

**CSIR GOLDEN JUBILEE
SERIES**

ARTIFICIAL INTELLIGENCE



K. D. PAVATE

ARTIFICIAL INTELLIGENCE

K.D. PAVATE



Publications & Information Directorate

Dr. K.S. Krishnan Marg

New Delhi 110 012

India

Artificial Intelligence

K.D. Pavate

©Publications & Information Directorate

First Edition : May 1992

Second Edition : March 1993

Third Edition : January 1996

ISBN : 81-7236-035-5

CSIR Golden Jubilee Series

Publication No. 7

Series Editor : Dr. Bal Phondke

Volume Editor : Parvinder S. Chawla

Cover Design : Pradip Banerjee

Illustrations : Pradip Banerjee, Neeru Sharma, Neeru Vijan,
Sushila Vohra, K.K. Bhatnagar and Anil Kumar

Production : V. Ramachandran, K.B. Nagpal, Vinod Sharma,
Sudhir Chandra Mangain and Radhe Shiam

Designed, Printed and Published by
Publications & Information Directorate (CSIR)
Dr. K.S. Krishnan Marg, New Delhi 110 012
India

Foreword

The Council of Scientific & Industrial Research (CSIR), established in 1942, is committed to the advancement of scientific knowledge, and economic and industrial development of the country. Over the years CSIR has created a base for scientific capability and excellence spanning a wide spectrum of areas enabling it to carry out research and development as well as provide national standards, testing and certification facilities. It has also been training researchers, popularizing science and helping in the inculcation of scientific temper in the country.

The CSIR today is a well knit and action oriented network of 41 laboratories spread throughout the country with activities ranging from molecular biology to mining, medicinal plants to mechanical engineering, mathematical modelling to metrology, chemicals to coal and so on.

While discharging its mandate, CSIR has not lost sight of the necessity to remain at the cutting edge of science in order to be in a position to acquire and generate expertise in frontier areas of technology. CSIR's contributions to high-tech and emerging areas of science and technology are recognised among others for precocious flowering of tissue cultured bamboo, DNA finger-printing, development of non-noble metal zeolite catalysts, mining of polymetallic nodules from the Indian Ocean bed, building an all-composite light research aircraft, high temperature superconductivity, to mention only a few.

Being acutely aware that the pace of scientific and technological development cannot be maintained without a steady influx of bright young scientists, CSIR has undertaken a vigorous programme of human resource development which includes, inter alia, collaborative efforts with the University Grants Commission aimed at nurturing the budding careers of fresh science and technology graduates.

However, all these would not yield the desired results in the absence of an atmosphere appreciative of advances in science

and technology. If the people at large remain in awe of science and consider it as something which is far removed from their realms, scientific culture cannot take root.

CSIR has been alive to this problem and has been active in taking science to the people, particularly through the print medium. It has an active programme aimed at popularization of science, its concepts, achievements and utility, by bringing it to the doorsteps of the masses through both print and electronic media. This is expected to serve a dual purpose. First, it would create awareness and interest among the intelligent layman and, secondly, it would help youngsters at the point of choosing an academic career in getting a broad-based knowledge about science in general and its frontier areas in particular. Such familiarity would not only kindle in them deep and abiding interest in matters scientific but would also be instrumental in helping them to choose the scientific or technological education that is best suited to them according to their own interests and aptitudes. There would be no groping in the dark for them. However, this is one field where enough is never enough.

This was the driving consideration when it was decided to bring out in this 50th anniversary year of CSIR a series of profusely illustrated and specially written popular monographs on a judicious mix of scientific and technological subjects varying from the outer space to the inner space. Some of the important subjects covered are astronomy, meteorology, oceanography, new materials, immunology and biotechnology.

It is hoped that this series of monographs would be able to whet the varied appetites of a wide cross-section of the target readership and spur them on to gathering further knowledge on the subjects of their choice and liking. An exciting sojourn through the wonderland of science, we hope, awaits the reader. We can only wish him Bon voyage and say, happy hunting.

Preface

Computers are here to stay. They have had a big impact on our everyday lives. With suitable software and programs they are being used to solve a wide spectrum of problems. For instance, they are being used in solving complex mathematical problems, preparing database, forecasting economic trends, assisting management in taking decisions and in helping engineers to operate large factories. So, it is not surprising when a casual onlooker says: "This computer is a smart machine. It is very intelligent!" Now, adjectives like 'smart' and 'intelligent' are invariably associated with human beings. A person who does all the work assigned to him in the shortest possible time is considered smart. But, intelligence is a quality which is much easier to recognize than to define. One runs into serious problems while attempting to measure or assign values to it.

Over the past four decades scientists have attempted to apply available technologies to design user-friendly computer systems which exhibit a semblance to human intelligence. So far they have been only partially successful. Expert systems are available which assist their users in finding solutions within certain well-defined areas of knowledge. Today, machines can recognize spoken words and simple sentences. There are those which analyze pictures and recognize specific patterns. The new subject of Artificial Intelligence (AI) is progressing rapidly. It fascinates people because it touches on subjects which are of intimate concern to human beings - choice, learning, cognition, etc. This book has tried to capture the flavour of AI-linked concepts and developments. Although written for the benefit of the younger generation, it is hoped that this book will also create an interest among our budding scientists and engineers.

Acknowledgments

Dr G.P.Phondke set the ball rolling by first suggesting that I write something for the CSIR Golden Jubilee series. Even as I was murmuring, he quickly followed it up by suggesting the theme. It was only when he confronted me with a deadline which had to be strictly adhered to that I woke up to the realization that I had made an irrevocable commitment. I must confess that I would have missed the excitement of writing had it not been for Dr Phondke.

The project brought me in close contact with the staff of PID. It has been simply great working with Mrs. Parvinder Chawla, the volume editor, and Shri S.K.Nag. Parvinder has spared no pains in processing the manuscript and in making valuable suggestions. The Art Section of PID, especially Shri Pradip Banerjee, also contributed substantially by providing illustrations to break the monotony of the text. My thanks to all of them.

I also owe gratitude to my colleagues and friends at the CEERI Centre in Delhi. They too have followed the progress of this book with considerable interest. They have, in fact, provided me with facilities to complete this writing assignment. I am grateful to Dr K.V.Ramakrishnan and to Dr. S.S.Agarwal for many useful discussions on the subject. Shri Sadashiv Sharma helped me with the sonographs and Shri D.J.Ray provided me with photographs. Shri S.N.Gupta assisted me with both the computer and the WordPerfect program. I must thank them for their cooperation.

Finally, I must thank my family for their whole-hearted support, particularly, my daughter, Dr Rita, who discussed with me on various aspects of the physiology of central nervous system and the psychology of learning. I am also grateful to her for reading the manuscript through the various stages of its evolution.

K. D. Pavate

Dedicated to my wife

Suvena

for all her patience

Contents

The Genesis	... 1
Learning and Memory	... 16
Scope and Extent	... 34
A Problem Solver	... 41
Programming Intelligence	... 61
User-Friendly Systems	... 79
An Overview	... 92
Glossary	... 97

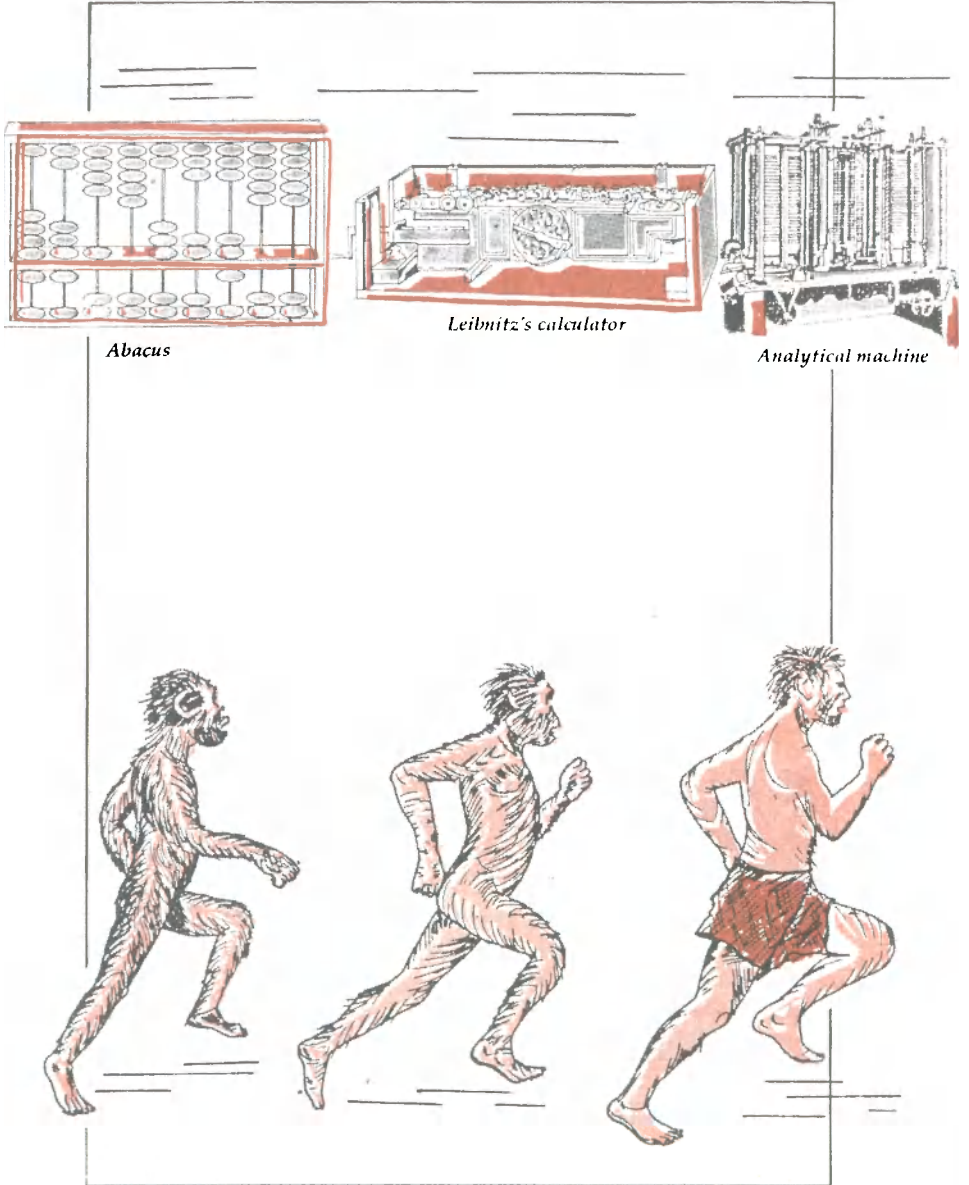
The Genesis



Since the dawn of civilization man has been interested in devices which could assist him in simple arithmetical exercises. Amongst the earliest known contraptions are the abacus and the soroban. These use beads which are mounted on stiff wires and can be moved up and down on them. They are used to count numbers. Surprisingly, they have survived through the ages and are being used even today in shops in USSR, China, Singapore and Hongkong.

The next known advancements were the table top counting machine invented by Blaise PASCAL (1623-1662) in 1642 and a calculating machine by the eminent German mathematician-philosopher Gottfried LEIBNITZ (1646-1716) in 1694. Whereas Pascal's version could only add and subtract, the latter machine was capable of even multiplying and dividing numbers by the process of repeated additions or subtractions. Both these machines operated on the basis of toothed gear wheels meshing into one another.

In 1801 a Frenchman named Joseph JACQUARD (1752-1834), introduced the concept of a "memory" in the conventional



Abacus

Leibnitz's calculator

Analytical machine

Evolution of computers in



Desktop personal computer

Supercomputer



parallel to human evolution

weaving loom. Earlier, the workers had to memorize the various designs or weaving patterns. The same information was now permanently stored on a series of stiff, large-sized cards with punched holes. These holes enabled the shuttles with different coloured threads to operate in a sequence and generate specific weaving patterns. By introducing a different set of cards, new patterns could be woven on the cloth. Jacquard's cards were adopted almost overnight by the entire weaving industry in France.

A hundred years ago, Herman HOLLERITH (1869-1926) in U.S.A extended Jacquard's principle of storing information on punched cards to another important application. In 1891 he redesigned the cards so that holes now corresponded to statistical data concerning individuals. He also designed and constructed machines to punch holes and to read the data off these cards. These machines could also perform the required additions so as to generate useful statistical information.

Hollerith's machines enabled the entire results of the 1891 census to be available in less than two years. Had manual methods been used, the results would have been ready only in 1901! Hollerith's machines were there to stay for a very long time. In fact, the company founded by him eventually transformed itself into the International Business Machines (IBM), which is one of the leaders in the field of computers even today.

Charles BABBAGE (1791-1871), an eminent English mathematician commenced on a venture in the 1830s to solve simple mathematical polynomial equations using mechanical machines. His first venture was to design a "difference engine". Later, he also designed an "analytical machine" which could solve any complex mathematical problem. The innovation in his design was the incorporation of a "memory". He called this a "store" and used it to hold the answers of various subcalculations till they were needed at a later stage. However, Babbage was not entirely successful with both his ventures as he lacked a precision workshop

with fine machining facilities. Yet, Babbage is credited with having thought of all the basic components of a modern computer.

By the 1930s, technology had advanced and Babbage's ideas and concepts were revived. In 1937, Howard AIKEN (1900-1973) at the Harvard University designed the first electromechanical computer. It used punched cards to feed in information and simple electric **relays** to store data. Aiken's machine (Mark 1) was used to generate tables for use by the artillery gunners of the army and also for solving equations for those who designed the first atomic bomb. Almost simultaneously, John MAUCHLY and Presper ECKERT at the Pennsylvania University commenced work on a similar machine but used **thermionic valves** rather than relays. Since it was an all electronic calculating system it turned out to be much faster than the Mark I. Nevertheless, both these machines had some inherent limitations. The most important of these being that in order to change a program to solve another set of mathematical equations, considerable rewiring had to be resorted to. The circuits had to be changed and this resulted in delays. The other shortcomings were their large size, a huge maintenance staff and an enormous cooling mechanism.

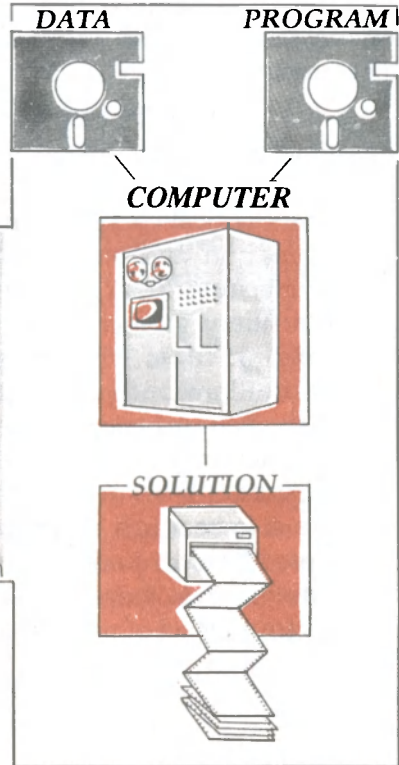
Immediately after World War II, a large number of engineers and scientists who had been working on these machines found themselves free to exchange ideas with other colleagues and to apply their new found expertise to solve problems in other fields. Two very important events took place during the late 1940s and these together enabled the original ideas of Babbage to be engineered into some very important products.

The first was the introduction of a new concept by John von NEUMANN (1903-1957). He suggested that all the detailed steps which the machine has to follow in performing its calculations, should also be coded and entered into the machine along with the data. Interestingly, at about the same

time, an English mathematician named Alan TURING (1912-1954) also suggested a similar idea. This was a major breakthrough since the con-



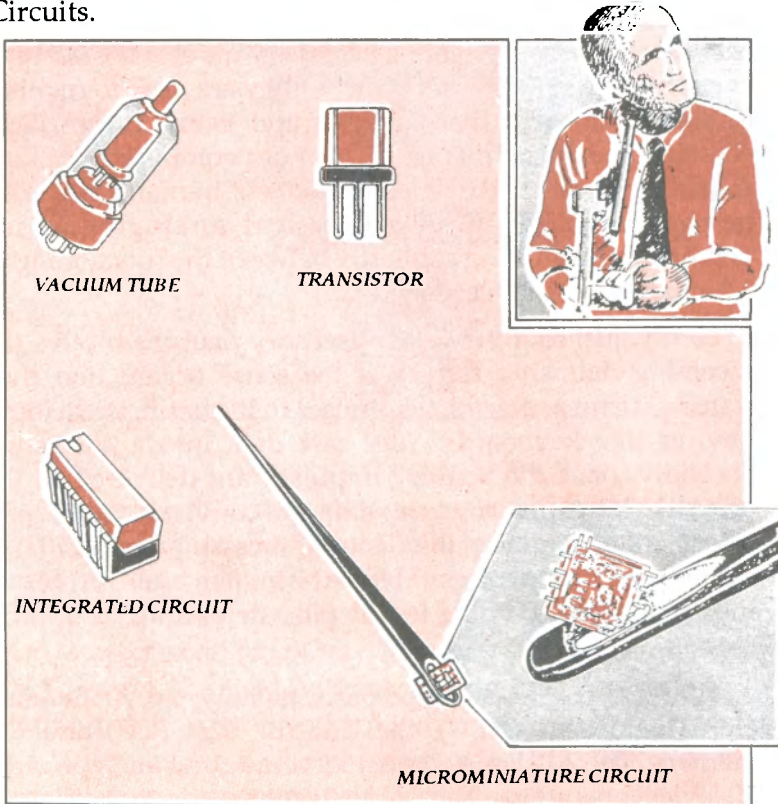
cept of a stored **program** gave such machines considerable flexibility. One no longer had to do the rewiring of machines in order to solve a new set of mathematical problems. All one did now was to write out a new program or in other words, prepare a new **software** and feed it into the machine along with the data.



von Neumann gave the concept of a stored program

Another useful suggestion von Neumann proposed was that all arithmetic be performed using the **binary** notation instead of the conventional units of ten. The ones and zeros used can be easily represented by the **electronic toggles**. The first computers to be built using electronic toggles rather than the electromagnetic relays were the EDVAC in U.S.A and the EDSAC in England and these were the forerunners of the modern computers.

The second advancement was a technological one. With the invention of the semiconductor transistor by William SHOCKLEY and his colleagues in 1948 it was only a matter of time before the vacuum tubes were entirely replaced by semiconductor devices. The foremost advantage of semiconductor devices is their very low power consumption. A revolutionary advance in semiconductor technology was ushered in a decade later when KILBY of Texas Instruments showed how an entire electronic circuit could be fabricated on the surface of a silicon **chip**. These are known as Integrated Circuits.



Shrinking computer components; William Shockley (inset)

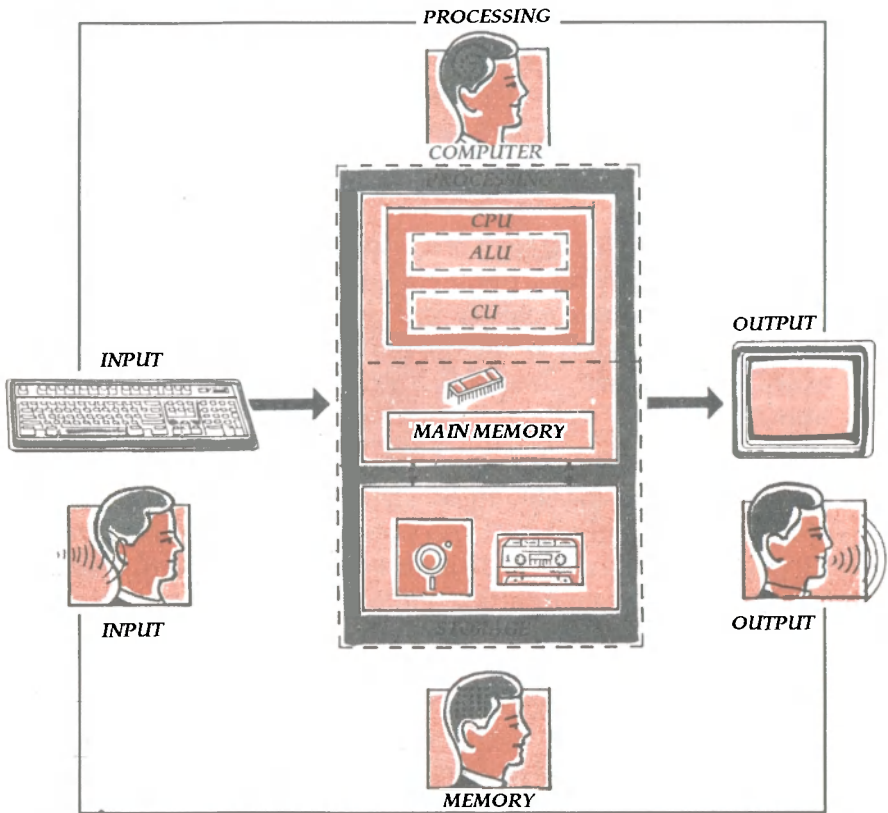
More recently, Very Large Scale Integrated circuits(VLSI) with literally thousands of semiconductor devices are being fabricated on single chips. The complexity of these VLSI circuits has grown year by year. The manufacturing technologies of such functional circuits have helped in reducing the costs of such chips. The reliability of semiconductor chips is also very high. Reasonably priced computers based on such VLSI circuits are now available.

Computer and Brain

Computers are being used today to assist man in several activities, such as navigation across oceans and the sky, and even through space. What is more, they are used to regulate human heart beat (pacemakers) and even control large chemical industrial plants. No wonder computers are considered to be the machine equivalents of human brain. This is supported by three anatomical analogies which demonstrate a striking similarity between the functioning of a human brain and a computer.

Firstly, just as a network of sensory neurons receive the incoming data that arrives at the sense organs and then transmit them as electrical impulses to the brain, so do input devices like keyboards enter raw data into a computer. Secondly, once the sensory impulses are delivered to the brain, the latter processes this data and converts it into useful information. Similarly, the Central Processing Unit (CPU) of a computer, that is comparable to the human brain, processes the input data according to a step by step sequence of commands.

The CPU consists of three parts, namely, the Arithmetic-logic Unit (ALU), the Control Unit (CU) and the memory. The ALU does the actual computing and performs all logical operations. The CU analyzes the command signals and sends instructions to various parts of the computer. Also, it regulates the overall flow of information. Computer



Parts of a computer system

memory comprises data storage systems similar to human memory which can store a wide range of information that could be retrieved as and when required. A computer has a Random Access Memory (RAM) where data can be read/stored and Read-Only Memory (ROM) from where data can only be read without writing new data into it. In addition, there are two other kinds of specialized memories in a computer. One, the programmable ROM(PROM) that records data only once and second, the erasable PROM (EPROM) which has an additional facility that helps the erasing of previous data and entry of new ones.

Finally, in the brain, the processed information is brought out through the motor nerves in the form of motor impulses and is further carried to the muscles and various other body organs. This function in a computer is performed by the output devices such as the Visual Display Unit (VDU), printer, etc.

The working mechanism of a computer is thus akin to that of the brain. However, a pertinent query that springs up from this brain-computer analogy is: Can a man-made machine having no biological element entertain conscious thoughts exactly as humans? Or simply, can computers simulate the human mind?

Despite major breakthroughs in computer technology, it has yet not been feasible to construct conscious machines - systems that could mimic brain power. The computers have till now only been used to do arithmetical calculations. There is no doubt that they are efficient slaves and do whatever they are instructed to do any number of times and without complaining. Nevertheless, the abilities of a computer are not at par with the ingenuity of the human brain. This has today catalyzed research for the creation of machines that would possess the intellectual abilities of human beings.

Thinking Machines

Many people have tried to compare the abilities of humans with those of computers. Suppose you were in charge of a large chemical manufacturing plant and it was your duty to ensure that this factory operates at its optimum capacity with minimum wastage of materials and human resources. You could consult a number of experts who have operated similar plants with success and use their combined expertise to write a computer program in order to operate the factory. This program would essentially imitate the thought process of a 'human expert' and manage the operation of the entire factory. Solutions to crisis situations would be anticipated and



Manager of a factory gets inputs from various departments to help him take decisions

included in the program. Most people would agree that this particular program demonstrates some (artificial) intelligence in its task of operating the factory. This is a situation where a computer simulates intelligent human behaviour as

far as this particular assignment of operating the factory is concerned. The same program would probably not be usable in operating a shipyard! Some researchers have proposed that the ultimate goal of artificial intelligence (AI) is to construct a machine which could actually think like human beings in all aspects. Such machines, known as Ultra Intelligent Machines (UIM) do not exist yet, but who knows, they may be feasible in the 21st century !

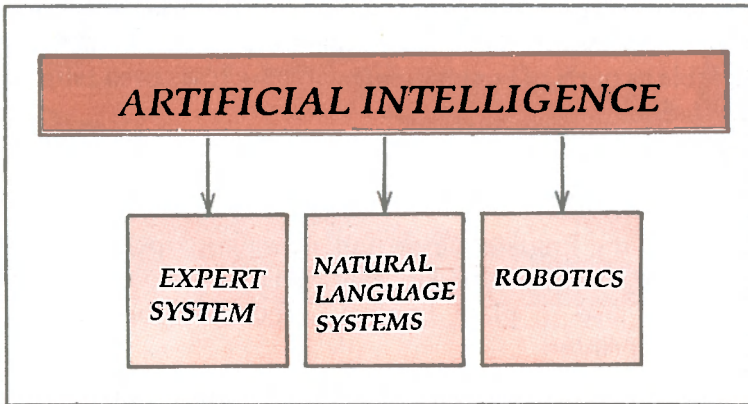
Human beings tend to think symbolically rather than numerically. Our intelligence seems to be based on our mental ability to manipulate symbols rather than numbers. On the other hand, computers were originally designed to process numbers. An algorithm is a step by step procedure with well-defined starting and ending points used in solving problems. The conventional computer architecture readily lends itself to this step by step approach. Human reasoning processes tend to be non-algorithmic i.e. our mental activities consist of something much more than merely following step by step procedures. Most of AI research has been devoted to symbolic non-algorithmic processing techniques with the help of computers.

In addition to symbolic processing AI researchers rely on heuristics — assumptions based on past experience — to solve problems. By using heuristics one does not have to completely rethink every aspect of a problem one is faced with. If a handy rule of thumb based on earlier experiences exists, then it is applied to the problem (or to a particular part of a problem).

Another aspect of AI is the use of pattern matching procedure to discover relationships between activities just as humans do. Here one tries to describe objects, events or processes in terms of their features and relationships. People tend almost instinctively to discover relationships between things. They sense qualities and spot patterns that explain how various items relate to each other. If computers are to become more intelligent they must be able to make similar

associations between objects, events and processes which people do in a natural manner.

AI continues to experience rapid changes in both concept and scope. Three major areas of AI are expert systems, natural language processing and robotics.



Branches of Artificial Intelligence

Expert systems: These are computer programs designed to emulate the reasoning process of an experienced professional in a particular area of expertise. These are also known as knowledge based systems. These systems are designed to assist experts and not necessarily to substitute them. Expert systems have proved extremely useful in such diverse fields as medical diagnosis, chemical analysis and geophysical resource exploration. Further applications of expert systems are likely to become available in other fields as well. Complex instruments would soon become user-friendly as they would contain programs to anticipate the difficulties a user might face in using the machine.

Natural Language Processing : At present, the use of computers is limited by communication difficulties. An effective use of computers will be possible if people could communicate with them through the use of natural languages. The

field of natural language processing is divided into two parts. Firstly, computers are trained to understand a natural language such as ordinary English. This will enable the users to communicate with the machine in a language with which they are already familiar. Secondly, the machines can be trained to produce outputs which are in English. This will enable the user to understand what the computer has to say very clearly. Some natural language interfaces are already available as parts of business softwares such as spreadsheets, database, etc.

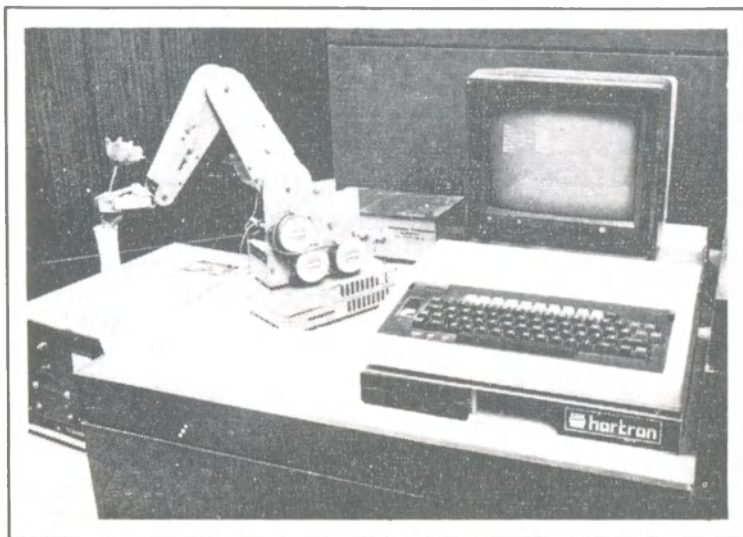
Natural language processing consists of the following aspects:

- *Processing English words and sentences typed onto the computer through the keyboard. The response of the machines, also in plain English, appear on the VDU or through a printer.

- *The machine can recognize words and sentences as spoken by the user. Speech synthesizers can be used for the machine to reply or to warn the user about a mistake he is about to make.

- *Another aspect is computer vision where the machine understands its environment with the help of tactile sensors or TV cameras.

Robotics: A robot is an electromechanical system which can be programmed to perform manual tasks. It can move materials, tools, perform a variety of tasks and all these can be programmed. There are two types of robots, viz. the fixed place, industrial assembly robots similar to those used in assembly of cars. The others are autonomous robots which operate by themselves in the real world. Software is the intelligence behind these devices. An intelligent robot includes sensory devices such as tactile sensors and TV cameras. It can be programmed to perform different functions. The system reprograms itself based on the signals received from its sensors.



A robot

Closely linked with AI is the process of learning. Learning as we all know, is a means of acquiring new knowledge. There is already much research underway to create such advanced programs which would enable machines to learn just as humans do so naturally. Humans learn by the process of memorizing facts and sequences of actions (to cycle, ride horses, swim, etc.). Each item one learns is a piece of information. Sequences are known as procedures. So when one programs a computer one feeds in both factual and procedural information. Another important way people learn is by the cognitive procedure. Here a human being analyzes, organizes and correlates specific pieces of knowledge in his mind. He subsequently learns to generalize after examining a few specific examples. If systems can learn just as humans do, it would be a very powerful tool as far as AI is concerned.



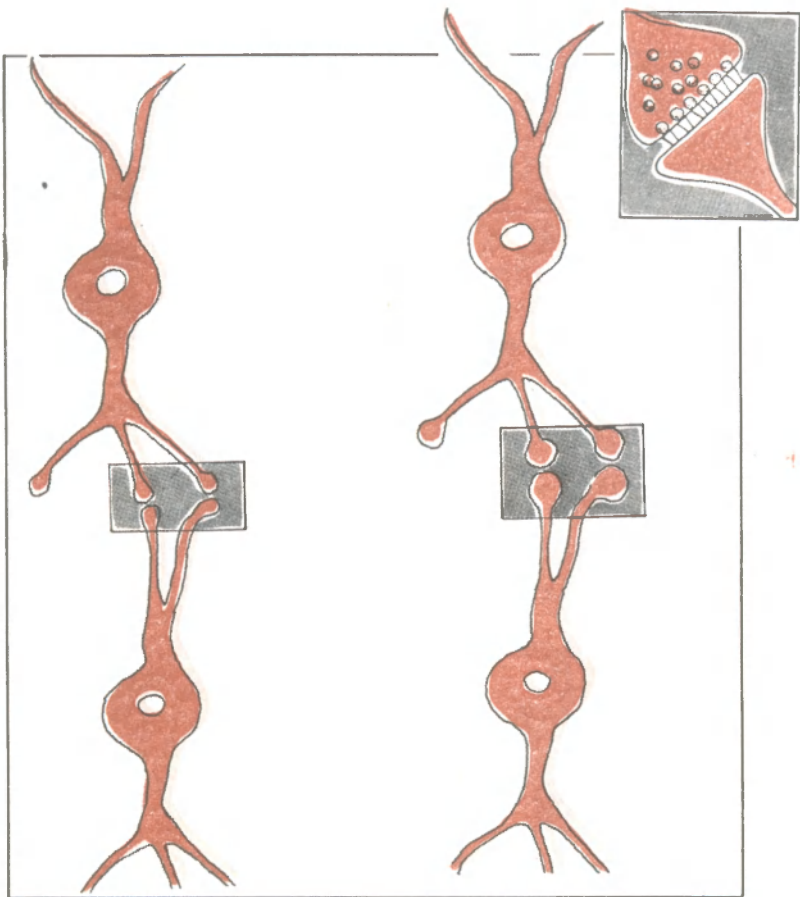
Learning and Memory



A human being is a remarkable animal as he is the only one to be endowed with such a large brain. One of the major functions of this organ is to process the external stimuli which arrive through the senses of vision, smell, hearing, touch and taste. The sensory organs correspond to the five different channels through which data is received from the outside world.

As soon as an external stimulus which could be a photochemical, smell, sound, touch or a biochemical signal arrives in the sensory organs, the associated neurons undergo complex chemical changes converting the stimuli into electrical impulses. The electrical signals are then passed on from one neuron to another by release of chemical substances at the synapse — the junctions between two neurons. On reaching the brain, these impulses are processed to generate appropriate responses.

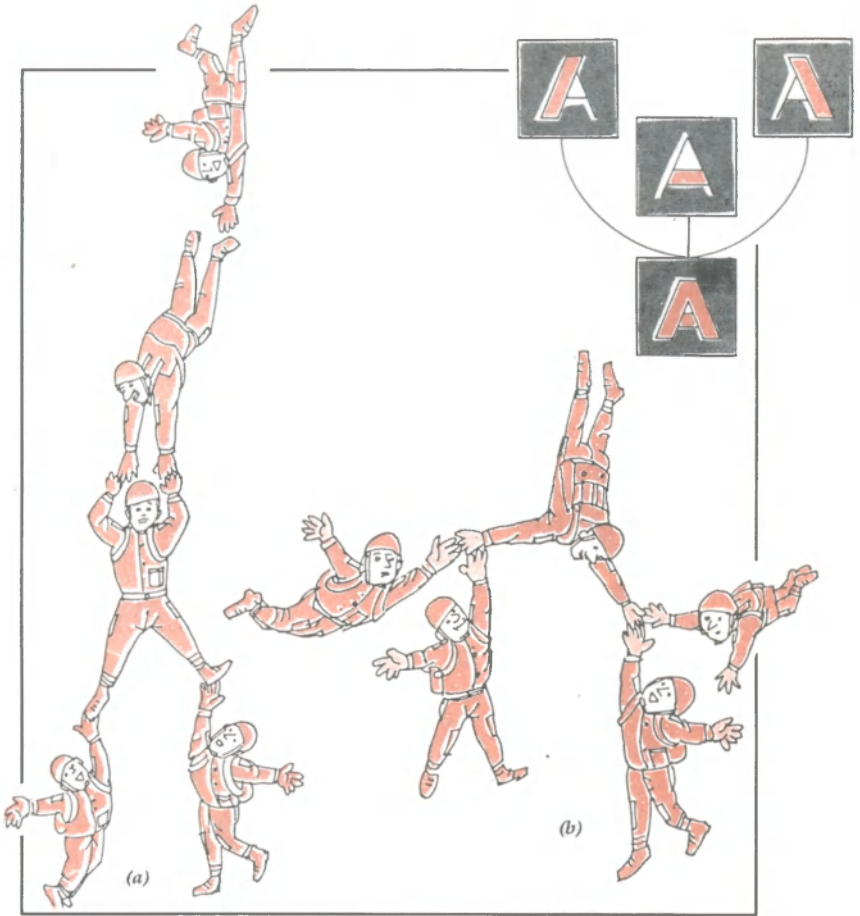
While processing information, the human brain is involved with vital activities such as 'pattern recognition', 'learning', 'thinking' and 'remembering'. All these processes occur in the cortex which is a thin layer on the surface of the brain. In addition to



Hebb's hypothesis; The release of neurotransmitter molecules at the synapse (inset)

these basic activities, processes such as language interpretation and speech generation also take place in the brain cortex.

Learning is associated with much more than mere acquisition of factual information. It is a process which causes a permanent change to take place in our behaviour. Physiologically, some corresponding change also takes place in the brain with the acquisition of knowledge. Donald HEBB has related the process of learning with some changes taking



A cell assembly resembles a group of sky-divers holding each other in patterns either (a) or (b); A cell assembly identifies letter 'A' by its three defined strokes (inset)

place at the synapse across two neurons. The chemical transmitter substance released at the synapse causes the next neuron to fire. If the two cells activate over and over again then the surface area of the synapse increases and the two neurons tend to act as a pair. With increased surface area there is an obvious strengthening of the connection at the synapse. Hebb's theory is a plausible hypothesis of the process of learning in the human brain.

An individual cell in the visual perception region of the brain may learn to identify say, a stroke in the letter 'A'. A large number of cells which together are capable of identifying the entire letter 'A' are collectively called a cell assembly. It is only when this association is formed that we say we have learnt to recognize the written letter 'A'. A cell assembly is a sort of **template** formed in the brain which responds when the appropriate visual signals are received. These groupings can be extended to include individual words, sentences, etc. Similar assemblies are also found to recognize acoustic **phonemes** to identify the basic components of speech sounds. Conventional learning therefore, involves strengthening of linkages between many thousands of neurons. There is evidence to believe that an actual learning task involves forming new links between individual cells or between association of cells which already exist. This also helps us to remember individual events, words, etc. by association with separate individual pieces of information already stored in the brain.

Learning to Learn

As they grow older, all men and women acquire in their own individualistic manner, ways of responding to unusual situations. It is remarkable how they select their responses which enable them to tide over a particular crisis or solve a problem. In fact, whenever they come across a new situation they subconsciously work out a strategy to arrive at a convenient solution. Harry HARLOW (1905–1981) conducted experiments which demonstrated that people can solve problems in areas they are already familiar with (at least in principle if not in detail). Harlow called these as "learning sets" as they help human beings to learn how to learn. He suggested that the method commonly referred to as "trial and error" is actually an orderly development of learning and thinking processes. When we first face an unusual situation we use the

trial and error approach. Subsequently, when a similar situation arises at a later point of time, we find the appropriate solution almost instantaneously (based on previous experience). This is commonly known as "insight".

Jean PIAGET (1896–1980) studied children over a long period of time. According to him, children go through a series of stages while developing their understanding and thinking skills. Cognitive development is related to the development of thinking, perception and making use of the memory. Piaget suggested that a child's intellectual growth takes place in a stepwise manner through the development of various "schemas". These are internal (or mental) representation of physical situations or actions. Usually, they are a set of rules which the child creates, stores in his memory and subsequently uses while interacting with his environment. It is only through activity and interaction with the environment that the child develops and modifies his schemas.

Schemas are formed by the processes known as assimilation and accommodation. The manner in which a child interacts with the environment depends upon his schemas at that moment. This process of understanding the world is known as assimilation. This process enables him to remember and to obtain more and more information about the world. He receives feedback from the environment which he uses to verify the accuracy of his perceptions. When there is a mismatch between his earlier schema and the feedback he has received then he is in a state of 'disequilibrium'. Quite often this is only a temporary phase and the child quickly evolves a new schema to cope with the changed situation. The earlier schemas may be modified to generate a new one or it is just discarded. Continued interaction with the environment brings new forms of stimulation to the child and this process of changing his mental structures is known as accommodation. The assimilation of information into schemas and the accommodation of schemas to cater to the demands of new experiences occurs all through the child's development.



Assimilation & Accommodation

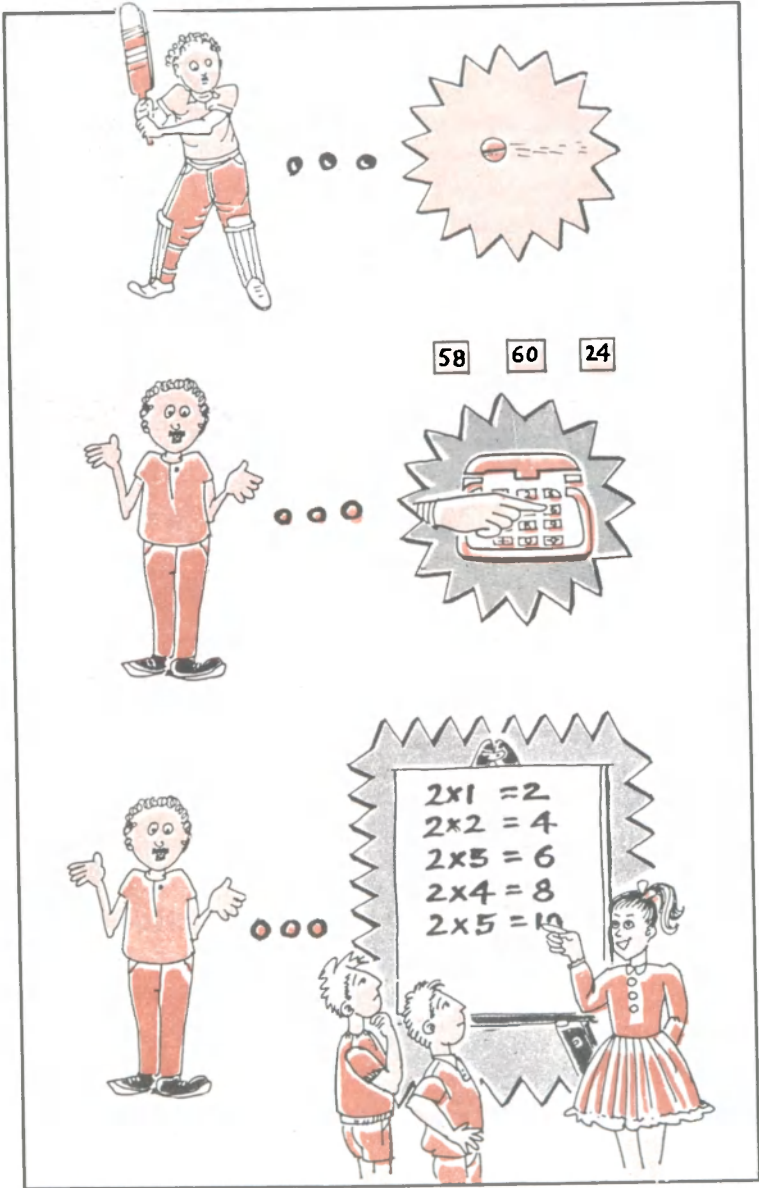
Intelligence is usually explained as being the ability to acquire new skills and knowledge. A child develops cognitive abilities in order to understand and interact more successfully with his environment. It also enables him to form new concepts and to grasp their significance.

Human Memory

It is important to realize that there are three basic activities associated with human memory, viz. receiving new data from the environment followed by its processing and conversion into usable information within the brain; storing this information in the memory; and retrieving it as and when required.

Humans have three basic forms of memory. One is the immediate memory. Here one retains all details of an object or a scene or a picture for about 1/10th of a second like a batsman facing a ball. After that, much is forgotten. Next is the short term memory where information is retained for a few minutes only. This memory is also known as the working memory and as such has a limited capacity. It is also used for rehearsal, i.e. for mentally repeating the information over and over again so that it enters the long term memory.

The information which is stored in the long term memory can be broadly classified into four categories. First is the information related to specific tasks, which includes riding a horse or a bicycle, driving a car, typing, painting, carving, etc. These skills often require several years of experience to perfect. Secondly, the information related to fear-inspired actions like being bitten by a dog, hurt by a falling wall, burnt by a fire, etc. By proper training one can prepare oneself to take specific actions under difficult circumstances or emergency conditions. Third is 'Episodic memory', in which information regarding dated episodes or personal experiences is stored date-wise. Finally, the 'Semantic memory', which relates to the use of words, grammar, metaphors, and all



The basic forms of human memory



Types of long term memory

other niceties of a language. However, these different groupings are conceptual in nature and no specific region of the brain as such is reserved for these categories. Information from the short term memory is transferred to

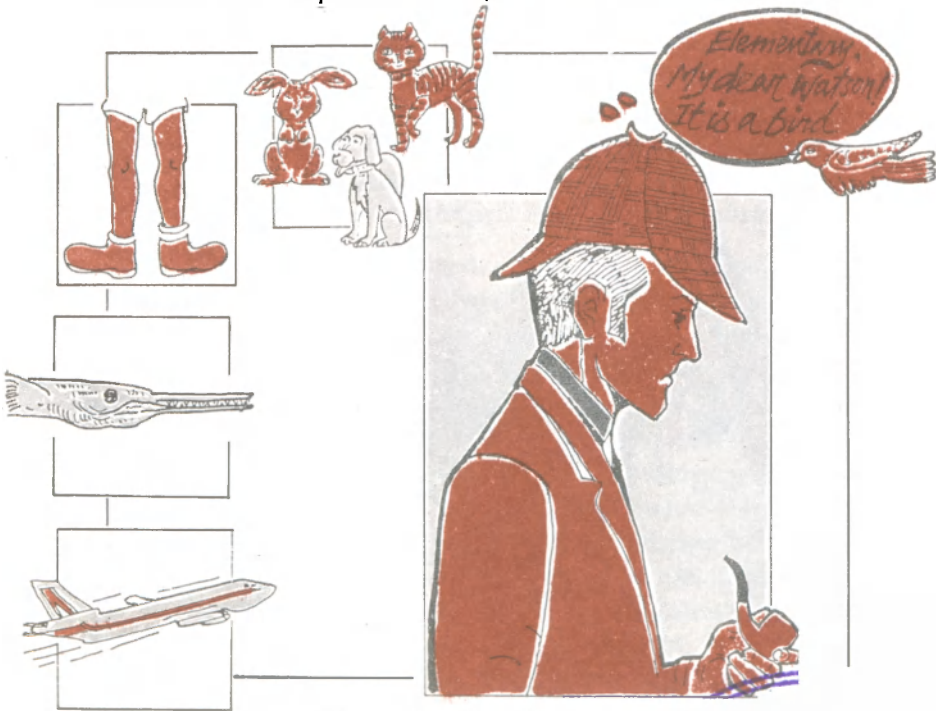


the long term memory by the process of repeating it a large number of times as one learns poetry in school.

It is common knowledge that people tend to forget. One explanation for this phenomenon is that unless information is frequently retrieved and used, it tends to

decay with time. Another explanation is that memory is stable but with time other information is stuffed in and the original contents of the memory are weakened or displaced.

Human beings have an ability to take cognizance of only some of the incoming information while discarding the rest. This phenomenon is known as "attention". Unless one consciously pays attention, most of the information arriving from the sensory organs is lost. For example, when one enters a room with a large gathering of people, the first impression is one of a buzzing confusion. However, if someone mentions your name or happens to be discussing a subject which is of particular interest then you suddenly concentrate and try to follow that conversation. This is a common phenomenon and is known as the 'Cocktail Party' effect. It is only when we pay attention that the processed data or information is sent to the



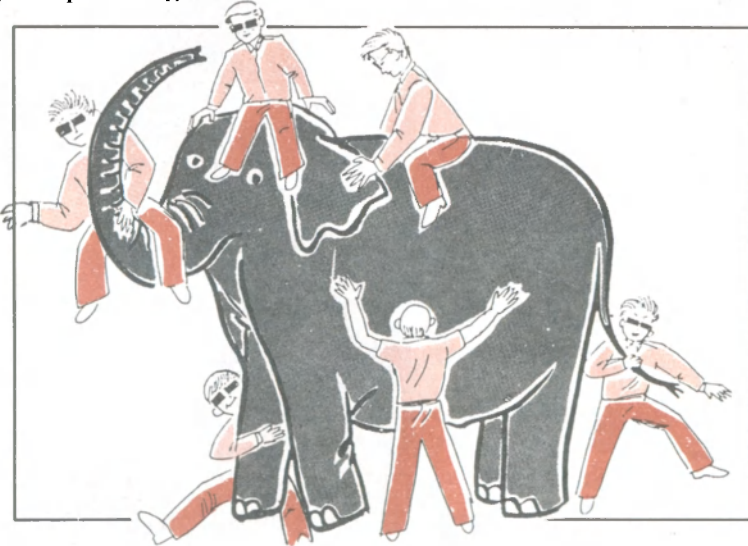
Perception: Guessing game

short term memory and then perhaps to the long term memory. It would be extremely taxing on the brain if we did not have this ability to be selective.

Perception is the creation of a visual schema in our mind based on clues arriving from our sensory organs. We translate the signals from the sensory organs into individual, meaningful and stable items in our mind. Since a schema is an implicit mental image or a theory which we create on the basis of worldly experiences, we are likely to have a large number of schemas already stored in our memory. These are created specifically to understand the world we live in.

According to Hebb, perception is an acquired characteristic. The sensory experience is registered in the form of cell assemblies. As the sensory patterns become more complex and stable, the entire combination of cells are activated together in response to a specific stimulation. These larger sequences are known as 'phase sequences'.

However, according to the Gestalt school of psychology, perceptual organization is an inborn characteristic. The word



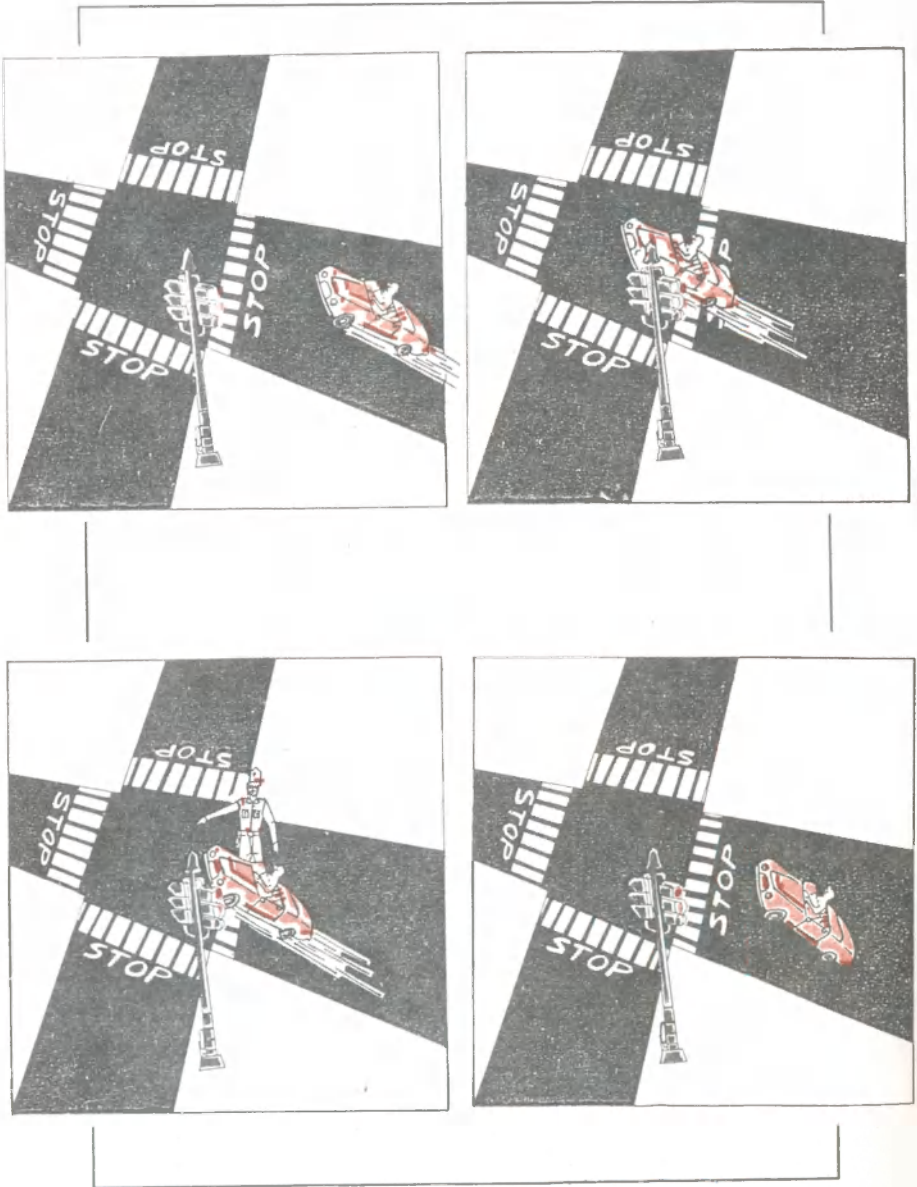
A contrast to Gestalt's theory of perception

gestalt is German for 'pattern' or 'form' and the theory emphasizes the ability of human beings to perceive patterns as a whole or in their entirety. When we perceive, we give meaning to objects by their characteristics as a whole (in toto) and not by considering them as a jumble of parts which go into making up a total figure. The clues from the sensory organs along with the rules or guidelines help humans to recognize these patterns which in turn will help in his perception.

The Thought Process

As we live and grow, we interact with our fellow human beings and also with the environment. Consequently, our collection or store of schemas keeps on increasing. The sum total of the schemas which we possess eventually becomes our "knowledge". This word is used in the widest sense of its meaning. Our schemas include a variety of visual descriptions such as identification of written characters, repertoire of written words, pictures, spoken phonemes, spoken words and so on. These are stored in our long term memory.

The schemas mentioned above are at the elementary level. We also possess a mental ability to link schemas together and produce new schemas which are at a higher level. This process is known as "thinking". For example, while driving early in the morning we may come across a red light at the traffic junction. Now, there is a problem. Should we risk it and proceed without stopping. We could undergo a thought process as follows. We visualize a situation where we do not stop but proceed. This is followed by visualizing ourselves being stopped by the traffic police stationed just around the corner who either fines us on the spot or asks us to appear before a magistrate. As yet we have seen neither the policeman nor the magistrate (but we know that they do exist). Not wishing to experience it, we decide to apply the brakes to the car and halt at the red light. This process of



Evaluating the different possibilities before arriving at a decision

mentally going through a series of actions is “thinking”. It ends in our deciding on a course of actions which, hopefully, is in the best interests of all concerned. The thinking process need not necessarily be concerned with real objects only but could involve abstract concepts as well!

We have yet another ability and that is to mentally classify things and events around us. Based on our earlier experience we learn how to group objects or events according to characteristics or attributes which are common to them. For example, there are certain types of plants which though come in all sizes and shapes are yet grouped as trees. So, we have a concept or an idea regarding the attributes or characteristics which are similar and which help us to group such plants as trees. We would certainly not classify grass or shrubs as trees. We learn of these facts by observation, discussing with other people and by reading books.

Concepts are symbolic for us human beings. They are very useful and help us in comprehending complex objects and situations. The concept of a “car” can be vividly recalled in our minds on just hearing the sound of the engine or its horn or the screeching of its brakes. This concept leads us, by our thought processes to recollect about the mechanism of the car, the need to buy petrol, the price of petrol, requirement of a license to drive a car, new traffic laws which are currently being enforced, etc. There can even be a hierarchial relationship amongst these new concepts as to which is more important from say, convenience of transportation or from the speed aspects or from the running cost points of view.

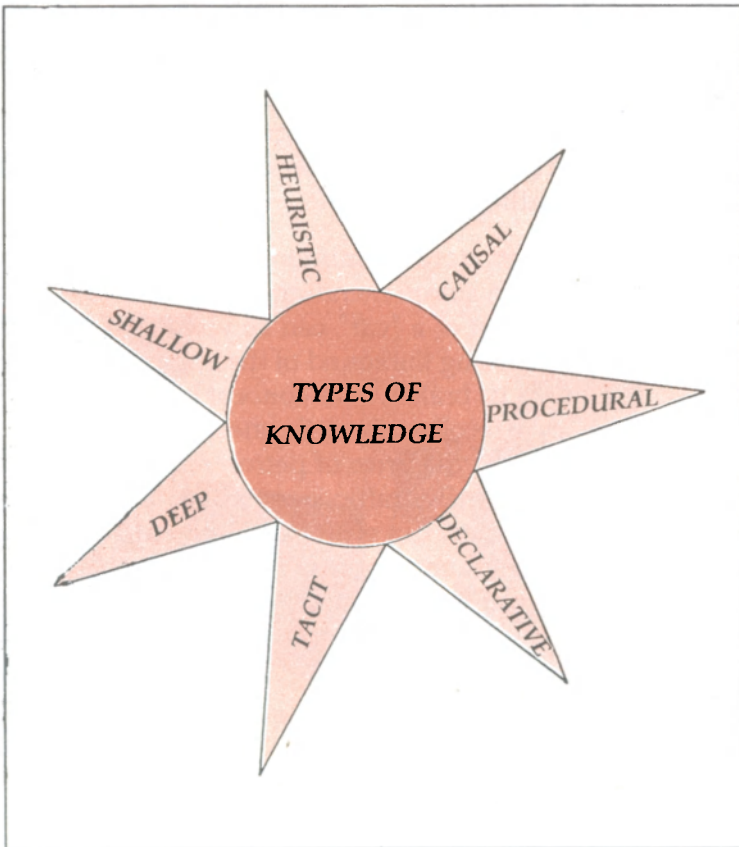
Epistemology is the study of knowledge. It is concerned with the nature, structure and origin of knowledge. Knowledge too has been sorted and classified under many groupings. For example:

* *Causal knowledge* is the understanding of the underlying cause and effect of various aspects of life.

* *Procedural knowledge* is knowing how to do something, e.g. how to make tea.

* *Declarative knowledge* refers to knowledge being expressed as a declarative statement e.g. "true"/"false".

* *Tacit knowledge* is also known as the unconscious knowledge since it cannot be expressed very specifically in



words. For example, how to ride a bicycle or a horse is something one may know but find difficult to explain.

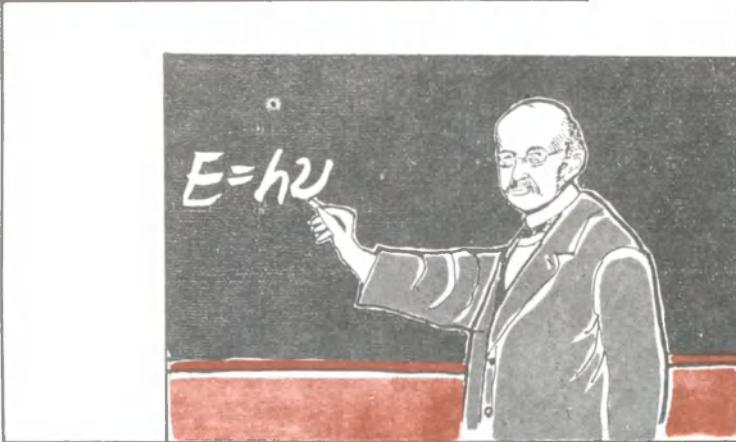
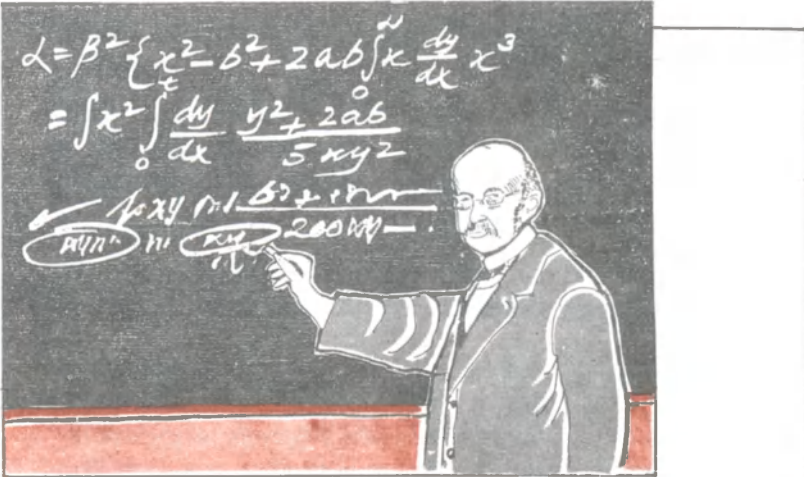
* *Deep knowledge* is acquired by exhaustive study and understanding of a subject. Usually it is restricted to one or two domains of knowledge as it is not possible for a single person to study all the different areas of learning.

* *Shallow knowledge* is based on a superficial or an empirical understanding of a subject.

* *Heuristic knowledge* is a type of shallow knowledge. It consists of various rules of thumb based on first-hand experience. They aid in solving problems but there is no guarantee that they will succeed. It is in areas such as medicine and practical engineering that heuristics play a useful role in problem solving.

Language, an invention of man, is the symbolic vehicle of our thought process. Besides externalizing his thoughts by speaking and writing, man also uses language to internalize his thoughts i.e. to think about his ideas or concepts in his own mind or to find solutions to problems. Language, therefore, is very helpful in widening the scope of human knowledge.

Yet another unique characteristic of human beings is creativity and this includes both creative thought as well as creative action. They both refer to generating something original, something new which no one else had thought of earlier. An example of the former is the concept which Max PLANCK (1858–1947) introduced regarding energy. He suggested that energy could be quantized at the atomic levels. This novel concept which he introduced in 1895 found immediate acceptance as it resolved a number of vexed problems in physics and chemistry. Examples of creative activities are the much appreciated wall paintings of the Ajanta caves, sculptures of the Ellora caves, poetry by Kalidasa, etc. Researchers have been intrigued as to what makes a man



Max Planck

creative and many theories have been put forward on this subject but these remain speculative.

Thinking, conceptualization, problem solving and decision taking are all intimately related to one another. They are closely linked with man's knowledge and intelligence. Quite often the process of thinking is followed by a decision which may have to be taken. Some decisions have far reaching consequences and once made cannot be retracted. There

are also a large number of decisions which we take one day and change the next day. This is often done on the basis of new data or information which becomes available. When we take a decision we can usually explain why our proposed course of action is the best one. Nevertheless, it is not so easy to give a detailed account of how we arrived at our conclusion. They are generated in the mind where emotion, creativity and intuition are all at work. The goal of AI is to simulate the human mind, particularly those aspects which relates to thought process and decision making.

In many problems requiring solutions, one finds that there are no black and white answers. Invariably, an element of probability is involved. A large science has grown up around this probabilistic concept in problem solving and is known as the Fuzzy Set Theory.



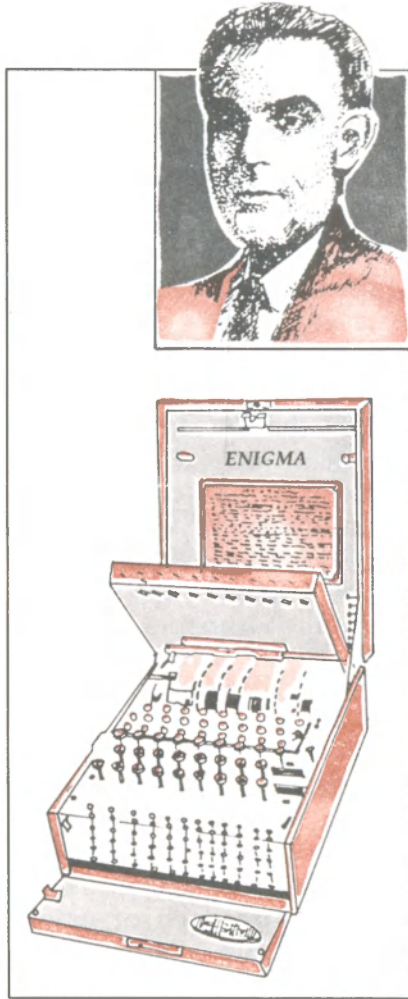
Scope and Extent



By the late 1940's machines to assist man with mathematical calculations had reached an advanced stage of development. At that time many scientists began to ponder over the possibility of constructing machines which could think. They got encouragement from the development of a machine called ENIGMA by Alan Turing. This machine was designed during the War to decipher coded German messages. Turing subsequently contributed to the design of a stored program computer in U.K. His ultimate aim was to design machines which understood the general logical operations like AND, NOT and OR. His idea was that these symbols used to describe mathematical processes and logic could also be used to manipulate words. Turing's extension of machines to process non-numerical symbols has been quoted in some quarters as the starting point of AI. Turing died prematurely in 1954, but the ball had already been set rolling.

Meanwhile Anthony OETTINGER had been attempting to use a computer to translate English words into Russian and vice versa. An English word fed into the machine would result in

a Russian equivalent to appear on the screen. Here, what took place was matching of equivalent words from the two languages. For every English word a Russian counterpart was selected. It was fondly hoped that this beginning would eventually lead to a machine which could translate continuous text from one language to another. Unfortunately, the *syntax* of the two languages had not been taken into account. The procedure which a common man uses in stringing words together to form meaningful sentences had yet not been properly investigated at that time. This led to serious problems in automatic translation of texts and often resulted in amusing situations. For instance, an English sentence, *The spirit was willing but the flesh was not* came back after retranslation as *The wine was good but the meat was spoilt!*



Alan Turing and his machine

With this background, a group of eminent computer scientists gathered at the Dartmouth College (U.S.A.) in the summer of 1956 to discuss about possible "thinking" machines. Many ideas were put forth and discussed. Subsequently, they

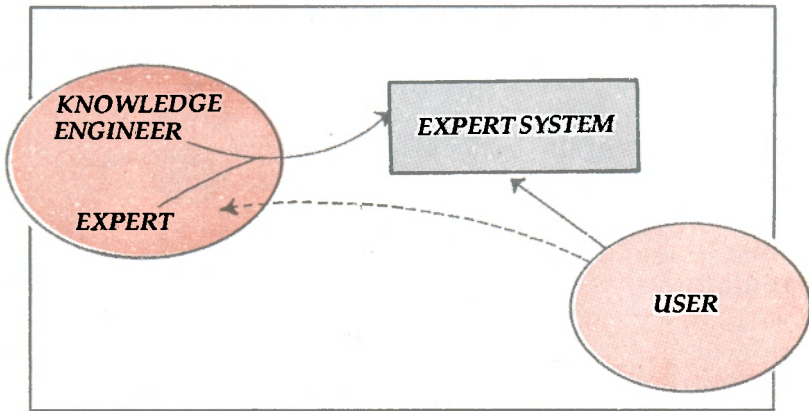
all went their own ways to develop new systems. Many of them founded their own research centres to develop such machines along with the necessary software. The term “Artificial Intelligence” came into common usage only after this meeting.

It is important to note that there is a difference between an AI system and AI. The former is a computer based system that lets you solve certain types of problems. On the other hand, AI is an academic discipline where one studies how human beings behave intelligently and how machines could emulate this intelligent human behaviour. An advantage of AI is that it is a useful discipline and enables us to solve problems in a practical manner. The procedure consists of first studying the problem, getting to know all the facts and the interrelated rules and then finding an appropriate solution.

Knowledge Engineering

AI concentrates on knowledge engineering. Here “knowledge” refers to a particular domain of expertise. This expertise could include highly specialized or technical information which would not normally be expected to be available with an ordinary person. The term “engineering” refers to a practical approach of using this knowledge to solve problems. This is indeed a commonsense way of doing things.

Knowledge engineering is the discipline that deals with the way a knowledge-base is organized so that it becomes available in the most useful manner possible. This knowledge is collected by a knowledge engineer by searching literature or even by asking people who are considered knowledgeable in that particular subject. His next task is to arrange this knowledge in some kind of an order so that it is easy to access and retrieve quickly. This structuring of information is however, much easier said than done. One way would be to place



Development of an expert system

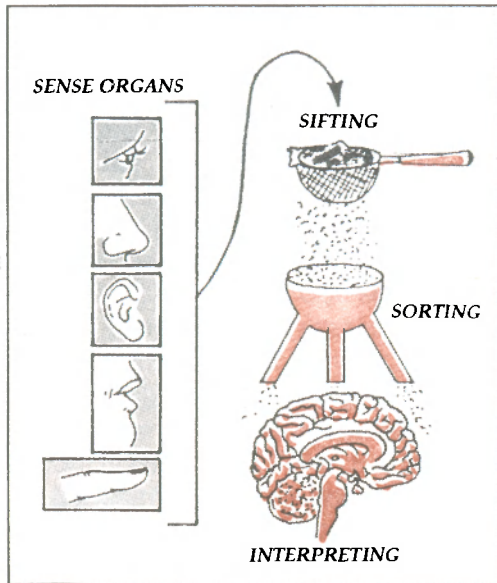
the information most likely needed at the top and the least likely at the bottom, but then who is to decide? It is often difficult to get human experts to unanimously agree as to which aspect is more important than others. Unlike computer systems which can scan through its entire memory in a matter of seconds, it takes a considerable time to extract information from a real human expert. Usually they are hesitant to tell what they know. This is one of the reasons why developing a knowledge-base for an expert system requires such a long time.

Can Computers Think?

When we say a computer is thinking what we really mean is that the computer is executing a program and is handling information and manipulating it just as humans do. Here a pertinent question is, do people know how they think? They can tell what they are thinking about and what is the result of their thought process, but certainly, they are unable to explain the mental steps involved.

A person acquires data, information and other details from his environment. This information keeps arriving through

his sensory organs on a continuous basis. Since all data received are not important, the brain sorts them out and rejects most of them. But the remaining data which are considered vital or important are then organized and stored as 'knowledge.' Finally, the stored knowledge is retrieved as and when required and is used to solve problems, make



Acquisition and processing of data

decisions and assimilate additional new information which has just arrived. When one wants to gain knowledge in a new area of study, one makes a conscious effort in learning everything about it. Once the existing body of knowledge has been acquired one seems to know how to apply it. The unconscious ease with which knowledge is applied is what makes intelligence so difficult to quantify and hence simulate. The reasoning mechanism in man is difficult to explain in concrete terms. With the goal of AI being to simulate intelligent behaviour, computers have to be taught how to analyze problems and to take decisions just as human beings do.

For a computer program to be considered intelligent it is necessary that it acts as a human being would. Its thought process however, need not be exactly same as that of a human being. For example, an assistant in a large office spends most of his time physically retrieving data from existing files,



Who's smarter?

reorganizing it, preparing new reports and so on. If he does a good job of extracting the desired information from the huge volume of data then his superiors do not hesitate in praising him and saying that he has used his brains (i.e. used his thinking process). If a computer can manipulate data in a logical manner then it too should qualify to be called intelligent (or at least as intelligent as the clerks) !

Finally, the scope of AI has grown so rapidly over the past 35 years that it has already branched off into many categories. Besides expert systems, natural language processing and robotics, it has even been extended to machine learning that helps a user to operate new machines and systems, knowledge representation, data search methods and logic programming. Each one of these have become important subjects of study in many research centres of the world.



A Problem Solver



The ability to solve problems is usually taken as an index of intelligence for human beings. It is tempting to extend this criterion to machines as well. There are two kinds of problems one comes across in real life. Firstly, there are those which can be solved by strictly adhering to well established procedures or algorithms. These are algebraic (or formula) type of problems. Here all one does is to feed data related to some unknown variables in a formula and the machine calculates and produces the correct answer. However, many problems do not quite come under this group. They relate to a search through a large volume of data to find the correct answer. AI is mainly concerned with such problems.

Soon after the Dartmouth Conference, a technique called the General Problem Solver (GPS) was put forth by Allen NEWELL and Herbert SIMON. Many other powerful techniques have subsequently been developed, which come under the general category of 'expert systems'. However, many practical AI programs continue to use the basic concept of the GPS technique.

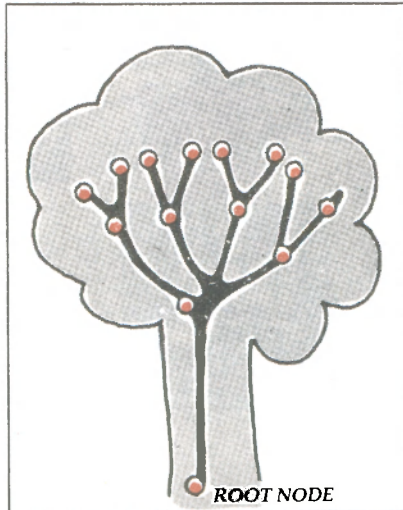


Alan Newell and Herbert Simon introduced the "IF-THEN" rule

Newell and Simon demonstrated that many of the human cognition or problem solving capabilities can be expressed by production rules of the type "If-Then". For example, "IF you are thirsty, THEN drink water". Or "IF you see a red traffic light, THEN stop the car". These production rules apply only to small but specific portions of knowledge which are known as 'chunks'. Such chunks are arranged in a loose manner in the long term memory. Moreover, Newell and Simon believed that these chunks are linked to other allied chunks of knowledge according to certain rules. An example of one chunk being connected with another related chunk is: "IF you are thirsty AND your glass is empty THEN fill up the glass with water". Thus, they popularized the use of IF-THEN rules to represent human knowledge.

Until the mid 1960s efforts were being made in the area of AI to design intelligent systems which relied less on domain knowledge and more on the methods of reasoning. Now arriving at a solution based on a search through a huge mass of data can be an extremely lengthy process. To overcome this drawback, alternate search procedures were developed and these continue to be used in the expert systems.

One of the methods used to simplify the search is to draw a tree diagram with possible solution paths. A tree diagram consists of nodes which correspond to a certain 'state' where the problem is at a particular moment. It can then move to other nodes or states by various paths. The task of the problem solver is to discover a se-



A simple tree diagram



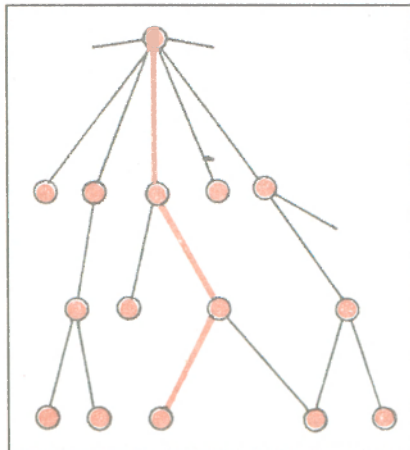
Search for the missing purse

quence of state changes which initiates from a 'starting state' (or from where the problem was originally posed) to a solution or the 'goal state'. Let us assume the problem is a search for your missing purse. You are reasonably sure it is somewhere in the house. So you draw a diagram indicating how you might go about the search. A possible path you take might first lead you into your bedroom and you then search the wardrobe, the writing table and then search the bathroom and finally the kitchen. A pictorial description is useful since it helps to visualize the search path even before it has been undertaken. One could prepare many alternate search programs in advance. These are then compared and a path most likely to be fruitful is selected. However, a heuristic approach for finding the missing purse would be to try and recollect which pair of trousers you were wearing last evening and then to search its pockets.

Search Strategies

Once the tree diagram is established, one has to evolve a strategy to proceed from the starting point to the goal or solution. If this search network is an extensive one, then keeping track of the path is difficult. There are four types of strategies which are commonly used to find a solution to a problem, viz. depth-first, breadth-first, bidirectional and means-end.

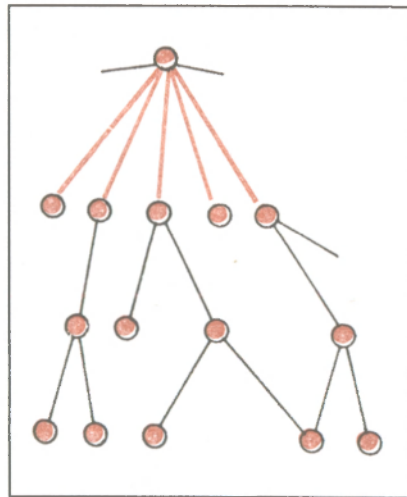
Depth-first search: To execute a simple search program, one could adopt the 'depth-first' strategy. In this method each possible



Depth-first search

path is explored to its end before another path is tried. One goes systematically from the beginning to the end of a path and investigates all the nodes and their linkages to the next nodes till the right solution has been found. This depth-first technique is bound to find the solution because, in the worst case, the solution will correspond to the last node. However, to find the optimal solution one has to traverse almost all the nodes. This is a very time-consuming method particularly when there are long branches to be searched and they do not contain the solution.

Breadth-first search: The 'breadth-first' strategy is the opposite of the depth-first method of searching. Here all the nodes at the same level are searched before proceeding to the next deeper level. The advantage of adopting the breadth-first technique is in situations where the solution is close to the surface. In case the solution is deep inside, then once again a considerable amount of effort will have to be expended in arriving at the solution.



Breadth-first search

Bi-directional search: This strategy consists of two searches conducted simultaneously. One works forward from the initial state and the other works backward from the goal state. New levels are generated alternatively and as soon as the two processes meet then the shortest possible route would have been found.

Means-end analysis: In this analysis, the problem is usually broken-down into a series of smaller sub-problems, each of

which is solved separately. This has the effect of keeping the size of each search-space within reasonable limits. One now moves from one sub-problem to another. The new tree will consist of nodes which are both the goal of one sub-problem and the root node from which the search for a solution to the next sub-problem would commence.

Generally speaking the programmer has to choose between the depth-first and the breadth-first search methods. The choice eventually depends on his earlier experiences in solving similar problems. The aim of AI is to show the programmer a superior methodology of choosing between either of the two techniques. Heuristic methods are of great help in solving such problems since they are based on personal experience and intuition. This search works much better than a blind search since it is used to decide which particular node to search next.

Knowledge-base

A knowledge-base contains all the facts, rules and procedures related to a specific field. It is important for solving problems in that field. In contrast to a human expert whose capacity to work diminishes with age, a knowledge based expert system can be made more and more efficient by collecting expertise from not one but a number of experts. The computer program along with the knowledge-base can be copied to create any number of expert systems and so the expertise is never really lost. Moreover, an expert system is available any time of the day unlike a human expert who needs to rest between work. Above all, a computer expert system has no preferences as to whom it provides the expertise. The user may be a very senior Director in an organization or a junior officer. The same expertise is provided to both.

Before loading information into the knowledge-base, the knowledge engineer has to define the relationships between facts, objects and groups of objects. Let us suppose that the

objects within a restricted knowledge-base relate to aeroplanes. There are many different kinds of aeroplanes: transport planes, military planes, passenger planes and so on. The relationship between the objects which are classified as aeroplanes is that, they satisfy the definition of being machines with fixed wings and engines and which fly in air. So, a passenger plane inherits this categorization and also has an additional definition of being capable of transporting people. This procedure is known as object-oriented programming and is commonly used in developing expert systems.

The knowledge which has been obtained from an expert has to be formalized and structured. Usually the knowledge engineer sits down with the expert who describes a large number of cases he has handled or problems he has solved. From these interviews relevant knowledge is extracted. Another way of obtaining the expertise is to observe an expert while at work.

The knowledge engineer has to make the ideas so explicit that the computer can accept them. An object is defined first by assigning a name to it, then its attributes are specified and finally values are assigned to these attributes.

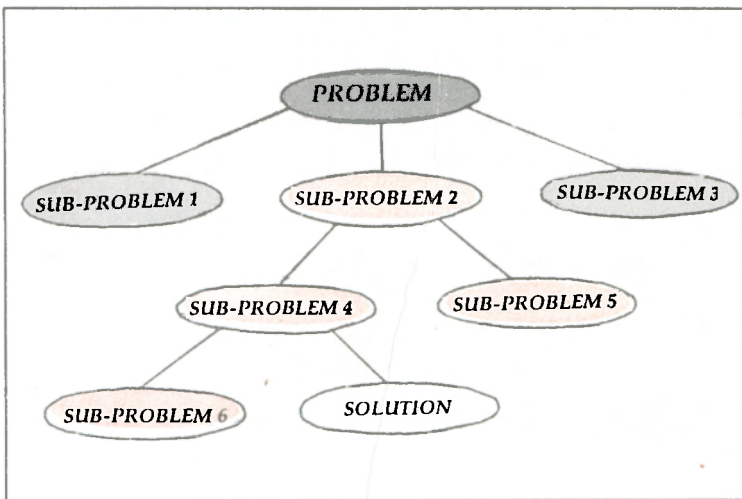
Name of the object:	Bicycle
Attributes:	Steel frame, steel handles, pair of pedals, two large wheels, chain drive to rear wheel and a leather or plastic seat.
Purpose:	For personal transport.
Operation:	(a) Hold the handles with both hands; (b) Put left foot on the pedal; (c) Push the bicycle with right foot; (d) Climb bicycle; (e) Pedal with both feet; (f) Steer with handle; and (g) Reach destination.



A guessing game defines an object for a knowledge-base

In a conventional computer, the data and the program are both loaded into the system. The program is in the form of algorithms and each individual processing step which the computer has to take is well defined. In an expert system, a knowledge-base comprising declarative, procedural and heuristic knowledge is created and entered into the machine. The ways by which knowledge can be expressed are by tree representation, semantic networks, production systems, frames, list and predicate logic.

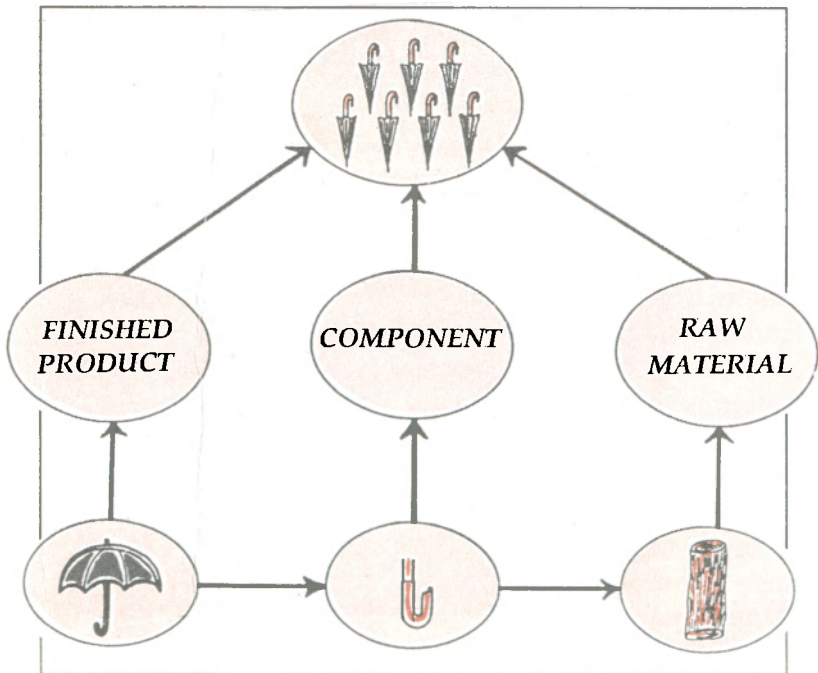
Tree representation: A tree is a graphical representation of the knowledge-base. It consists of nodes which store information and branches which connect them. The root node is the highest node in the hierarchy of knowledge. From this root node are paths or links which lead to other nodes. A simple example of such a graph is a road map of a country. The cities are the nodes and the roads are the links. Trees also store hierarchial knowledge. An expert system which uses this knowledge tree would ask its user relevant questions as it moves through the tree.



Pathways to decision

Another application of the tree structure is in arriving at decisions (these are called decision trees). It begins with the statement of the problem as the root node. From this node decision paths lead to new sub-problem nodes. Heuristic rules are used extensively in arriving at decisions. In a way, these rules prune the decision tree of its unproductive branches.

Semantic networks: A semantic network is a traditional method of representing propositional information. A proposition is a statement that is either true or false. It is a form of declarative statement since it represents facts. (A proposition is sometimes called 'atomic' since it cannot be further divided). Semantic networks are graphic representations of knowledge based on objects and their relation-



A simple semantic network

ships. The nodes of the networks correspond to objects and the link between the two nodes represents the relationship between the objects. Such networks are used to analyze the meaning of sentences in natural language processing.

Production Systems: In an AI program where a course of action is to be recommended it is more efficient to use a procedural system of knowledge (i.e. facts and rules which tell what to do with the facts). A popular procedural system of knowledge representation is with the help of production rules which describe condition-dependent actions and this includes the IF-THEN rules.

Some advantages of production rules are:

* Explanations become simple. To explain a particular reasoning all one has to do is to keep track of the rules which have been used and display them on the screen.

* Modifications are easy. In rule-based systems the knowledge-base can be modified by adding, deleting or changing the rules.

* Understanding is also easy since the production rules are extremely readable.

Frames : A frame is a method of representing objects and their relationships in the form of columns and rows. It is useful since the method envisages breaking down objects or situations into their constituent parts. These parts are entered in corresponding slots in the frames. The

FLTN	ROUTE	ETO	REMARKS
483	LEH	0630	ONTIME
479	BOM-100-DPR	0650	ONTIME
815	LKO	0700	ONTIME
433	BHO-10R	0815	LATE
401	DEL-DIV	0911	ONTIME
402	DEL-DIV	1000	ONTIME

ETO : EXPECTED TIME OF DEPARTURE

BROUGHT TO YOU BY
INDIAN AIRLINES
NATIONAL INFORMATICS CENTRE

slots can be further subdivided for more detailed structuring. The frame is not a universally defined data structure. In each case the frame structure and the slot have to be clearly redefined. The frame provides greater flexibility in representing knowledge than the net. Frames were originally developed to represent stereotyped knowledge (i.e. those with well-defined features). Their columns and rows are easier to understand compared to logic and production rules.

List : Knowledge can also be represented as lists like a card catalogue in a library. The cards contain all details such as the authors' names, title, publisher, contents, keywords, etc. One



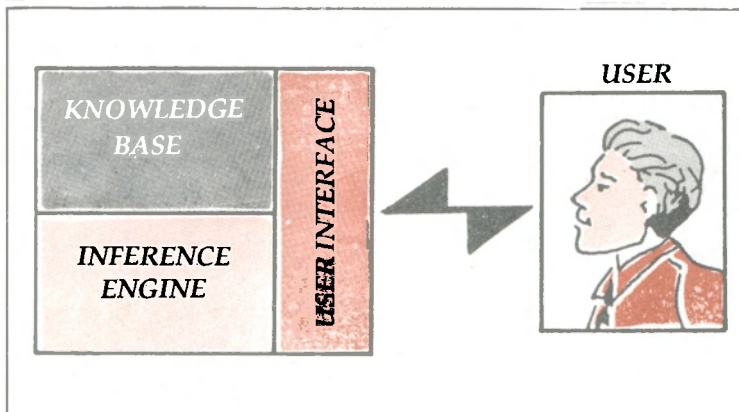
A library catalogue list

can go through each card, rejecting those which do not match what one is searching for or select those which do. By proper indexing one can make lists as efficient as the trees.

Predicate logic: The fourth way of representing knowledge is in the form of statements or predicates. It is a formal way of describing objects and their logical relationships. It also includes grammar or syntax for generating valid logical statements. Predicate logic has semantic rules (meanings) which relate the symbols of the formal language to the objects and the processing rules which can generate valid logical expressions from other logical expressions (e.g., **PROLOG**).

Inference Engine

Just having a great deal of knowledge does not make one an expert. One must know how to select the appropriate knowledge and how to apply it. Similarly, a knowledge-base alone does not make an expert system. The system must possess another component which directs the implementation of the knowledge. This component is the inference engine. It decides which heuristic search technique is to be used



Major components of a knowledge based system

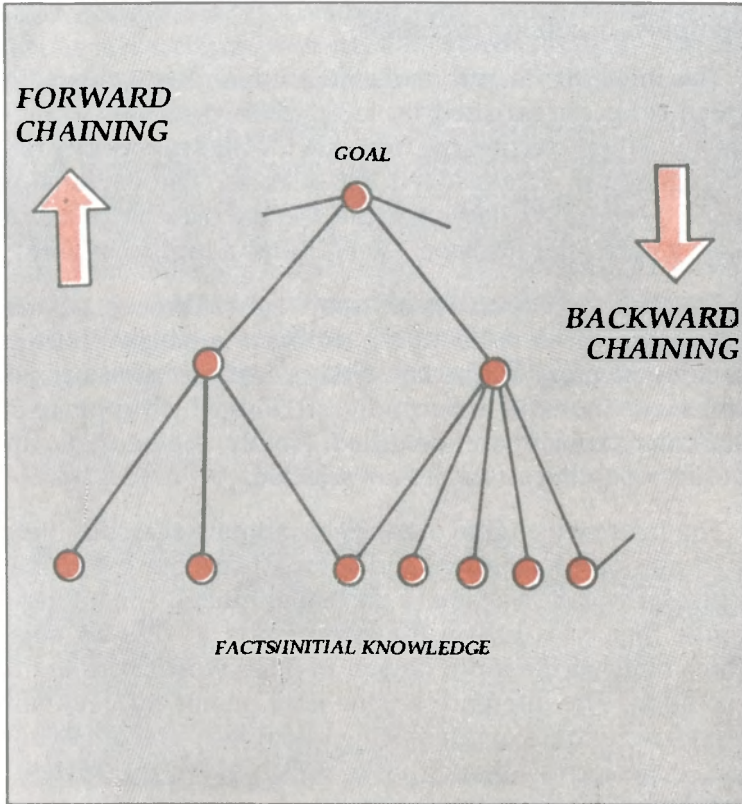
and to determine how the rules of the knowledge-base are to be applied in solving problems.

The inference engine makes the inferences by deciding which rules are satisfied by facts, gives priorities to these rules and then executes the rule with the highest priority first. The agenda is a prioritized list of rules. The explanatory facility keeps on explaining to the user the reasons for having taken a particular decision. (Very unlike a human expert!)

Therefore, an expert system operates in a three-stage manner. Firstly, all data about the problem is entered into the machine memory by the knowledge engineer after consultations with the experts. Secondly, all rules which apply to the particular problem are identified. Finally, the heuristic and conflict resolving strategies are selected.

The inference engine organizes elaborate searches, helps in sorting of information, and comparisons between a particular information against a particular model. The inference engine then tries to use the information to find an object which matches the input data or in other words tries to find a solution. The inference engine uses one of the few well-proven techniques of inferencing to arrive at a recommendation. It performs all such tasks which gives the system a semblance to human intelligence.

There are two general strategies for inferencing which are used in expert systems. These are the 'forward-chaining' and the 'backward-chaining' methods. There are other strategies as well which include means-end, backtracking, etc. A combination of inferences which connect a problem to its solution is known as a 'chain'. A chain which searches from the problem statement to its solution is known as forward-chaining. It is also known as "data-driven" since the inference engine uses the information provided by the user to move through facts and rules in the knowledge-base to reach a conclusion. The inference engine in this case starts from the set of data provided by the user and then tests all the



Inferencing strategies used in expert systems

hypothesis in which this data plays a part. If at the end the machine does not reach a conclusion, it can then ask for additional information or would simply indicate that there is no solution. The system operates on the basis that all rules must be satisfied without exception. Hence the forward-chaining method starts with the data which has been fed-in and it tries to find a solution in conformity with the rules.

Backward-chaining is the reverse of forward-chaining. It is a chain which proceeds from a hypothesis back to the facts or data. It really involves working backwards from an as-

sumed hypothesis to see whether the given facts can make that possible and valid. It sees whether the data conforms to this particular hypothesis. Backward-chaining can result in an expert system coming up with more than one solution to a problem.

In practice, a solution by either of these two methods often involves working forward to a conclusion, testing that conclusion for reasonableness, working backward to make adjustments, working forward to another conclusion and so on until an acceptable end-solution is reached. This is how human beings solve complex problems. In the same manner, expert systems also through combined forward and backward chain methods reach the solution to a particular problem. This process is known as sideways-chaining.

Success Stories of AI

Both expert systems and conventional database programs involve retrieval and processing of information. However, there is a difference between the two. In medicine, for example, a database is useful to enumerate the symptoms of different illnesses. An expert system, on the other hand, helps to diagnose an illness, determine its causes and suggests a line of treatment. The database program may reorganize the data but does not reason about it. In expert systems, it is a human expert who has provided the information about specialized areas of knowledge such as say, water management, diseases of the blood, methods of teaching, etc. into a computer.

One of the first successful expert system was the *Mycin*. It was developed in the mid 1970s by Edward FEIGENBAUM and Edward SHORTLIFFE at the Stanford University. It is a medical diagnosis expert system. Its purpose is to help the doctors or physicians to diagnose and to prescribe medication to those suffering from infectious diseases of the blood. The problem is to first decide what bacteria has caused the

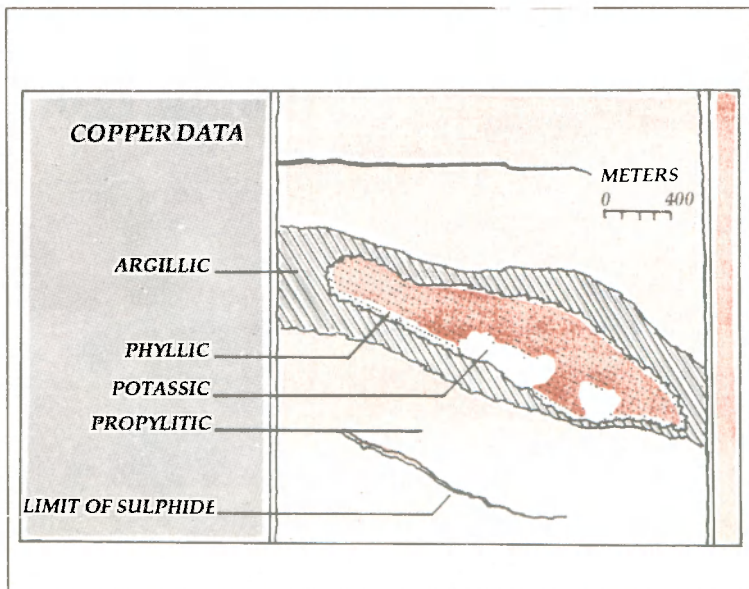


Medical diagnosis by an expert system

illness and then to prescribe an appropriate medicine which will be effective on the bacteria and at the same time have minimum side-effects on the patient. The physician first puts the general data including past medical history about the patient into the computer expert system. The *Mycin* then proposes a hypothesis and then goes ahead testing it. First it searches for clues to support that hypothesis. In case the expert system is unable to confirm or reject the hypothesis, then it can ask the physician additional questions and even

request for certain tests to be performed. The diagnosis is made by comparing the symptoms of the patient with those of the various possible infections the patient might be suffering from. This process goes on till a match is found. Later, taking into account previous medications prescribed and the patient's known responses to certain drugs, the expert system offers a prescription which should be just appropriate for the patient. This program has been very much appreciated as the user-computer interaction is in the form of a dialog. The program gives a reason as to why it is going to ask the next question. Even though the scope of *Mycin* remained restricted to a particular type of illness it paved the way for many similar programs in the field of medicine. *Puff* is an expert system for lung diseases. *Mycin* and *Puff* work just as well as any alert physician would.

Another expert system known as the *Prospector* is used in the field of geology and mineral prospecting. It is used to



Prospector helps locate mineral deposits

predict whether certain precious minerals and natural resources (including petroleum and gas) are found in particular regions under the earth. All information about which particular mineral ore would be found in what specific geological regions forms the domain of knowledge for this program. The data consists of results of analysis of rock samples. A number of other expert systems would be forthcoming and interestingly, some of these would even provide advice on matters related to tax, insurance and legal advice!



Program- ming Intelligence

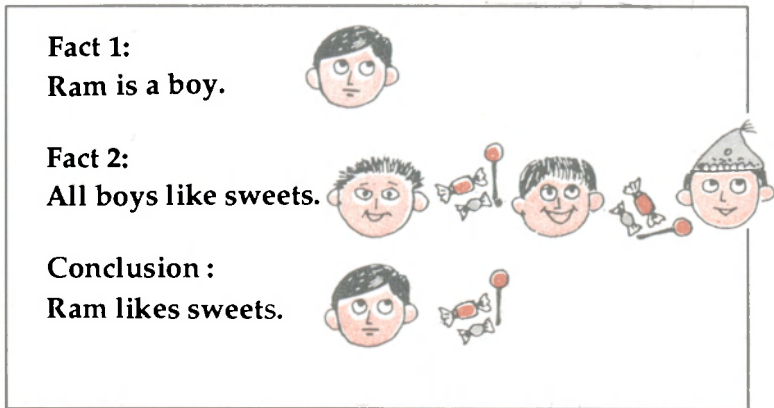


The first person in recorded history, to have defined logic was the Greek philosopher ARISTOTLE. His logic dealt with deriving the truth (T) or falsehood (F) of statements based on arguments. His methodology is referred to as the classical logic or syllogistic logic. Arguments based on classical logic are categorized as “syllogisms”.

Syllogism is a simple example of formal logic. The expression – ‘formal’ means that the logic is associated with the form of the statement rather than the meaning of the words. Algebra which one learns in school is also a kind of formal logic. In this form of logic, the procedure is important than the values assigned to the variables. In many ways syllogistic logic is formalization of commonsense. This type of logic continues to be the basis of legal arguments in our courts.

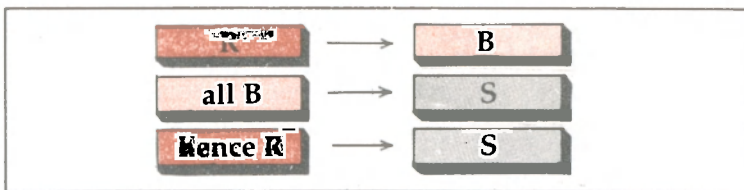
An argument based on syllogistic logic consists of premises (or facts) and a conclusion. The latter is derived from the premises.

Consider for example the following elementary argument relating to a child named Ram and his love for sweets:



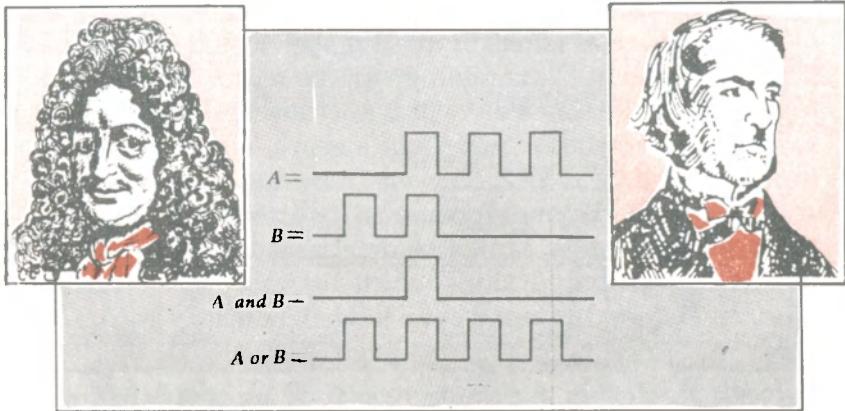
Symbolic logic

The above example can also be converted into a symbolic form where the symbol R stands for Ram, B for Boy and S for Sweets:



Symbolic logic was first introduced by Gottfried Leibnitz about 300 years ago but probably it was premature and did not catch on. Hundred and fifty years ago, George BOOLE (1815-1864) rediscovered symbolic logic. It dealt with abstraction of concepts into symbols and connecting them to form logical statements. To be able to reason we have to say things like "if 'this' is true then 'that' must be true" or "if both this and that are true then the third thing must also be true". To link the different facts together, one uses operators or connectives such as 'or', 'and', 'not', etc. A symbol is usually represented by a single letter or a string of characters but they

must commence with a letter. The validity of logical statements are best described by 'truth tables'. The connectives enable us to determine the truth or falsity of the logical statements.



Gottfried Leibnitz and George Boole: Pioneers of symbolic logic

Within symbolic logic there are two distinct but interrelated branches. These are 'propositional' logic which deals with the truth or falsehood of propositions made and 'predicate' logic which goes one step further by including the relationships between individual objects or classes of objects.

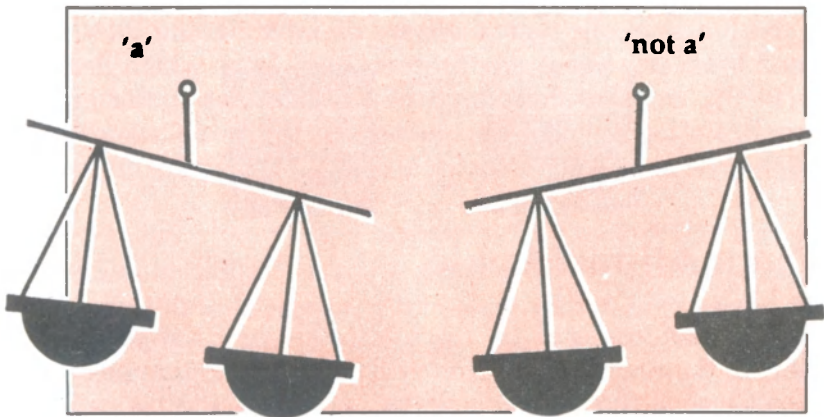
Propositional Logic.

A computer operates on the basis of binary signals. The electronic circuits inside a computer consists of components which are capable of assuming one of the two possible stable states. This basic concept has been used to represent longer numbers, characters, words and even statements about the real world in a digital format. Since propositional logic relates to whether a particular proposition is true or false, it could also be expressed in a digital format. Therefore, a similarity exists between the stable states in digital electronic circuits and propositional logic. Statements such as "It is raining

outside" or "The traffic light is red" can only be true or false and in a computer can be represented by a 'one' or 'zero' states respectively.

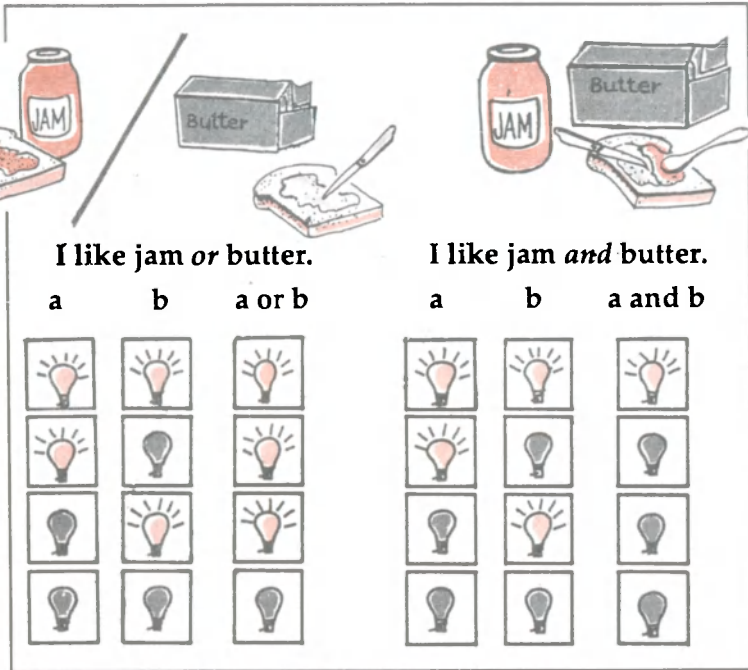
Propositional logic can be programmed into computers quite easily. The propositions can be represented by means of symbols such as letters of the alphabet, just as in algebra. If the proposition, "It is raining" can be represented by the letter 'a' then the statement, "It is not raining" can be represented by the symbol 'not a'. In algebra, we come across formulae such as, $X=Y+Z$. Now we know that $Y=2$ if, say, $X=5$ and $Z=3$. Similarly, in propositional logic the truth or falsity of the proposition 'a' cannot be determined by itself. Other related facts or propositions which we know to be true or false, are required before we can judge whether 'a' is true or not.

In simple algebra, the connectives used are operators such as $+$, $-$, \times and \div . Likewise, in propositional logic the connec-



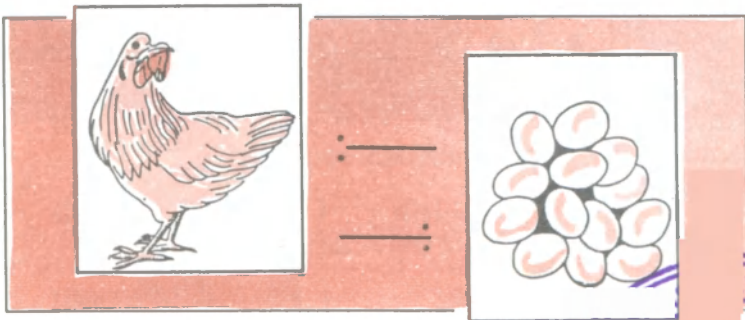
tives used between symbols are 'not', 'or', 'and', 'implies', 'equivalent', etc. The simplest connective is of course the 'not'. If 'a' is true then 'not a' is false. This can be represented by means of a truth table which is self explanatory.

The truth tables for the other connectives 'or' and 'and' are shown opposite. Here the symbol 'a' represents "I like jam"



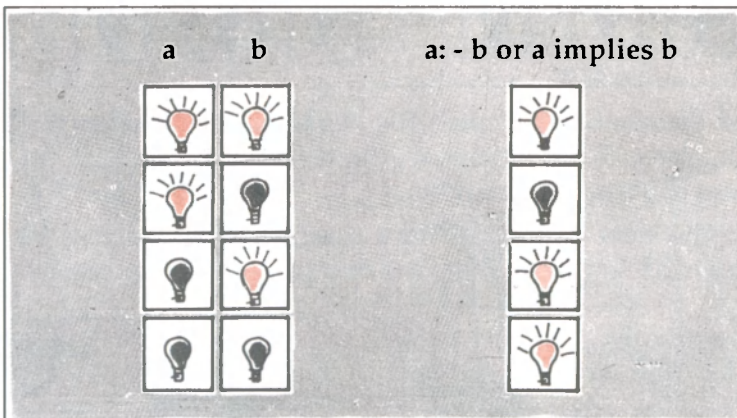
and the symbol 'b' represents "I like butter". The truth (T) and falsity (F) of a statement is represented by "on" and "off" states of a bulb respectively.

The next connective is a conditional connective also known as 'implies'. There are two parts to the statement "a



A Chicken implies egg or vice versa

implies b". These are the 'head' which comes before the symbol "implies" and the 'body' which comes after. The statement says that if the head is true then the body is also true. The statement can operate in the opposite direction also. If the body is true then the head must be true. Suppose we are out in the countryside for a walk (without an umbrella). The statement, "IF it rains THEN I will be wet" can be interpreted as "IF it rains" implies that "I will be wet". Symbolically this can be represented as "a implies b". The only way this statement can be false is when 'It rains' and 'I will not be wet'. Otherwise this statement is always true. The third statement is, 'It does not rain' and 'I will be wet' is true because you can get wet by falling into a pond even when it is not raining. The fourth statement is, 'It does not rain' and 'I will not be wet' and it is also true. This example which defines the "a implies b" connective can be expressed in a truth table shown below. The "on" and "off" states of a bulb represent the truth



(T) and falsity (F) of a statement respectively.

PROLOG and Predicate Logic

In order to generalize propositional logic, the concept of predicate logic was developed. It happens to be a convenient

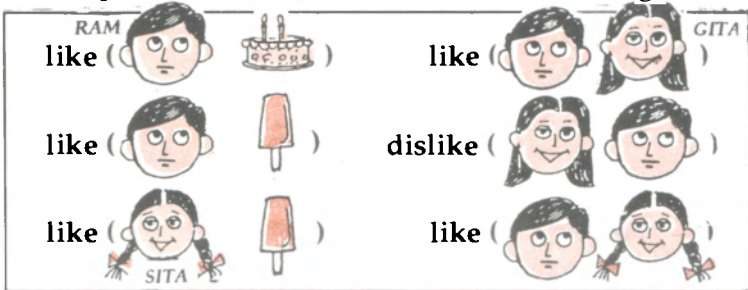
method of representing relationships within a knowledge-base. Predicate logic is the basis of programming the AI language, PROLOG, in which knowledge can be easily represented in the form of production rules.

In predicate logic, the items which are being discussed in a proposition are taken out of the statement and separated from those which explain the relationship between them. In the simple statement, "The boy plays tennis" there are two nouns, namely the *boy* and *tennis*. The relationship between these two items is through the verb *play*. The items are known as 'arguments' and the relationship as the 'predicate'. In predicate logic, the statements can be expressed in a very simple format:

Normal statement: 'The boy plays tennis.'
Predicate format: play (boy, tennis).

The relationship or the predicate is written first and then within brackets are the two arguments separated by a comma. According to the rules of PROLOG, the words commencing with capital letters are variables and those with lower case are constants or proper names. Also, it does not matter whether the facts which are being declared are true or not in real life. The predicate statement: *eat (tiger, grass)*, is a valid predicate statement but then it is obviously false in the real world.

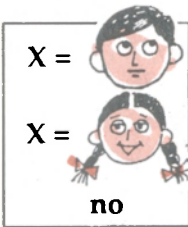
Next, let us collect some facts for our knowledge-base. As an example these relate to who likes what. Having entered



these pieces of information into the computer memory we can query and ask for information by typing the symbol `?-`. PROLOG will now answer logical questions, such as, `?- like (ram,sita)`. In plain English this question is, "Does Ram like Sita?". The PROLOG program will produce an answer 'yes'. This literally means that the PROLOG program has searched all the entries, commencing from the top of the knowledge-base, working downwards and has found a match and so gives the result 'yes'.

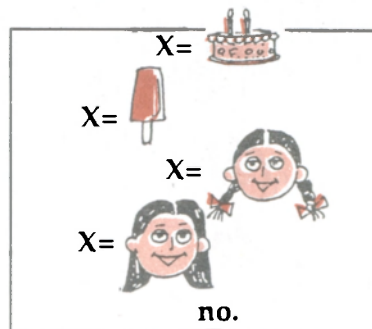
For the query; `?- like (sita,cakes)`, the answer is 'no' which means that the statement cannot be proved from the facts present in the knowledge-base.

The greatest improvement that predicate logic has over propositional logic is in the use of variables. These are expressed by symbols which always commence with a capital letter. A query could be made, such as; `?- like (X,icecream)`, which in plain English is, "Is there any one, call him X, who likes icecream?" The PROLOG program commences the search from the top of the knowledge-base and comes to the first match which is: $X = \text{ram}$. If a semicolon is now pressed on the keyboard, the search continues for the next person who likes icecream.



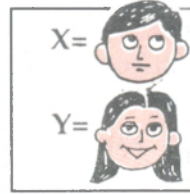
The last 'no' means that there are no more facts within the knowledge-base to satisfy this query and so the search has stopped.

Once we get used to this round about way of asking questions, we can ask even more complex questions such as, `?- like (ram, X)`. This is asking the question, "What or whom, call it X, does Ram like?" PROLOG will promptly reply:



Clearly, 'no' here means that from within the knowledge-base there are no more items which Ram likes!

Another question one could ask is: *?-likes (X,Y),dislikes (Y,X)*, which means that,"Does anyone called X like anyone else called Y and this second person Y does not like X?" The comma between the two parts on the right hand side indicates the conjunction 'and'. PROLOG will provide the following reply:



Interestingly, predicate statements can be extended to combine propositions.

Proposition: "IF Ram likes icecream THEN Ram buys icecream."

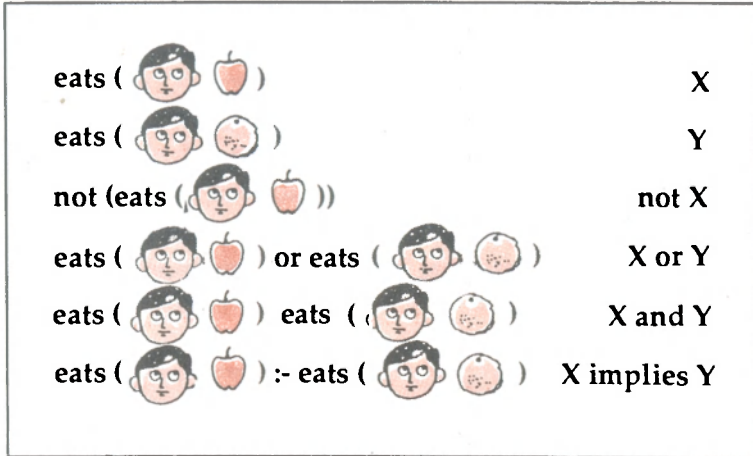
Predicate logic

Step 1: like (ram, icecream).
Step 2: like (ram, icecream) implies buy (ram, icecream).

Conclusion: buy (ram,icecream).

Each part of a proposition that is complete (i.e. cannot be broken down) is called an 'atom'. A predicate and its arguments constitute an atom. For example, *eat (ram,apple)* is an atom. A 'literal' is an atom with or without a 'not' in front of it. Both *eat (ram,apple)* and *not (eat (ram,apple))* are literals.

They are known as positive and negative literals respectively. Connectives are used with predicates as shown below.



Now, how does PROLOG go about finding answers? Take the case of two simple statements such as, *like (ram,icecream)* and *like (sita,icecream)*. In the same knowledge-base a rule can be introduced, that is, *like (X,Y) :- like (X,icecream), like (Y,icecream)*. For asking the question, "Do Ram and Sita like icecream?", one keys into the computer :

?- like (ram,icecream), like (sita,icecream)

Thus predicates, conjunctions, rules and functions can be used to manipulate predicate logic and the AI program PROLOG. The question asked by us had one goal whereas the question keyed into the computer has two goals. The program searches the knowledge-base from top to bottom to see whether the first goal, namely, *like (ram,icecream)* can be found or satisfied. The moment it is found, the program begins a search for the second goal. With these two goals being satisfied the main goal of the original question is answered 'yes'. If a goal fails then the entire conjecture has failed and the answer is 'no'. This technique is known as 'backtracking' and is a powerful one in PROLOG.

Reaching Inference

The first thing to do on receiving an input is to convert it from English into an equivalent predicate logic. Putting it in the correct format is very important. The choice of the predicate, the function and the constants have to be carefully made in AI applications.

If the input sentence is a question then the answer is somewhere in the knowledge-base but it may not be explicit. We may have to infer the answer from the facts. This is made clear in the following example:

Fact 1: Ram has fever.
Fact 2: Infection brings on fever.
Conclusion: Therefore Ram has an infection.

Inference and deduction are the strong points of predicate logic. The initial facts which lead to the deduction are known as 'axioms'. In algebra and geometry we are used to axioms which are explicitly stated statements. However, in AI the information received by the sensory organs constitutes the axioms for the discussion which is current at that time. Therefore, unlike in mathematics, in AI our set of axioms are not stable.

Deduction is logical reasoning in which conclusions are drawn from knowledge or facts by logic. In order to deduce new facts from the axioms we make use of the rules for inference. The most important rule is the 'modus ponens' which is as follows:

Given: If A then B
Given: A is true
Deduction: ∴ B is true

The above line of argument can even be expressed as:

A, A → B; therefore B

We are reasonably familiar with this deduction procedure as we use it in our every day life.

Another rule with which we are familiar is that of 'universal instantiation'. We use it quite often without being aware of it. For example, if something is true for Delhi then it must be true for Karol Bagh.

Yet another rule for inference is called 'abduction'. It is the process of reasoning backwards from a true conclusion to the premises that may have caused the conclusion. Abduction generates explanations as the following example demonstrates.

Given: If A then B
Given: B is true
Inference: A is true

This is not a legal inference as one could easily arrive at a wrong A (in the above example). This is demonstrated in an extreme example below:

If X has cancer of the lungs then he will be short of breath.
Y is known to be short of breath.
Therefore Y is suffering from cancer.

What has happened is that given the rule and the fact that Y is suffering from short breath we have jumped to the conclusion that Y is suffering from cancer. The fact that Y is short of breath has been explained away by assuming lung cancer. This conclusion is a drastic one and hopefully a wrong one (at least for the sake of the doctor's reputation). Though abduction could lead to wrong conclusions, the method is widely followed in many fields including medicine and economics. Abduction is the process of providing an explanation. Experienced medical doctors use this method with judicious caution and are able to make satisfactory diagnosis.

Another kind of inference is 'induction'. It is the process of generalizing from a few specific cases. Suppose we meet a

boy from a particular school and he plays hockey. We meet another boy from the same school and he too plays hockey. Then we could infer that all boys of that school play hockey. This process of reasoning is known as induction and it is accepted just as we do abduction, with a certain amount of restraint and control.

In the case of 'intuition' an inference is arrived at by unconsciously recognizing a pattern. It is also known as the *sixth sense*. It is usually correct if one is very alert and aware of what is happening around us. If someone chokes while eating a banana, we immediately slap him on the back, having witnessed a similar situation long ago.

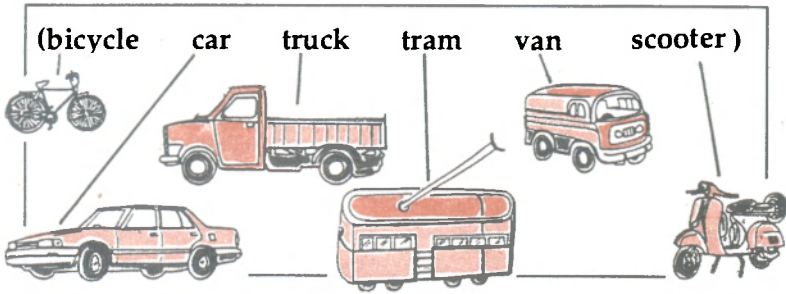
Inference by 'analogy' is arriving at a conclusion based on a similarity to another situation. This is a common procedure one adopts in providing explanations to phenomena in scientific subjects. A similarity is drawn between a new phenomenon and another which already has an explanation. Based on this analogy we suggest or infer an explanation for the new phenomenon.

LISP

LISP or List Processing was a programming language created by John McCARTHY. It is a functional language that manipulates both symbols as well as logical statements. It is a simple language and has been extensively used, particularly in the U.S.A. for developing expert systems. The LISP rules for manipulation of information are easy to understand and to learn.

LISP was developed mainly to process lists. To put items in the form of a list is an important part of its structure.

For example, in LISP the list of all possible road transport vehicles one can think of can be written in a line with brackets around them. A gap is left between neighbouring elements.



There are three ways of assigning meanings to symbols in LISP. First, a symbol can represent a fact, an object, a piece of data, or it can represent an action whose result could once again be data or knowledge. Second, a symbol can be a variable which has a value or a meaning. The analogy here is that symbols which represent variables are similar to the unknowns x and y we use in algebra where the values are determined after applying the rules of algebra. The same variable has different values or meanings in different problems. Finally, a symbol can represent functions. The above list can be rewritten as: $(a1\ a2\ a3\ a4\ a5\ a6)$. Each element of a list can be either an atom or another list. A complex list would look like: $(a1\ (a2\ a3)\ a4)$. This is a list consisting of $a1$ which happens to be an atom, $(a2\ a3)$ is a list of two atoms and $a4$ is another atom. The above list has three components. On the other hand, $((a1)\ (a2\ a3)\ a4)$ is a list where $a1$ is no longer an atom but a list. This situation of having lists within lists is known as 'deep-nesting' and this makes it difficult to estimate how many elements are actually contained in a list.

Both programs as well as data are represented as lists. So what actually happens in LISP is that the programs manipulate lists. Also, since all programs are in the form of lists they can manipulate the programs. This is a very useful criterion in AI as one of its characteristics is to design programs which learn from experience. (Learning here means modifying the knowledge-base and the reasoning algorithms to suit new information)

However, since both programs and data have similar list structures, every expression has to be evaluated. In LISP the format of a list would be:

- * The first symbol inside the left hand bracket is the name of the function. In case it has not been defined or given a name then the program indicates a syntax error.

- * All numbers are considered as constants and retain their values.

- * All other symbols or atoms are considered variables and are to be evaluated and their values determined.

There are a large number of functions used in LISP. Some of the commonly used LISP functions are quote, setq, cdr, car, cons, append, predicate, cond and defun.

Quote: This function takes an argument (which may be a symbol, a number or a list) and returns it unaltered. In literature it is expressed in two alternate forms as illustrated below:

(Quote Bicycle) or (Bicycle) returns Bicycle
(Quote 1 2 3) or (1 2 3) returns 1 2 3

This function might appear to be very elementary but has its uses. Its real importance is that it prevents any evaluation of the argument from taking place. Whatever follows the 'quote' is returned as it is verbatim.

Setq: This is another very common system function. Its effect is to bind the second symbol and the third in the list. The function has two arguments. The first element is the function setq. The second is a symbol variable and the third

(setq letters' (a1 b1 c1)) returns (a1 b1 c1)
(setq one 1) returns 1

can be any symbolic expression. `Setq` returns the value of the second argument to the variable.

The 'quote' mark in the first example ensures that the value of the variable, 'letter', is the list (a1 b1 c1). In the second example the third element is a constant 1 and LISP returns the value 1 to the variable 'one'. If one thinks carefully the variables 'letters' and 'one' in the example above should also have a quote in front of them to indicate no operation or evaluation. However, convention in LISP has it that the *q* after the word *set* (in *setq*) represents a quote and as such there is no need to mention it specifically.

Similarly, in example: (*setq birds '(crow sparrow ostrich)*), the variable 'birds' is a list. So if the word 'birds' is keyed into the computer it returns the list as: (*crow sparrow ostrich*).

Cdr: This function is used to remove elements from a list. The function returns a list minus its first element or atom.

(cdr' (a1 b1 c1)) returns (b1 c1)

Here *cdr* has been applied to the list (a1 b1 c1). LISP returns the same list with the first element a1 removed. The value of the list has been changed. It has been made one element shorter and therefore, is different from the original list.

Considering the same example: (*setq birds (cdr birds)*), LISP starts with the innermost part of the statement, i.e. (*cdr birds*). Since there is no quote, the *cdr* operates on the variable 'birds' and so it returns the list corresponding to 'birds' but with *crow* eliminated. The new 'birds' list therefore, becomes (*sparrow ostrich*)

Car: This function consists of an argument and a non-empty list. It returns as its value the first element from the list.

(car' (a1 b1 c1)) returns a1
(car' ((a1) (b1) (c1))) returns (a1)

The usefulness of the quote is seen here in the example. It has prevented the argument of *car* from being evaluated. If the quote were removed then LISP would try to evaluate $(a1\ b1\ c1)$. Since this is a list it expects *a1* to be the name of a function. If *a1* is not a function then an error in syntax is indicated by the computer.

Cons. LISP functions are useful in adding new elements or lists to existing lists. *Cons* enables an element to be added. It takes two arguments, the first may be a symbol, a list or a number, but the second must be a list. *Cons* inserts its first argument as the new first element of its second argument and returns the new list.

```
(cons 'a'(b c)) returns (a b c)
(cons '(a)'((b) (c))) returns ((a) (b) (c))
```

Cons permits one to construct new predicate function from existing ones.

Again going back to the previous example; $(setq\ birds\ (cons\ 'parrot\ birds))$, *cons* returns as the new definition of 'birds' as: $(parrot\ crow\ sparrow\ ostrich)$.

Append: This function joins a new list on to the end of an existing list. Here, $(setq\ birds\ (append\ birds\ ' (eagle\ myna)))$ would be returned as a new list, i.e., $(parrot\ crow\ sparrow\ ostrich\ eagle\ myna)$

Predicate: LISP has a number of predicate functions. Some of them are *atom*, *zerop* and *greaterp*. These return a value true (T) or false (F). The function *atom* takes one argument of any type and returns T if the argument is a symbol or a number and a F if the argument is a list.

```
(ATOM '4) returns T
(ATOM 'A1) T
(ATOM '(a1 b1 c1)) F
```

The function *zerop* is a simple function used with a numeric argument. It returns a T if the number is 0 and F if not.

(ZEROP 14) returns F
(ZEROP 0) T

The function *greaterp* takes a list of numbers and returns T if the numbers are in descending order, a F otherwise.

(GREATERP '(10 9 8 7)) returns T
(GREATERP '(9 3 6 5)) F

Cond: LISP enables one to write a conditional clause with the help of the function *cond*.

(cond ((condition 1) (action1 action2 action3)
(condition 2) (action4 action5 action6)
(condition 3) (action7 action8 action9)))

LISP evaluates the first condition and if it is satisfied (T) then action 1, action 2, action 3 are executed. If the first condition turns out to be false (F) then the second condition is evaluated and so on. The last condition is designed so as to be always true and the corresponding actions are carried out. The conditions and the actions can be any symbolic expressions.

Defun: Defun is a LISP function which can be used to create new functions.

(DEFUN function_name ((argument_list) (function body)))

This returns as its value the name of the function being defined.

This is only a brief introduction to LISP but it shows how useful it can be in AI. Interestingly, in writing LISP programs no distinction is made between the structure of data and programs. Moreover, this language is capable of manipulating symbols and structures rather than just numerical calculations.



User-Friendly Systems



Ever since the world's first general purpose computer was put into operation, an anxiety was expressed in many quarters, that one day such machines would become far too complex and difficult for a casual user to operate. To begin with, he would have to make special efforts to familiarize himself with the intricacies of the new computer. This would include acquiring some knowledge regarding the details of its hardware, learning some new programming languages or commands used to communicate with the computer and above all matching the problem solving capabilities of the machines with his actual requirement. This trend which continued for over three decades, made it necessary for users to take pains in matching their skills with those of the computers. However, with the advent of AI this trend is being reversed.

With the popularization of knowledge based systems, the computer designers began to think in terms of introducing new interfaces between machines and their users. These would enable the two to interact on a more familiar basis. The systems which

emerged were labelled “user-friendly” and they are currently proving to be very popular.

Natural Languages

Advances in knowledge based techniques have led to many conceptual changes in computers. These have been redesigned so that even ‘computer illiterate’ users can use such advanced machines. This in turn has stimulated academic research in subjects such as natural language processing, speech recognition, synthetic speech generation, vision systems, etc.

With knowledge-base being added, the status of a computer has risen to become one of a ‘partner’ with its user. This

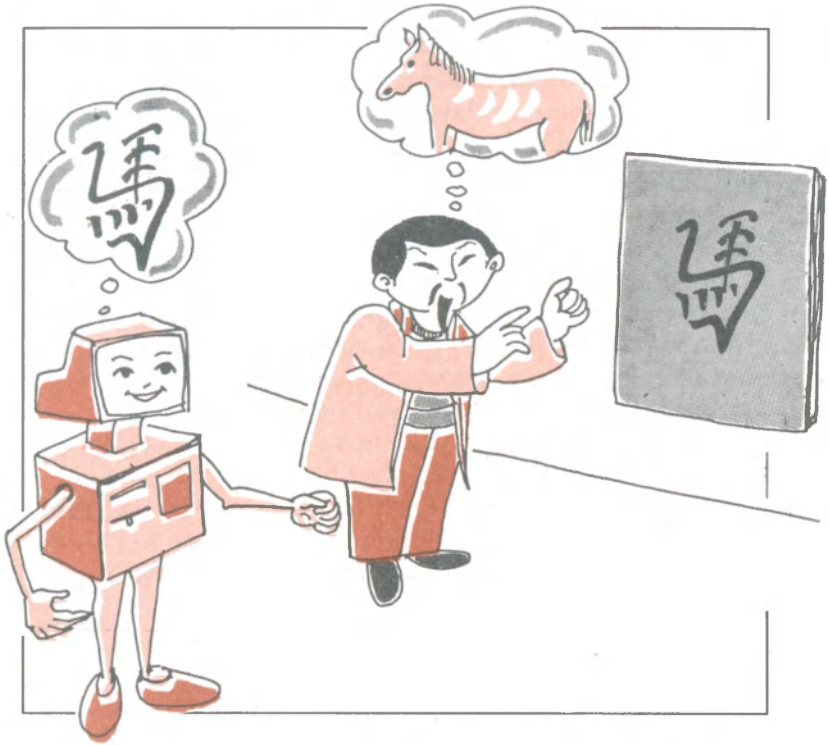


Man-machine interaction

is not surprising, since the knowledge-base is basically derived from human experts. Both the human user and the computer can now initiate interaction with one another by asking questions, setting goals or generally by responding to each other's queries. Such situations are generally known as possessing mixed initiative. It is, therefore, even more important that the dialogue between the two should be in a language which the user is already familiar with. The problem now reduces to one of making the computers familiar with natural languages! They have to learn how to analyze sentences and to extract meaning from them.

A popular way of introducing instructions and data into a computer has been through keyboards. At first sight the common computer languages, COBOL and BASIC appear similar to spoken English. However, the user has to learn them and become familiar with their rules and restrictions. On the other hand, a natural language, say English, is learnt at school. When we hear or read a simple sentence in English we understand it straight away as we are familiar with its vocabulary and rules of grammar. If it happens to be a long or a complicated sentence with difficult words thrown in, then we mentally analyze the sentence into its different grammatical clauses and then try to understand the message part by part. 'Natural Language Understanding' is a study of the different processes by which a computer tries to comprehend the instructions given in ordinary English.

The core of such a 'natural language processor' is the '*Parser*'. It reads each sentence and then proceeds to analyze it. This process can be divided into three tasks viz. dividing the signal into its acoustic-phonetic, morphological-syntactic and semantic-pragmatic aspects. The first part comes into operation only if the input consists of spoken words. It converts the utterances or the sound waves into words. The second establishes the syntactic form (grammatical arrangement) of the sentence. The third tries to extract the meaning from these analyzed patterns. It takes into account the com-



Computer programs are formal (syntactic); Human minds have mental associations (semantic)

mon usage of words in a language. This entire process is known as 'syntactic analysis' or 'parsing' the sentence. Subsequently, it can be represented in a tree type of a format based on the parsing rules which govern the language. Extracting the meaning is known as 'semantic analysis'. The syntactic analysis or the parse indicates whether a word is a noun, a verb, an adjective, a determiner, etc. The meaning becomes far easier to extract or derive after parsing. This process can be extended to longer and complex sentences as well.

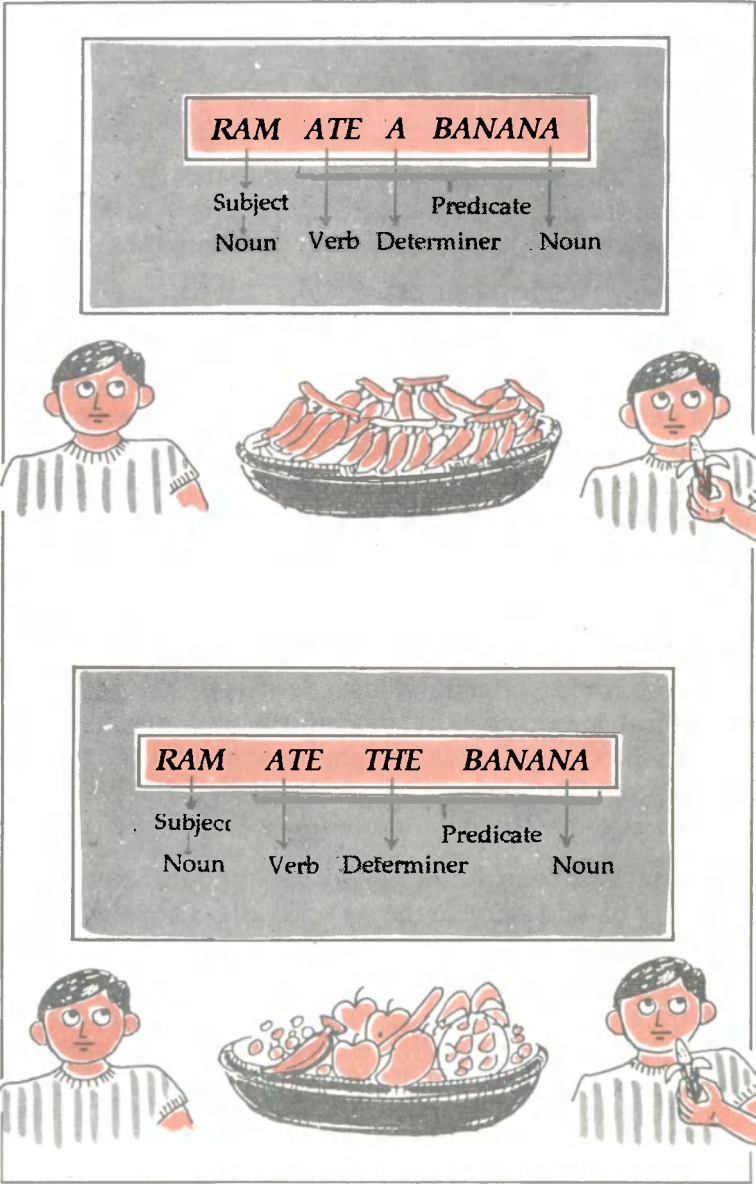
In our normal conversation, we often use ungrammatical structures as the following dialogue would reveal:

Doctor: When did you experience pain this week?
Patient: On Tuesday and Thursday.
Doctor: Last week ?
Patient: Wednesday.

The doctor's second question is not grammatically correct but such expressions are normally accepted without much fuss. No one bothers about grammar in such situations. The listener responds as though he has heard a perfectly correct sentence. Humans accept such mistakes or errors when they already have a gist of the discussion based on the earlier part of the conversation. If the doctor's second question had come after a delay of five minutes or so, the patient may not have been able to understand the question at all. One of the most difficult aspects of designing a natural language processor is the complexity and flexibility of the human languages. What happens when a machine, which has been trained to accept statements in accordance with the correct rules and format, has to interpret the above conversation? It may simply not understand it. Many of the first generation natural language processors could understand just a few types of sentences of the natural language. The machine would understand a spoken sentence only if it was grammatically correct. If the rules of grammar were not constrained, then, the procedure for analysis or parsing would become extremely lengthy.

Parsing had been investigated by linguists quite independently of, and prior to the AI scientists. They developed tree diagrams for parsing sentences. For example, in the sentence: *Ram ate a banana*; 'Ram' constitutes a proper noun and is the subject and 'ate a banana' constitutes the predicate. The linguists would parse the sentence with the help of a tree diagram.

Many AI scientists believe that the parse trees are important intermediaries between a sentence and the mental (or internal) representation of the concept it is supposed to create. If in the above sentence, the determiner 'a' were to be



Parsing sentences

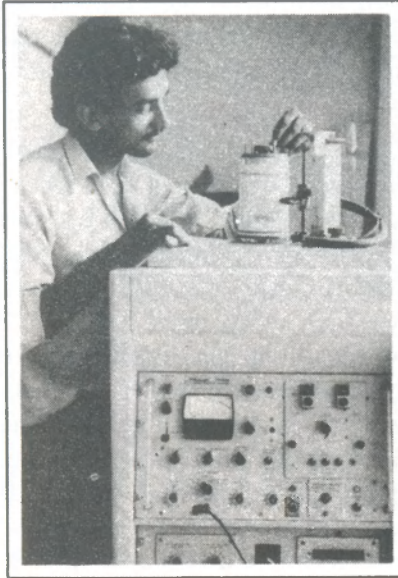
substituted by *'the'*, then the meaning or the impression created by the sentence is different. *'The banana'* makes it a very specific banana where as *'a banana'* refers to any banana lying in the basket.

Similar parse trees could be prepared for multi-clause sentences as well. These tree structures, in turn, have led to the development of networks consisting of nodes or states with arcs which tell us what is the next parse (noun, verb, etc.) The transition networks operate by working through a sentence from left to right. For each word there are only a limited number of options of words which can grammatically follow it. For example, consider the sentence, *The small fat man ate a cake*. It consists of two parts viz. the noun phrase *'the small fat man'* and the verb phrase *'ate a cake'*. The adjective *'small'* can be followed by another adjective such as *'fat'* or by a noun like *'man'* but cannot be followed by a verb. Rules for each of these phrase structures can be noted and used in analyzing a sentence.

Each phrase structure can be written as a route from the beginning to the end of a grammatical item. The network with possible routes from a given starting point to a completed noun or a verb phrase is described as a transition network.

Automatic speech recognition

Automatic speech recognition has fascinated engineers, scientists and linguists alike. It has kept them interested for the past half a century, first in trying to understand the production and perception of speech sounds in human beings and then to simulate the process using electronic systems. They have designed machines called sonographs which can analyze speech signals. A careful study of the machine's output called a *'sonogram'*, helps to recognize the spoken words and even extract their meaning. This continues to be an exciting challenge and vast sums of money are



A sonograph used to analyze speech

being spent on its research. Speaking into machines would appear to be the most natural and efficient means of man-machine communication.

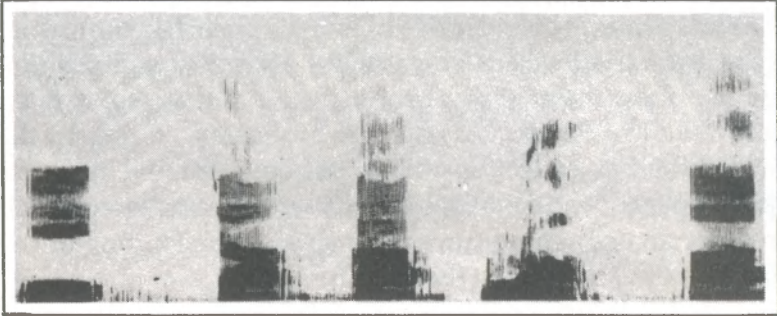
Speech is a natural means for human communication particularly, in situations where both the eyes as well as the hands are occupied or busy (eg. a pilot of an aeroplane). Automatic speech recognition systems are electronic devices which receive the spoken words from human speakers and produce an output which consists of coded digital signals ready to be fed into an applica-

tion unit. These signals correspond to what the Automatic Speech Recognizer (ASR) has recognized of the speech message. The application unit takes appropriate action on receiving these signals from the ASR. They can be used to initiate a specific machine activity. For example, a robot may be given oral instructions to walk ahead, to turn and so on. In the commercially available ASRs, no comprehension of the speech sounds as such is undertaken. The output signals strictly correspond to the recognized words.

Human speech is characterized by three types of sounds viz. *voiced* sounds where the acoustic source is the vibrating vocal cords in the larynx, *fricative* sounds (like *s*, or *sh*) which are essentially white noise. They are generated by forcing air through narrow constrictions in the mouth. *Plosive* sounds are produced by first allowing an air pressure to be build up

behind a closure and then released (as in /p/ or /b/ or /k/ etc.).

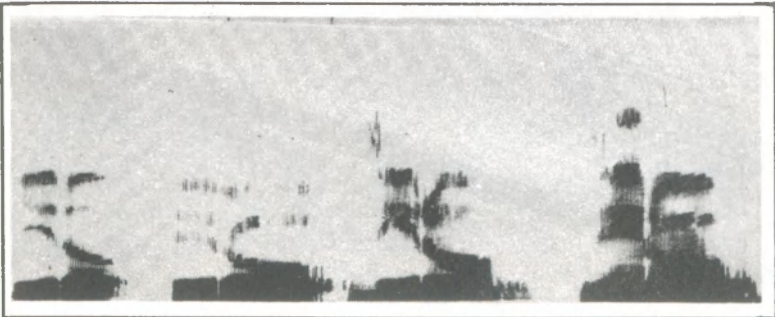
Linguistic studies of speech sounds suggests that there are basic discrete sound segments in any language and these are



Sonogram of phonemes

strung together to form words while speaking. These segments are known as 'phonemes'. Most spoken languages possess, on an average, about 40 phonemes. Each one of these phonemes can be characterized by a set of unique properties.

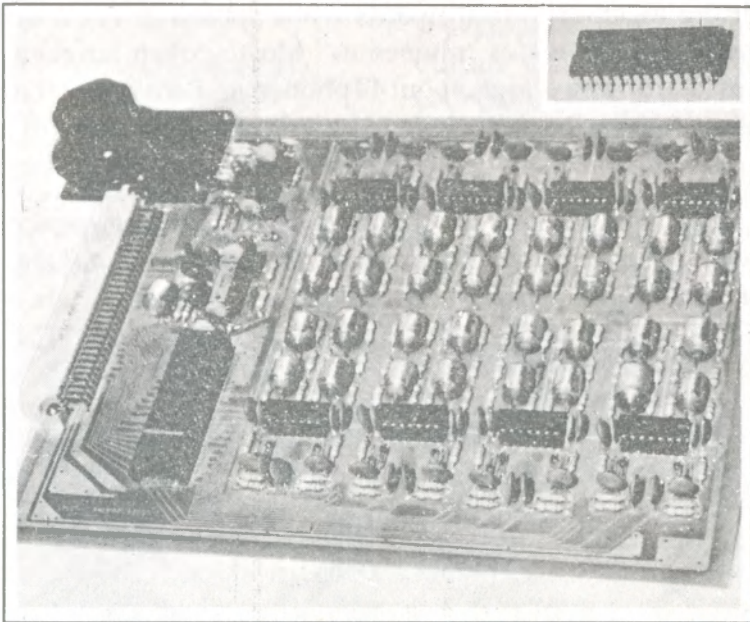
However, recognition of phonemes by computer systems is yet not a successful proposition. The contextual influence of continuous speech on individual phonemes are so



Sonograms of the word 'zero' as spoken by four speakers

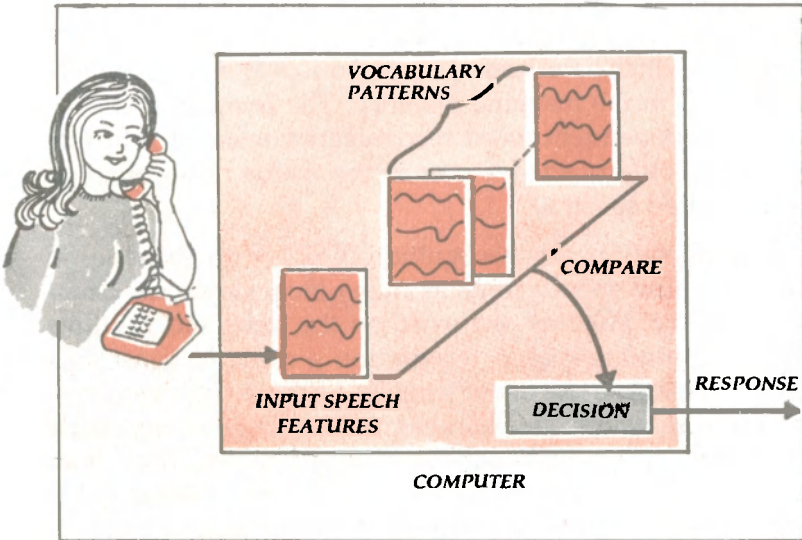
considerable that it is difficult to get machines to recognize them with certainty. There are enormous variations in the styles of speaking by men and women. A word spoken by the same speaker on different occasions has minor variations. The same word spoken by different speakers also has variations depending on individual accent or dialect. In the former case, the variations arise on account of context and position of the phoneme within a word. In the latter case, it is on account of factors such as the size of the vocal tract and the individualistic style of speaking. The development of a complete speaker independent speech recognition system will require an even more comprehensive understanding of the sources of variability in the speech signals.

Isolated word recognizers with a pause of at least 300 milliseconds between successive words are available com-



Electronic band-pass filters used in speech analysis;
A VLSI chip equivalent to the above circuit (inset)

mercially. These machines typically deal with the recognition of a small set of isolated words by a known speaker. Most systems produce a running spectrum with the help of a bank of **band-pass filters**. The spectra of the words are examined every 20 milliseconds. The energy content in the different frequency bands are compared with those in corresponding bands already available and stored in the memory. The template which has the best match with the input is selected as the recognized word. Such machines operate very well when the vocabulary is between 20 and 200 acoustically distinct words. A technique known as 'dynamic programming' is extensively used in the isolated word recog-



Pattern recognition approach to isolated word recognition systems. It is a path finding algorithm which attempts to seek the best alignment between two similar patterns. Time alignment is necessary as the duration of the spoken words, even if by the same speaker, can vary when spoken on different occasions. If the system caters to larger vocabularies then not only the time required to produce a match is propor-

tionately longer but also the training period for each new speaker would be time-consuming.

Continuous speech signals can be reduced to a sequence of isolated words by intentionally introducing pauses between words. In this case, the individual words will be recognized as such. This method is adequate to get the spoken word to appear on the screen or where instructions consisting of single word commands are to be recognized. The speaker has to produce samples of the spoken words before he can use the system. This is known as the 'training phase'. Each new user has to have templates for his speech prepared in advance before he can use the machine. However, speaker independent systems require no such individual training schemes. Such systems usually have templates which have been averaged out over many speakers and stored in the machine memory. The templates in such speaker-independent word recognizers consist of information based on acoustic features of the words rather than on their detailed spectra.

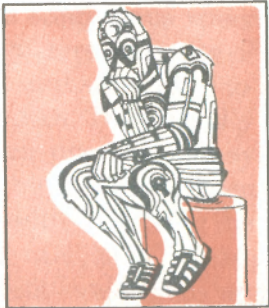
Continuous Speech Recognition (CSR) is much wider in scope. Natural speech contains many inconsistencies. Coarticulation (i.e. effect of one word on the next) and rhythms make continuous speech recognition a difficult task. Such recognizers make use of our linguistic knowledge (both syntactic and semantic) to characterize the range of acceptable sentences. They usually consist of large vocabularies, both in their phonetic and syllabic representations. Extensive research has already been undertaken in this area of technology, with CSR being the ultimate goal. Systems claiming 1% error rate have already been produced. However, independent evaluations mention error rates between 12% and 0.5% which depends on the complexity and cost of the systems.

While speech signals carry linguistic information regarding the message to be conveyed, they also possess extralinguistic information about aspects such as the speaker's

identity, dialect, his psychological and physiological states and also the prevailing environmental conditions such as noise, room acoustics, etc. In order to develop high grade speech recognition systems one has to learn how to extract the message-bearing components from the signal while discarding the rest. Understandably, considerable fundamental research in speech science is needed before artificial speech recognition systems can approach anywhere near human performance.



An Overview



Artificial Intelligence is barely 40 years old. However, these four decades have witnessed phenomenal advances in this field which tempts one to say that machines would soon perhaps be able to think and perform intellectual tasks as human beings do. Ideally, the objective of AI is to design machines or non-biological systems with which man should be able to interact exactly as he does with another human being. An interesting example from ancient Greece is the *Oracle of Delphi*. It was a statue which listened to the prayers of pilgrims and often spoke back, providing the listener with some wholesome advice! In actual fact, there was a hollow duct leading from the mouth to its back where a priestess sat. She remained out of sight, heard the prayers and answered through this tube, astonishing the visitor! Today, thanks to AI, we have machines which speak back to their operators. You walk into some banks in Japan or U.S.A. and speak your account number into a machine and it replies in a synthetic speech, the exact amount of money you have in your account. Also, electro-mechanical robots are now available which perform difficult tasks such as welding.

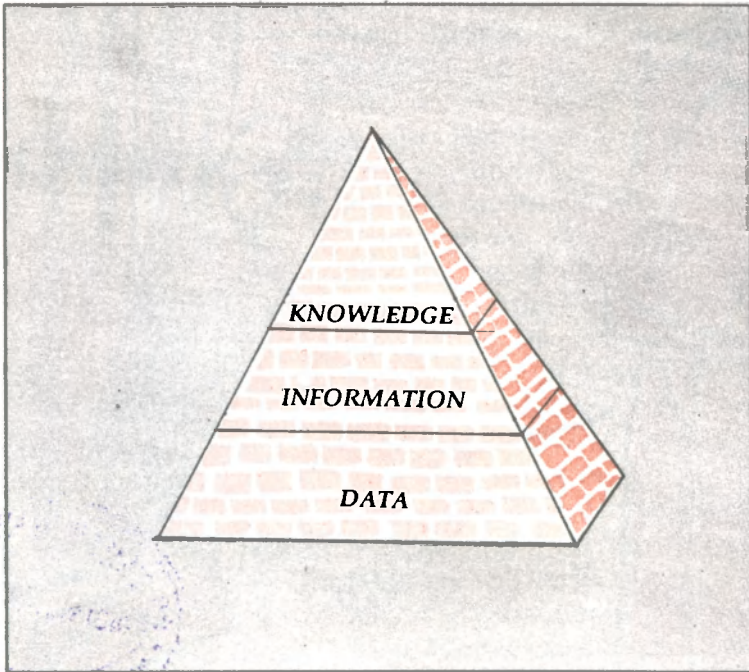


The modern 'Oracle of Delphi'

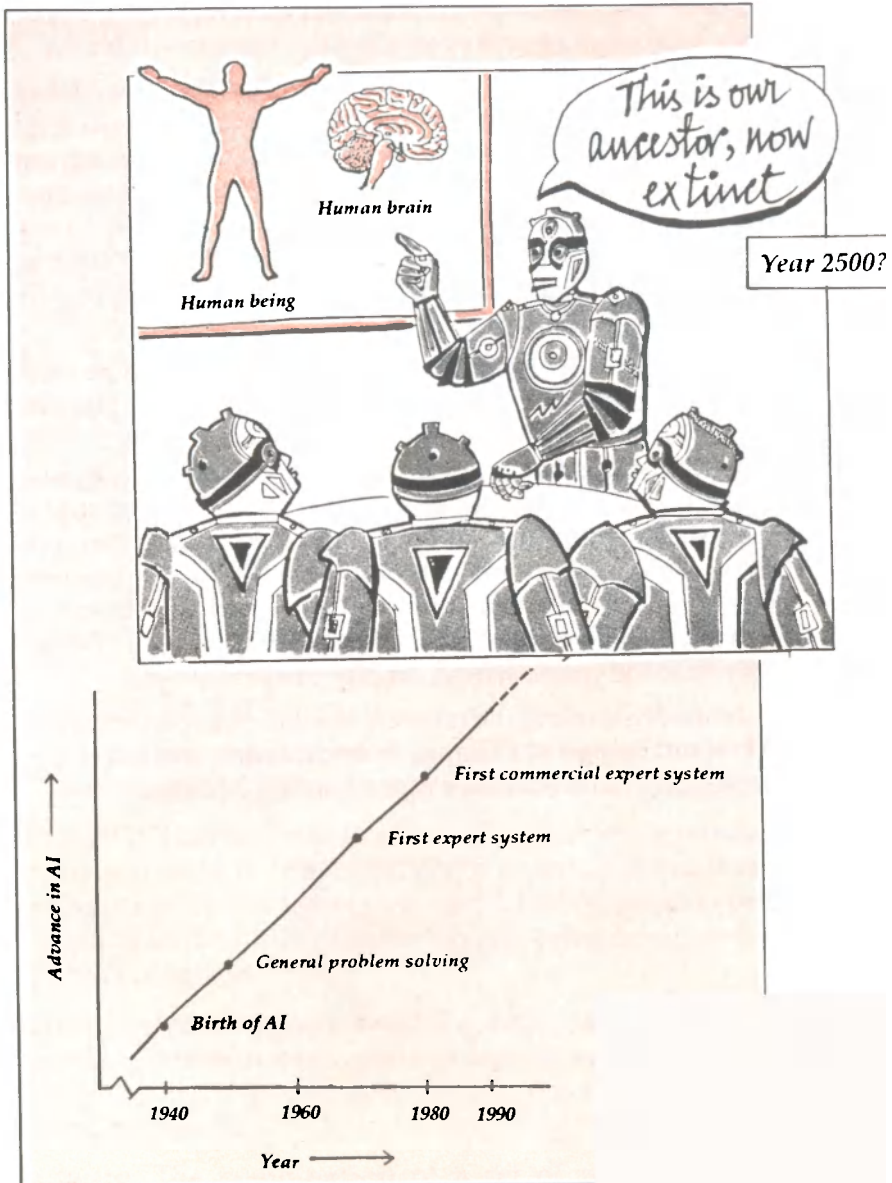
Such robots are currently replacing human labour in the Japanese automobile industry.

Intense research on how the human mind functions is essential before one asks whether its electronic equivalent can be constructed. Though one may not be able to describe the mind in terms of a set of rules and frames, yet there is a possibility that the current research on parallel processing and neural networks might lead AI scientists to simulate the fabric and process of the mind.

Here one comes across a question as to what constitutes knowledge. This is a very elusive concept and even more difficult to quantify. Nevertheless, this aspect has been the single greatest preoccupation of AI research. At present, its



Hierarchy of knowledge



Trends in AI with a peep into its future

scope is confined to the collection of data regarding objects, their relationships and events within restricted domains.

Expert systems, the best known manifestations of AI, have today gained immense credibility and acceptance in many professional fields. Several complex programs are commercially available to help analyze data fed into the machine, diagnose diseases and perform various tasks which have so far required a high level of human expertise. Further attempts are also being made to widen the scope of each such system to cover wider domains.

AI is rapidly making an impact on our lives. New machines taking advantage of AI have become irreplaceable in commerce, industry and research. What was science fiction a decade ago is fast turning into reality. Nonetheless, man-made machines are yet not capable of mimicing in entirety the sophistication and subtlety of the human mind. So far all the knowledge-base and data which are fed into the machines have been carefully planned in advance and various exigencies taken into account. In a sharp contrast, a human being is born into the world with an ability to learn.

Notwithstanding the inherent limitations, it is certain that AI would bring a sea-change in information and knowledge processing the world over in the coming decades.



Glossary

Band-pass filters: An arrangement of electronic components which enable only signals of a specific frequency to pass through them. At the same time, they weaken the rest of the unwanted signals.

Binary: A number system with a base of two. It involves only two digits : 1 and 0. In a binary code, any data is encoded through the use of these two digits called 'bits'.

Chip: Integration of thousands of electronic devices and components on a single, rectangular slice of silicon. It is the key to the microelectronic revolution in computers.

Electronic toggle: A circuit having elements which are capable of assuming either one of the two stable states, i.e. ON or OFF states, at a given time. This device is synonymous with a 'flip flop' circuit.

Phonemes: A set of basic sound signals used in a particular spoken language. A change in a phoneme may alter the meaning of the word.

Program : A set of coded instructions that direct a computer to follow a given logical sequence to perform a particular task.

PROLOG: Short for Programming in Logic. A programming language used in the development of expert systems. It can be easily represented in the form of production rules. Expert systems based on PROLOG are currently being developed in Japan and England.

Relay: An electromagnetic switching device having multiple electrical contacts. It is energized by passing electrical current through its coils. Relays were used in pre-electronic computers, e.g. *Mark I*.

Software: The combination of a set of programs and documentation that enables the computer to perform a set of defined tasks.

Syntax: A set of rules which enables us to correctly arrange words when we either write or speak a sentence. It is also known as 'grammar' of a language.

Template : A set of features which together help in correctly identifying a particular sensory signal

Thermionic valves: A system of electrodes arranged in an evacuated glass or a metal envelope. On heating, the negative electrode called 'cathode' emits electrons which are attracted by the positive electrode, 'anode'. Most valves also contain a perforated grid interposed between the two electrodes. The grid controls the flow of current through the valve.

Biotechnology has taken rapid strides in recent years and has emerged as one of the most promising branches of science. Committed to rooting out hunger by the creation of high-yielding, disease-resistant crops and pledged to resurrect the earth's verdant green cover, plant biotechnology holds the key to the world of tomorrow. Digging out closely guarded secrets from the inner recesses of the vault of life, and unleashing and channelizing nature's power of self-renewal, this particular brand of biotechnology has virtually rewritten the story of botanical genesis and prospects.

This attractive, profusely illustrated book written in a popular language highlights human creativity and imagination in coaxing life to spring from non conventional sources and takes the reader on a trip down the woods of tomorrow.

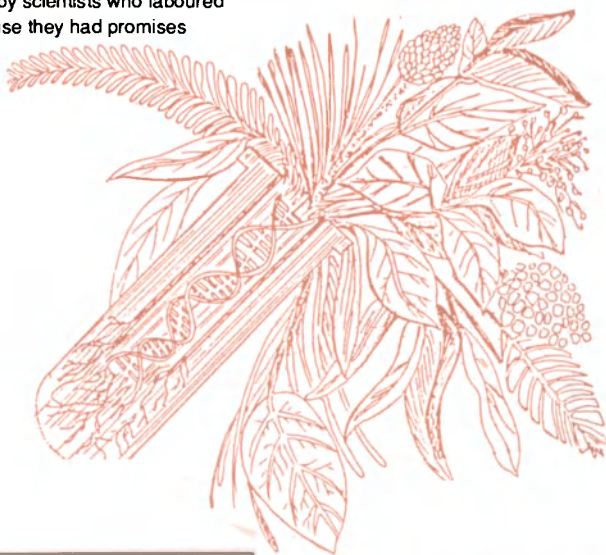
Lovely, dark and deep forests fashioned in test tubes by scientists who laboured long because they had promises to keep.

TEST TUBE FORESTS

**A.F.MASCARENHAS
RAJANI NADGAUDA**

Pages : 72

Price : Rs.15.00



FORTHCOMING TITLES

ABC of Biotechnology
Vaccines for Tommorrow
Biology of Behaviour
Omnipresent Hormones

Orders should be accompanied by Money Order or Demand Draft made payable to "Publications & Information Directorate" New Delhi and sent to:

Sales & Distribution Officer
Publications & Information Directorate, (CSIR)
K.S.Krishnan Marg, New Delhi 110 012

PROMISING in concept, scope and utility, the "fifth" generation computers are at the threshold of ushering in an era of machines that could simulate the human mind. In keeping with its interdisciplinary nature, the instilling of intelligence into computers holds a vast potential in a plethora of fields. Even in its infancy it has revolutionized database management, information retrieval, programming languages and system design. Expert systems are fast gaining laurels in the wide domains of medicine and industry. Equally fascinating is the insight into the future role of increasingly intelligent robots saving precious manpower in all walks of life.

This attractive and lavishly illustrated book, targeted for the non-specialist, explicitly unveils the many facets of artificial intelligence research. The ability to surpass human thinking is apparently not far away, thanks to computer technology that is poised for creating machines possessing artificial intelligence.

About the Author

K.D.Pavate (b.1930), after graduating with Physics and Mathematics from Fergusson College, Poona, studied Physics at the Cavendish Laboratory, University of Cambridge. For the next three years he worked with the Metropolitan - Vickers Electrical Co. Ltd., Manchester. On returning to India, he joined the Central Electronics Engineering Research Institute (CEERI), Pilani.

In 1976, he moved over to Delhi as the head of the newly formed CEERI Centre. Here he established R & D groups to work on various user-oriented aspects of electronics. These included development of audiovisual systems, electronic communication and animation systems for museums, electronic systems to assist the National Literacy Mission, development of small capacity EPABX systems, secure speech communication, speech analysis and speech synthesis technologies, among others.

Pavate's interest in recent years has shifted to science communication and in writing articles on science and technology. *Artificial Intelligence* is his first popular science book.

ISBN : 81-7236-035-5

