Computers Cannot Do. BYTE, January 1980.

Lewis, J. (Ed.) (1974):
Beyond Chance and Necessity: A Critical Inquiry into Professor Jacques Monod's 'Chance and Necessity'. Garnstone Press.

Lindsay, P.H. and Norman, D.A. (1972):
Human Information Processing. Academic Press.

Locker, A. and Coulter, N.A. (1975):
An Outline of Teleogenic Systems Theory. Progress in Cybernetics and Systems Research, Volume-2.

Lofgren, L. (1982):
Excerpts from the Autology of Learning. Cybernetics and Systems Research, Trappl, R. (Ed.), 1982.

Loofbourrow, T. (1978):
How to Build a Computer-Controlled Robot. Hayden.

Luria, A.R. (1973):
The Working Brain. Penguin.

MacKay, D.M. (1956):
The Epistemological Problem for Automata. Automata Studies, Shannon, C.E. and McCarthy, J. (Eds.), Princeton University Press.

MacKay, D.M. (1962):
Self-Organization in the Time Domain. Self-Organizing Systems, Yovits, M.C., Jacobi, G.T. and Golostein, G.D. (Eds.), 1962, Spartan Books.

Mago, G.A. and Locker, A. (1975):
Towards a Mathematical Model of Goal Generation. Progress in Cybernetics and Systems Research, Volume-2.

Malik, R. (1973):
Why Build Robots. The New Scientist, March 1973.

Manaster, A.B. (1975):
Completeness, Compactness and Undecidability: An Introduction to Mathematical Logic. Prentice-Hall.

Marce, L., Julliere, M., Place, H. and Perrichot, H. (1980):
A Semi-Autonomous Remote Controlled Mobile Robot. The Industrial Robot, December 1980.

Marce, L., Julliere, M. and Place, H. (1981):
An Autonomous Computer-Controlled Vehicle. Proceedings of the 1st International Conference on Automated Guided Vehicle Systems, Jun-1981.

Marler, P. and Hamilton, W.J. (1966):
Mechanisms of Animal Behavior. Wiley.

Marsland, T.A. and Rushton, P.G. (1974):
A Study of Techniques for Game-Playing Programs. Advances in Cybernetics and Systems, Volume-1, Rose, J. (Ed.), 1974.

Maturana, H.R., Varela, F.G. and Uribe, R. (1974):
Autopoiesis: The Organization of Living Systems, its Characterization and Model. BioSystems, Volume-5.

Mays, W. (1970):
Philosophy and Cybernetics. Progress of Cybernetics, Volume-1, Rose, J. (Ed.).

McCorduck, P. (1977):
History of Artificial Intelligence. Proceedings of the 5th IJCAI-1977.

McCorduck, P. (1979):
Machines Who Think. W.H. Freeman.

McCulloch, W.S. and Pitts, W. (1943):
A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, Volume-5.

McCulloch, W.S. and Pitts W. (1947):
How We Know About Universals: The Perception of Auditory and Visual forms. Bulletin of Mathematical Biophysics, Volume-9.

McCulloch, W.S. (1959):
The Reliability of Biological Systems. Self Organising Systems: Proceedings of an Interdisciplinary Conference, 1959, Yovits, M.C. and Cameron, S. (Eds.), Pergamon Press.

McCulloch, W.S., Pitts, W.H., Lettvin, J.Y. and Maturana, H.R. (1959):
What the Frog's Eye Tells the Frog's Brain. Proceedings of the IRE, Volume-47, 1959.

McMurtry, G.J. (1970):
Adaptive Optimization Procedures. Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications, Mendel, J.M. and Fu, K.S. (Eds.), 1970, Academic Press.

Mednick, S.A., Pollio, H.R. and Loftus, E.F. (1973):
Learning. Prentice-Hall.

Meieran, H.B. and Gelhaus, F.E. (1986):
Robotics: A Technology on the March. Nuclear Engineering International, April 1986.

Meitus, V.Y. (1970):
Robots with Memory: Structure and Simulation. Progress of Cybernetics, Volume-2, Rose, J. (Ed.).

Mendel, J.M. and Fu, K.S. (Eds.) (1970):
Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications. Academic Press.

Mendel, J.M. and McLaren, R.W. (1970):
Reinforcement-Learning Control and Pattern Recognition Systems. Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications, Mendel, J.M. and Fu, K.S. (Eds.), 1970, Academic Press.

Mesarovic, M.D. (1962):
On Self Organizational Systems. Self-Organizing Systems, Yovits, M.C., Jacobi, G.T. and Golostein, G.D. (Eds.), 1962, Spartan Books.

Messick, D.M. (1968) (Ed.):
Mathematical Thinking in Behavioral Sciences: Rending from Scientific American. W.H. Freeman.

Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. (Eds.) (1983):
Machine Learning: The Artificial Intelligence Approach. Tioga Publishing.

Michalski, R.S., Carbonell, J.G. and Mitchell, (1983):
An Overview of Machine Learning. Machine Learning: The Artificial Intelligence Approach, Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. (Eds.), Tioga Publishing, 1983.

Michie, D. (1968):
Machines that Play and Plan. Science Journal, October 1968.

Michie, D. (1974):
On Machine Intelligence. Edinburgh University Press.

Michie, D. (Ed.) (1982):
Introductory Readings in Expert Systems. Gordon & Breach.

Michie, D. (Ed.) (1982):
Machine Intelligence and Related Topics. Gordon and Breach.

Michie, D. and Johnson, R. (1984):
The Creative Computer. Penguin Books.

Millar, P.H. (1971):
On Defining the Intelligence of Behaviour and Machines. Proceedings of the 2nd IJCAI, 1971.

Miller, S.L. and Orgel, L.E. (Eds.) (1974):
The Origins of Life on Earth. Prentice-Hall.

Miller, J.A. (1977):
Autonomous Guidance and Control of a Roving Robot. Proceedings of the 5th IJCAI, 1977.

Milner, P.M. (1957):
The Cell Assembly: Mark II. Psychological Review, Volume-64.

Milner, P.M. (1959):
Learning in Neural Systems. Self Organising Systems: Proceedings of an Interdisciplinary Conference, 1959, Yovits, M.C. and Cameron, S. (Eds.), Pergamon Press.

Minsky, M. (1961):
Steps Toward Artificial Intelligence. Computers and Thought, Feigenbaum, E.A. and Feldman, J. (Eds.), 1963, McGraw-Hill.

Minsky, M. (Ed.) (1968):
Semantic Information Processing. MIT Press.

Minsky, M. and Papert, S. (1969):
Perceptrons. MIT Press.

Mitchell, T.M. (1983):
Learning and Problem Solving. Proceedings of the 8th IJCAI, 1983.

Moravec, H.P. (1979):
Visual Mapping by a Robot Rover. Proceedings of the 6th IJCAI, 1979.

Moravec, H.P. (1981):
Rover Visual Obstacle Avoidance. Proceedings of the 7th IJCAI, 1981.

Moravec, H.P. (1982):
Stability of Automatic Guidance for a Mobile Robot. Proceedings of the National Conference on A.I., 1982.

Mott, D.H. (1981):
Sensory-Motor Learning in a Mobile Robot. Ph.D. Thesis, University of London, Queen Mary College, 1981.

Munn, N.L. (1971):
The Evolution of the Human Mind. Houghton Mifflin.

Nagata, T., Yamazaki, M. and Tsukamoto, M. (1973):
Robot Planning System Based on Problem Solvers. Proceedings of the 3rd IJCAI, 1973.

Nemes, T. (1969):
Cybernetic Machines. Gordon and Breach.

Newell, A., Shaw, J.C. and Simon, H.A. (1958):
Chess-Playing Programs and the Problem of Complexity. Computers and Thought, Feigenbaum, E.A. and Feldman, J. (Eds.), 1963, McGraw-Hill.

Newell, A., Shaw, J.C. and Simon, H.A. (1959):
A Variety of Intelligent Learning in a General Problem Solver. Self Organising Systems: Proceedings of an Interdisciplinary Conference, 1959, Yovits, M.C. and Cameron, S. (Eds.), Pergamon Press.

Newell, A. and Simon, H.A. (1972):
Human Problem Solving. Prentice-Hall.

Nillson, N.J. (1965):
Learning Machines. McGraw-Hill.

Nilsson, N.J. and Raphael, B. (1967):
Preliminary Design of an Intelligent Robot. Computer and Information Sciences, Volume-(II), Tou, J.T. (Ed.), Academy Press.

Nilsson, N.J. (1969):
A Mobile Automaton: An Application of Artificial Intelligence Techniques. Proceedings of the 1st IJCAI, 1969.

Nillson, N.J. (1971):
Problem-Solving Methods in Artificial Intelligence. McGraw-Hill.

Nillson, N.J. (1980):
Principles of Artificial Intelligence. Tioga Press.

Norman, D.A. (1982):
Learning and Memory. Freeman.

Nowakowska, M. (1976):
The Modern Origins of the Term "Cybernetics". Eight International Congress on Cybernetics, 1976.

Oestricher, H.L. and Moore, D.R. (Eds.) (1968):
Bionics Symposium 1966: Cybernetic Problems in Bionics. Gordon & Breach.

Open University (1971):
Evolution and Natural Selection. Science Foundation Course (Unit 19), Pentz, M.J. (Gen. Ed.), The Open University Press.

Open University (1972):
Biological Bases of Behaviour (Units 1–4, 11–13 and 16–17). Rose, P.R. (Gen. Ed.), The Open University Press.

Open University (1978):
Cognitive Psychology (units 3–4 and 22–28). The Open University Press.

O'shea, T. and Eisenstadt, M. (Eds.) (1984):
Artificial Intelligence – Tools and Techniques. Harper & Row.

Optner, S.L. (Ed.) (1973):
System Analysis. Penguin.

Palfi, A.F. (1976):
Controlling a Robot by PURR-PUSS, a Teachable Machine. Man-Machine Studies Report UC-DSE/9, Univ. of Canterbury, Christchurch, New Zealand.

Palfi, A.F. (1977):
A Real Tome "Body" and "World" for PURR-PUSS. Man-Machine Studies Report UC-DSE/10, University of Canterbury, Christchurch, New Zealand.

Palm, G. (1982):
Neural Assemblies: An Alternative Approach to A.I. Springer-Verlag.

Palvolgyi, L. (1982):
Network Modelling of Learning. Cyb. & Systems Research, Trappl, R. (Ed).

Paritsis, N.C. and Stewart, D.J. (1982):
Adaptational Problems in Natural Intelligent Systems with Changes in Environmental Variety. Cybernetic Systems Research, Trappl, R. (Ed.).

Partridge, D. (1985):
Input-Expectation Discrepancy Reduction: A Ubiquitous Mechanism. Proceedings of the 9th IJCAI, 1985.

Pask, A.G.S. (1962):
A Proposed Evolutionary Model. Principles of Self-Organization, Von Forester, H. and Zopf, W. (Eds.), 1962, Pergamon Press.

Pask, A.G.S. (1963):
A Discussion of the Cybernetics of Learning Behaviour. Progress in Brain Research, Volume-15, 1963.

Pask, A.G.S. (1967):
Man as a System That Needs to Learn. Automaton Theory and Learning Systems, Stewart, D.J. (Ed.), 1967, Academic Press.

Pask, A.G.S. (1969):
The Cybernetics of Behaviour and Cognition Extending the Meaning of "Goal". Internal Monograph, Cybernetics Department, Brunel University.

Pask, A.G.S. (1970):
Cognitive Systems. Cognition: a Multiple View, Garvin, P.L. (Ed.), Spartan.
Pask, A.G.S. (1970):
Learning & Teaching Systems. Survey of Cyb., Rose, J. (Ed), Butterworths.

Penrose, L.S. (1968):
Self-Reproducing Machines. Scientific American, Volume-200, 1959.

Penrose, R. and Isham, C.J. (Eds.) (1986):
Quantum Concepts in Space and Time. Clarendon Press, 1986.

Penrose, R. (1987):
Logic, Philosophy of Science and Epistemology. Proceedings of the 11th International Wittgenstein Symposium, Austria, H.P.T., 1987.

Penrose, R. (1988):
On the Physics and Mathematics of Thought. Oxford University Press.

Phelps, R.I. and Musgrove, P.B. (1986):
Knowledge Representation for Basic Visual Categories. Artificial Intelligence & its Applications, Cohn, A.G. & Thomas, J.R (Eds), 1986, Wiley.

Piaget, J. (1953,1977):
The Origin of Intelligence in the Child. Penguin.

Pitrat, J. (1976):
A Program for Learning to Play Chess. Pattern Recognition and Artificial Intelligence, Chen, C.H. (Ed.), 1976, Academic Press.

Popplestone, R.J. (1969):
FREDDY in Toyland. Machine Intelligence, Volume-4, Meltzer, B. and Michie, D. (Eds.), Edinburgh University Press.

Porter, B.W. and Kibler, D.F. (1985):
A Comparison of Analytic and Experimental Goal Regression for Machine Learning. Proceedings of the 9th IJCAI, 1985.

Postman, L. and Keppel, G. (Eds.) (1969):
Verbal Learning and Memory. Penguin.

Powers, W.T. (1978):
The Nature of Robots. BYTE, June 1978.

Prendergast, D., Slade, B. and Winkless, N. (1984):
A General-Purpose Robot-Control Language. BYTE, January 1984.

Pribram, K.H. (Ed.) (1969):
Perception and Action. Penguin.

Raibert, M. (1977):
A Model for Motor Control and Learning. Proceedings of the 5th IJCAI.

Raphael, B. (1968):
Programming a Robot. Proceedings of the IFIP Congress, Edinburgh, Aug-68.

Raphael, B. (1970):
The Relevance of Robot Research to Artificial Intelligence. Proceedings of the 4th Systems Symposium at Case Western Reserve University, Banerji, R.B. and Mesarovic, M.D. (Eds.), 1970.

Raphael, B. (1976):
The Thinking Computer. W.H. Freeman.

Rapoport, A. (1955):
Applications of Information to a Theory of Visions. Bulletin of Mathematical Biophysics, Volume-17.

Rapoport, A. (1962):
Some Self-Organization Parameters in Three-Person Groups. Principles of Self-Organization, Von Forester, H. & Zopf, W. (Eds), 1962, Pergamon Press.

Rauch-Hinden, W.B. (1985):
Artificial Intelligence in Business, Science, and Industry. Volume-II: Applications, Prentice-Hall.

Razran, G. (1971):
Mind in Evolution: An East West Synthesis of Learned Behavior and Cognition. Houghton Mifflin.

Reggia, J.A. (1985):
Virtual Lateral Inhibition in Parallel Activation Models of Associative Memory. Proceedings of the 9th IJCAI, 1985.

Reichard, J. (1978):
Robots: Fact, Fiction and Prediction. Thames and Hudson.

Restle, F. (1976):
Learning: Animal Behavior and Human Cognition. McGraw-Hill.

Reuver, H.A. (1978):
Learning within the Context of General Systems Theory. Progress in Cybernetics & Systems Research, Volume-IV, Trappl, R. & Pask, G. (Eds).

Reuver, H.A. (1980):
The Theory of Learning by Gal'perin and the Theory of Learning Systems. Progress in Cybernetics and Systems Research, Volume-(VII), Pichler, F.R. and Hanika, F.de.P. (Eds.).

Rhodes, F.H.T. (1976):
The Evolution of Life. Penguin.

Rich, E. (1983):
Artificial Intelligence. McGraw-Hill.

Roberts, L.G. (1960):
Pattern Recognition With an Adaptive Network. IRE International Convention Record, Part 2, 1960.

Rolls, E.T. (1975):
The Brain and Reward. Pergamon.

Rose, J.S. (1978):
A Course on Group Theory. Cambridge University Press, 1978.

Rose, S. (1970):
The Chemistry of Life. Penguin.

Rose, S. (1976):
The Conscious Brain. Penguin.

Rosen, C.A. and Nilsson, N.J. (1967):
An Intelligent Automaton. IEEE International Convention Record, Vol-15.

Rosen, C.A. (1968):
Machines that Act Intelligently. Science Journal, October 1968.

Rosenberg, R.S. and Rowe, P.F. (1981):
Spatial Problems for a Simulated Robot. Proceedings of the 7th IJCAI.

Rosenblatt, F. (1958):
The Perceptron, a Probabilistic Model for Information Organization and Storage in the Brain. Psychological Review, Volume-65.

Rosenblatt, F. (1959):
Two Theorems of Statistical Separability in the Perceptron. Proceedings of the Symposium on Mechanization of Thought Processes (I).

Rosenblatt, F. (1960):
Perceptual Generalization Over Transformation groups. Self Organising Systems: Proceedings of an Interdisciplinary Conference, 1959, Yovits, M.C. and Cameron, S. (Eds.), Pergamon Press.

Rosenblatt, F. (1962):
Principles of Neurodynamics and the Theory of Brain Mechanisms. Spartan.

Rosenblatt, F. (1962):
A Comparison of Several Perceptron Models. Self-Organizing Systems, Yovits, M.C., Jacobi, G.T. and Golostein, G.D. (Eds.), 1962, Spartan Books.

Rosenblatt, F. (1962):
Strategic Approaches to the Study of Brain Models. Principles of Self-Organization, Von Forester, H. & Zopf, W. (Eds), 1962, Pergamon Press.

Rosenblueth, A., Wiener, N. and Biggelow, J. (1943):
Behaviour, Purpose and Teleology. Philosophy of Science, Volume-10, 1943.

Ross, T. (1933):
Machines That Think. Scientific American, April 1933.

Rushby, R.J., (1975):
CAESAR in Rome. Man-Machine Studies Report UC-DSE/7, University of Canterbury, Christchurch, New Zealand.

Russell, D.W. and Rees, S.J. (1975):
System Control - A Case Study of a Statistical Learning Automaton. Progress in Cybernetics and Systems Research, Volume-2.

Sacerdoti, E.D. (1973):
Planning in a Hierarchy of Abstraction Spaces. Proc. of the IJCAI-3, 1973.

Safford, E.L. (1978):
The Complete Handbook of Robotics. TAB Books.

Safford, E.L. (1982):
Handbook of Advanced Robotics. TAB Books.

Sampson, J.R. (1976):
Adaptive Information Processing - An Introductory Survey. Springer-Verlag.

Samuel, A.L. (1959):
Some Studies in Machine Learning Using the Game of Checkers. Computers and Thought, Feigenbaum, E.A. and Feldman, J. (Eds.), 1963, McGraw-Hill.

Samuel, A.L. (1960):
Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, Volume-3.

Sandewell, E. (1974):
Biological Software. Proceedings of the 6th IJCAI, 1979.

Savage, J.M. (1977):
Evolution. Rinehart and Winston.

Schaffer, J.D. and Grefenstette, J.J. (1985):
Multi-Objective Learning Via Genetic Algorithms. Proceedings of the 9th IJCAI, 1985.

Schank, C. (1982):
Looking at Learning. Proceedings of the European Conference on Artificial Intelligence, 1982.

Schank, R. and Hunter, L. (1985):
The Quest to Understand Thinking. BYTE, April 1985.

Schindler, M. (1984):
Software Technology: Artificial Intelligence. Electronic Design, August 1984.

Schutzenberger, M.P. (1954):
A Tentative Classification of Goal-Seeking Behaviours. Systems Thinking, Emery, F.E. (Ed.), Penguin, 1969.

Selfridge, O.G. (1955):
Pattern Recognition and Learning. Information Theory - Symposium on Information Theory held at the Royal Institute, Cherry, C. (Ed.).

Selfridge, O.G. (1956):
Pattern Recognition and Learning. Information Theory, Cherry, E.C. (Ed.), Butterworth, 1956.

Selfridge, O.G. (1959):
PANDEMONIUM: A Paradigm for Learning. Proceedings of the Symposium on Mechanization of Thought Processes (I).

Selfridge, O.G. and Neisser, U. (1960):
Pattern Recognition by Machine. Computers and Thought, Feigenbaum, E.A. and Feldman, J. (Eds.), 1963, McGraw-Hill.

Selfridge (1962):
The Organization of Organization. Self-Organizing Systems, Yovits, M.C., Jacobi, G.T. and Golostein, G.D. (Eds.), 1962, Spartan Books.

Shannon, C.E. and Weaver, W (1949):
The Mathematical Theory of Communication. Illinois Press.

Shannon, C.E. (1951):
Presentation of a Maze-Solving Machine. Transactions of 8th Cybernetics Conference, Josiah Macy Junior Foundation.

Shannon, C.E. (1953):
Computers and Automata. Proceedings of the IRE, Volume-41, 1953.

Shannon, C.E. (1956):
A Universal Turing Machine with Two Internal States. Automata Studies, Shannon, C.E. and McCarthy, J. (Eds.), Princeton University Press.

Shannon, C.E. and McCarthy, J. (Eds.) (1956):
Automata Studies. Princeton University Press.

Shaw, D.E. (1985):
NON-VON's Applicability to Three AI Task Areas. Proceedings of the 9th IJCAI, 1985.

Shih, L.Y. (1982):
Stability of Automatic Guidance for a Mobile Robot. Proceedings of the National Conference on A.I., 1982.

Silklossy, L. and Dreussi, J. (1973):
An Efficient Robot Planner Which Generates its Own Procedures. Proceedings of the 3rd IJCAI, 1973.

Simon, H.A. and Siklossy, L. (1972):
Representation and Meaning: Experiments With Information Processing Systems. Prentice-Hall.

Simon, H.A. (1956):
Rational Choice and the Structure of the Environment. Systems Thinking, Emery, F.E. (Ed.), Penguin, 1969.

Simon, H.A. (1983):
Why Should Machines Learn. Machine Learning: The Artificial Intelligence Approach, Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. (Eds.), Tioga Publishing, 1983.

Simons, G. (1983):
Are Computers Alive? Harvester Press.

Singer, J.R. (1961):
Electronic Analogue of the Human Recognition System. Journal of the Optical Society of America, Volume-51.

Skinner, B.F. (1961):
Cumulative Record: Enlarged Edition. Appleton-Century-Crofts.

Sklansky, J. (1964):
Adaptation, Learning, Self-Repair, and Feedback. IEEE Spectrum, May 1964.

Slagle, J. (1971):
Artificial Intelligence: The Heuristic Programming Approach. McGraw-Hill.

Sloman, A. and Croucher, M. (1981):
Why Robots have Emotions. Proceedings of the 7th IJCAI, 1981.

Smith, M.H. (1973):
Design of a Low Cost, General Purpose Robot. Proceedings of the 3rd IJCAI, 1973.

Smith, W.A. (1974):
Information Theory and Cybernetics. Journal of Cybernetics, Volume-4.

Smith, J.M. (1975):
The Theory of Evolution. Penguin.

Soloway, E.M. and Riseman, E.M. (1977):
Levels of Pattern Description in Learning. Proceedings of the 5th IJCAI.

Sommerhoff, G. (1969):
The Abstract Characteristics of Living Systems. Systems Thinking, Emery, F.E. (Ed.), Penguin, 1969.

Sommerhoff, G. (1974):
Logic of the Living Brain. Wiley.

Stanfield, D.E. (1979):
My Computer Runs Mazes. BYTE, June 1979.

Staudigal, P. (1980):
Applied Model of Learning. Progress in Cybernetics and Systems Research, Volume-(VII), Pichler, F.R. and Hanika, F.de.P. (Eds.).

Steele, E.J. (Ed.) (1960):
Bionic Symposium (1st): Living Prototypes - The Key to New Technology.

Steinbach, K. and Piske, U.A.W. (1963):
Learning Matrices and Their Applications. IEEE Transactions on Electronic Computers, December 1963.

Stepankova, O. and Havel, I.M. (1976):
A Logical Theory of Robot Problem Solving. Artificial Intelligence, Volume-7.

Stevens, J.K. (1985):
Reverse Engineering the Brain. BYTE, April 1985.

Stevens, L. (1985):
Artificial Intelligence - The Search for a Perfect Machine. Hayden.

Stewart, D.J. (1959):
A Notation for Logical Nets. Cybernetica, Volume-2, 1959.

Stewart, D.J. (1967):
Logical Nets and Organisms That Learn. Automaton Theory and Learning Systems, Stewart, D.J. (Ed.), 1967, Academic Press.

Stewart, D.J. (Ed.) (1967):
Automaton Theory and Learning Systems. Academic Press.

Sugeno, M. and Terano, T. (1978):
A Model of Learning Based on Fuzzy Information. Progress in Cybernetics and Systems Research-III, Trappl, R. and Klir, G.J. (Eds.).

Sutherland, N.S. (1968):
Machines Like Man. Science Journal, October 1968.

Takase, K. (1979):
Skill of Intelligent Robot. Proceedings of the 6th IJCAI, 1979.

Taylor, W.K. (1968):
Machines that Learn. Science Journal, October 1968.

Teyler, T.J. (Ed.) (1978):
Brain and Learning. D.Reidel Publishing.

Thathachar, M.A.L. and Oommen, B.J. (1983):
Learning Automata Processing Ergodicity of the Mean: The Two-Action Case. IEEE Transactions on Systems, Man, and Cybernetics, Vol:SMC-13.

Thathachar, M.A.L. and Ramakrishnan (1984):
A Cooperative Game of a Pair of Learning Automata. Automatica, Vol:20.

Thompson, A.M. (1977):
The Navigation System of the JPL Robot. Proceedings of the 5th IJCAI.

Thorpe, C.E. (1984):
Path Relaxation: Path Planning for a Mobile Robot. Proceedings of the AAAI-1984.

Thorpe, W.H. (1963,1956):
Learning and Instinct in Animals. Methuen.

Thorpe, W.H. (1978):
Purpose in a World of Chance: A Biologist's View. Oxford University Press.

Thring, N.W. (1984):
Robots and Telechirs. Wiley.

Toda, M. (1962):
The Design for a Fungus-Eater: A Model of Human Behavior in an Unsophisticated Environment. Behavioural Science, Volume-7.

Tolman, E.C. (1959):
Principles of Purposive Behavior. Psychology: A Study of a Science, Koch, S. (Ed.), Volume-2, McGraw-Hill, 1959.

Torrance, S.B. (Ed.) (1984):
The Mind and the Machine: Philosophical Aspects of Artificial Intelligence. Ellis Horwood.

Touretzky, D.S. and Hinton, G.E. (1985):
Symbols Among the Neurons: Details of a Connectionist Inference Architecture. Proceedings of the 9th IJCAI, 1985.

Trappl, R. (Ed.) (1983):
Cybernetics: Theory and Applications. Hemisphere Publishing.

Trappl, R. (Ed.) (1985):
Impacts of Artificial Intelligence. Elsevier Science.

Tsypkin, Ya.Z. (1973):
Foundations of the Theory of Learning Systems. Academic Press.

Turing, A.M. (1947):
Turing and the Origins of the Computer. New Scientist, February 1980.

Turing, A.M. (1950):
Computing machinery and Intelligence. Computers and Thought, Feigenbaum, E.A. and Feldman, J. (Eds.), 1963, McGraw-Hill.

Uhr,L. and Vossler, C. (1963):
A Pattern Recognition Program that Generates, Evaluates, and Adjusts its Own Operators. Computers and Thought, Feigenbaum, E.A. and Feldman, J. (Eds.), 1963, McGraw-Hill.

Uhr, L, (Ed.) (1966):
Pattern Recognition. Wiley.

Uhr, L. and Kochen, M. (1969):
MIKROKOSMOS and Robotics. Proceedings of the 1st IJCAI, 1969.

Uhr, L. (1973):
Pattern Recognition, Learning and Thought. Prentice-Hall.

Uragami, M., Mizumoto, M. and Tanaka, K. (1976):
Fuzzy Robot Controls. Journal of Cybernetics, Volume-6.

Uttly, A.M. (1956):
Temporal and Spatial Patterns in a Conditional Probability Machine. Automata Studies, Shannon, C.E. and McCarthy, J. (Eds.), Princeton University Press.

Uttly, A.M. (1956):
Conditional Probability Machines and Conditioned Reflexes. Automata Studies, Shannon, C.E. and McCarthy, J. (Eds.), Princeton University Press.

Veelenturf, L.P.J (1981):
An Automata-Theoretical Approach to Developing Learning Neural Networks. Cybernetics and Systems: An International Journal, Vol:12.

Vigilone, S.S. (1970):
Applications of Pattern Recognition Technology. Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications, Mendel, J.M. and Fu, K.S. (Eds.), 1970, Academic Press.

Von Forester, H. (1959):
On Self-Organizing Systems and Their Environments. Self Organising Systems: Proceedings of an Interdisciplinary Conference, 1959, Yovits, M.C. and Cameron, S. (Eds.), Pergamon Press.

Von Forester, H. and Zopf, W. (Eds.) (1962):
Principles of Self-Organization. Pergamon Press.

Von Forester, H., White, J.D., Peterson, L.J. & Russel, J.K. (Eds.) (1968):
Purposive Systems: Proceedings of the First Annual Symposium of the American Society of Cybernetics. Spartan Books.

von-Neumann, J. and Morgenstern, O. (1947,1953):
Theory of Games and Economic Behaviour. Princeton University Press.

von-Neumann, J. (1951):
The General and Logical Theory of Automata. Cerebral Mechanisms in Behavior: The Hixon Symposium, Jeffress, L.A. (Ed.), Chapman and Hall.

von-Neumann, J. (1956):
Probabilistic Logic and the Synthesis of Reliable Organisms from Unreliable Components. Automata Studies, Shannon, C.E. and McCarthy, J. (Eds.), 1956, Princeton University Press.

Walker, J.J. and Young, J.F. (1977):
A Learning Machine Which Learns, Remembers and Reproduces Sequences without the Use of Delays. Kybernetes, Volume-6.

Wallace, R.A. (1952):
The Maze-Solving Computer. Proceedings of the Association for Computing Machinery (ACM), Rimbach.

Walter, W.G. (1950):
An Imitation of Life. Scientific American, Volume-182, 1950.

Walter, W.G. (1951):
A Machine That Learns. Scientific American, Volume-185, 1951.

Walter, W.G. (1953,1961):
The Living Brain. Penguin.

Waterman, D.A. and Hayes-Roth, F. (Eds.) (1978):
Pattern-Directed Inference Systems. Academic Press.

Watson, W.C. (1981):
Physiological Psychology: An Introduction. Houghton Mifflin.

Webster, J. (1978):
Robot Simulation on Microcomputers. BYTE, April 1978.

Wee, W.G. and Fu, K.S. (1969):
A Formulation of Fuzzy Automata and its Application as a Model of Learning Systems. IEEE Transactions on Systems Science and Cybernetics, Volume-SSC-5.

Weinstein, M. (1975):
Structured Robotics. Proceedings of the 4th IJCAI, 1975.

Wendt, S. (1968):
A Trainable Classifier with Piecewise Linear Separation Using the Learning Matrix. Proceedings of the IFIP Congress, Edinburgh, August 1968.

Widrow, B. (1962):
Generalization and Information Storage in Networks of ADALINE "Neurons". Self-Organizing Systems, Yovits, M.C., Jacobi, G.T. and Golostein, G.D. (Eds.), 1962, Spartan Books.

Widrow, B., Gupta, N.K. and S. Maitra (1973):
Punish/Reward: Learning with a Critic in Adaptive Threshold Systems. IEEE Transactions on Systems, Man, and Cybernetics, Volume-SMC-3.

Wiener, N. (1948):
Cybernetics. MIT Press.

Wilson, M.J.D. and Aleksander (1976):
Arraylike Learning Systems. Journal of Cybernetics, Volume-6.

Wilson, M.J.D. (1978):
Artificial Perception in Adaptive Arrays. Technical Memorandum, N/R/065, Brunel University, Department of Electrical Engineering.

Winograd, T. (1973):
Understanding Natural Language. Edinburgh University Press.

Winograd, T., Davis, R., Dreyfus, S., and Smith, B. (1985):
Expert Systems: How Far Can They Go? Proceedings of the 9th IJCAI.

Winston, P.H. (1983):
Learning Physical Descriptions from Functional Definitions, Examples, and Precedents. Proceedings of the AAAI-1983.

Winston, P.H. (1977,1984):
Artificial Intelligence. Addison-Wesley.

Witkowski, C.M. (1983):
A Parallel Processor Algorithms for Robot Route Planning. Proceedings of the 8th IJCAI, 1983.

Witten, I.H. (1977):
In Search of PURR-PUSS: A Visitor's Experience. Man-Machine Studies Report UC-DSE/11, University of Canterbury, Christchurch, New Zealand.

Witten, I.H. and Gaines, B.R. (1978):
The Role of Randomness in System Theory. Progress in Cybernetics and Systems Research, Volume-(III), Trappl, R., Klir, G.J. and Ricciardi, L. (Eds.), 1978.

Wooldridge, D.E. (1963):
The Machinery of the Brain. McGraw-Hill.

Yazdani, M. and Narayanan, A. (Eds.) (1984):
Artificial Intelligence - Human Effects. Ellis Horwood.

Young, J.F. (1973):
Cybernetic Engineering. Butterworths.

Young, J.Z. and Margerison, T. (Eds.) (1969):
From Molecule to Man. Thames and Hudson.

Yovits, M.C. and Cameron, S. (Eds.) (1960):
Self Organising Systems: Proceedings of an Interdisciplinary Conference, 1959. Pergamon Press.

Yovits, M.C., Jacobi, G.T. and Golostein, G.D. (Eds.) (1962):
Self-Organizing Systems. Spartan Books.

Zadeh, L.A. (1973):
Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Transactions on Systems, Man, and Cybernetics, Volume-SMC-3.

Zadeh, L.A. (1973):
The Concept of a Linguistic Variable and its Application. Learning Systems and Intelligent Robots, Fu, K.S. and Tou, J.T. (Eds.), 1973, Plenum Press.