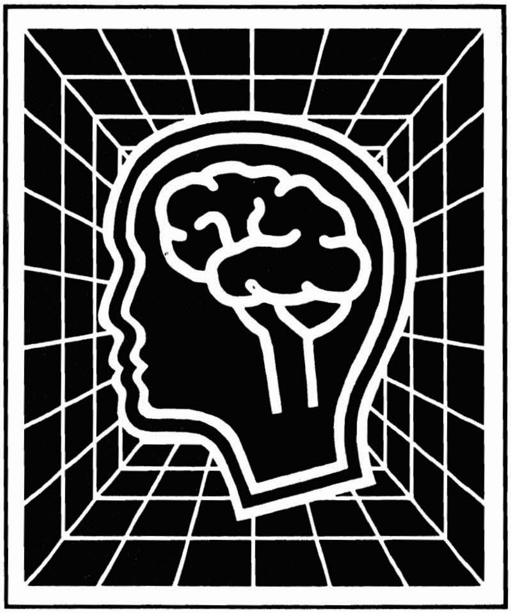
Part One

Foundations



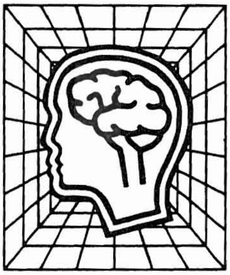
1. Intelligence
2. The Brain and the Computer
3. The Representation of Knowledge

In the first portion of this book we examine, in a very general way, the nature of intelligence and the principal mechanisms by which it is achieved.

Our subject matter includes the attributes of intelligent behavior, the structure of the "reasoning engines" devised by both nature and man, and the critical role played by the way in which knowledge is encoded. These topics provide a foundation for our discussion, in Parts II and III, of cognition and perception, the two major faculties by which intelligence is exhibited.

1

Intelligence



In this book we will explore some of the central aspects of intelligent behavior, and the approaches employed in creating machines that can exhibit such behavior. Our purpose in this chapter is to address three broad questions about the nature of intelligence:

- What is intelligence, and to what extent is it a unique attribute of the human species?
- How can intelligence be measured or evaluated?
- What is the nature of the mechanisms that are capable of intelligent behavior? In particular, can a machine be designed to display intelligent behavior?

WHAT IS INTELLIGENCE?

Intelligence is easier to recognize than to define or measure. While the word "intelligence" is used in ordinary conversation, and has a dictionary definition, it has no agreed-upon scientific meaning, and no quantitative natural laws relating to intelligence have as yet been discovered. In view of this situation, the concept of intelligence is subject to change as our understanding of human intelligence increases. Further, without a scientific definition, much of the social debate over matters relating to intelligence (e.g., contentions about racial differences with respect to intelligence) cannot be rationally resolved.

A dictionary definition of intelligence

includes statements such as *(1) the ability to meet (novel) situations successfully by proper behavior adjustments*; or *(2) the ability to perceive the interrelationships of presented facts in such a way as to guide action toward a desired goal.* We can associate the word "learning" with the first statement, and goal-oriented behavior, problem solving, and understanding with the second. Some additional attributes of intelligence (see Tables 1-1 and 1-2) include reasoning, common sense, planning, perception, creativity, and memory retention and recall.

Theories of Intelligence

Theories of intelligence are primarily concerned with identifying the major independent components of intelligent behavior, and determining the importance of, and interactions between mechanism, process, knowledge, representation, and goals. In particular, such theories address the following issues:

- *Performance theories*: How can one test for the presence or degree of intelligence? What are the essential func-

TABLE 1-1 ■ Attributes of an Intelligent Agent

We expect an intelligent agent to be able to:

- Have mental attitudes (beliefs, desires, and intentions)
- Learn (ability to acquire new knowledge)
- Solve problems, including the ability to break complex problems into simpler parts
- Understand, including the ability to make sense out of ambiguous or contradictory information
- Plan and predict the consequences of contemplated actions, including the ability to compare and evaluate alternatives
- Know the limits of its knowledge and abilities
- Draw distinctions between situations despite similarities
- Be original, synthesize new concepts and ideas, and acquire and employ analogies
- Generalize (find a common underlying pattern in superficially distinct situations)
- Perceive and model the external world (see Box 1-1)
- Understand and use language and related symbolic tools

TABLE 1-2 ■ Attributes Related to, but Distinct from, Intelligence

There are a number of human attributes that are related to the concept of intelligence, but are normally considered distinct from it:

- Awareness (consciousness)
- Aesthetic appreciation (art, music)
- Emotion (anger, sorrow, pain, pleasure, love, hate)
- Sensory acuteness
- Muscular coordination (motor skills)

tional components of a system capable of exhibiting intelligent behavior?

- *Structural/function theories*: What are the mechanisms by which intelligence is achieved?
- *Contextual theories*: What is the relationship between intelligent behavior and the environment with which an organism must contend?
- *Existence theories*: What are the necessary and/or sufficient conditions for intelligent behavior to be possible?

(A separate set of issues is associated with the question of how theories of intelligence can be validated.)

Theories are statements, circumscribed by definitions, about objects and their relationships that are implicit in a body of knowledge. Thus, definitions and theories of intelligence cannot be separated. Quantitative definitions of intelligence range from implicitly defining intelligence as that human attribute which is measured by IQ tests, to assuming that the total information processing capacity of the brain is measured by its size.[1] However, the dimension along which definitions of intelligence differ most is the structural (internal) versus the contextual (external). At the structural extreme, intelligence is viewed as the competence of the human (or animal) nervous system to reason, while at the contextual extreme, intelligence is viewed as the ability of an organism to adapt to its physical and social environment. In the latter case, goals, expectations, stored knowledge,

BOX 1-1 Visual Thinking

The idea that "visual thinking" and artistic creation are part of intelligent behavior has been discussed by Arnheim as follows [Arnheim 69]:

> My contention is that the cognitive operations called thinking are not the privilege of mental processes above and beyond perception but the esssential ingredients of perception itself. I am referring to such operations as active exploration, selection, grasping of essentials, simplification, abstraction, analysis and synthesis, completion, correction, comparison, problem solving, as well as combining, separating, putting into context. These operations are not the prerogative of any one mental function; they are the manner in which the minds of both man and animal treat cognitive material at any level. There is no basic difference in this respect between what happens when a person looks at the world directly, and when he sits with his eyes closed and "thinks."

Another aspect of visual thinking is the concept that the artist constructs his drawings by a reasoning process. Gombrich [Gombrich 61] describes the task of setting down a pictorial likeness on a flat surface as resembling the method used by scientists in arriving at a theoretical description of the natural world. The artist does not simply trace an outline of their visual contours to represent the appearance of things, but instead prepares a hypothetical construction to be matched and then modified in the light of further evaluation. Through an iterative process, the artist gradually eliminates the discrepancies between what is seen and what is drawn, until the image on the flat surface begins to resemble a view of the world as it might be seen through a pane of glass. The iterative process of the artist corresponds to the conjectures and refutations of the scientist in creating a theory of nature.

---

[1] Beyond that needed to support normal body functions.

and prior experience are as important and relevant as the internal reasoning machinery.

Theories of intelligence are largely dependent on whether we define intelligence to be a natural phenomenon appearing in living organisms (especially man), or whether we define it to be an abstract facility with certain specified properties. If intelligence is viewed as an outgrowth of specific biological structures, then it is reasonable to ask whether a single or coherent mechanism produces intelligent behavior, or whether intelligence is the result of a number of relatively independent processes. From a practical standpoint, we might also ask what kinds of measurements are needed to predict human performance in specified tasks requiring intelligence.

For example, if intelligence is a highly integrated process, then it is quite

BOX 1-2 Psychological/Performance Theories of Intelligence

Plato drew a distinction between the cognitive aspects of human nature (thinking, reasoning, problem solving) and what he termed the *hormic* aspects (emotions, feelings, passions, and the will). He theorized on the cause of individual differences in intellect and personality: *The God who created you has put different metals into your composition—gold into those who are fit to be rulers, silver into those who are to act as their executives, and a mixture of iron and brass into those whose task it will be to cultivate the soil or manufacture goods.*

The modern concept of intelligence was formulated by Herbert Spencer and Sir Francis Galton in the nineteenth century—they believed in the existence of a general ability distinct from, and in addition to, more specialized cognitive abilities. Galton also introduced some of the tools and methodology by which statistical correlation between tests of performance became the basis for answering questions about the relationships between different cognitive skills and general intelligence. Galton, and in 1890, James Cattell devised "intelligence tests" based largely on sensory and motor functions (e.g., color discrimination, time perception, accuracy of hand movement, description of imagery) under the assumption that these easily measured quantities were highly correlated with intelligence. In 1895, Alfred Binet argued for more direct testing of cognitive skills (e.g., verbal comprehension, moral sensibility, aesthetic appreciation). Binet, in 1904, also introduced the concept of "mental age," closely related to the idea of intelligence quotient (IQ: 100 times mental age divided by chronological age).

Early in this century, C. Spearman employed a technique called "factor analysis" to provide statistical evidence for the predominance of a general cognitive ability. Spearman proposed a "two-factor" theory of intelligence: Every intellectual activity has two underlying components, one specific to that particular activity, and one common to all intellectual activities. This second factor was called "general intelligence" or *g*. Following Spearman, L.L. Thurstone developed and employed a more advanced form of factor analysis to argue that Spearman's general factor *g* might be an artifact arising out of a set of primary mental abilities: spatial visualization, perceptual ability, verbal comprehension, numerical ability, memory, word fluency, and inductive and deductive reasoning. There was also evidence that these primary mental abilities were the base of a hierarchy in which the primary abilities first cluster into verbal, numerical, and logical groupings, and then finally into Spearman's *g*.

In the same time period as the work of Thurstone (1930–1950), Cyril Burt* used new statistical methods in an attempt to determine the relative contributions of heredity and environment to IQ test perform-

---

*See "The Real Error of Cyril Burt" in Gould (Gould 81) for a description of how Burt faked some of his data.

possible that a single number, such as an IQ test score, could be a good predictor of a human's ability to perform in almost any intellectual task domain. To the extent that intelligence arises from a loosely integrated combination of different mechanisms, prediction of human performance would depend on tests much more closely related to the specific task of interest. Most psychological theories of intelligence, and intelligence tests that implicitly arise from these theories, assume that intelligence is a composite of a relatively small number of component factors, possibly dominated by a single integrating factor. These theories can be called "performance theories," since they are based on measurements of performance and make assertions about relationships and correlations between different tests of performance (see Box 1-2). Such theories are largely empirical and, while they have

---

BOX 1-2 *(continued)*

ance, and by implication, to human intelligence.

Between 1950 and 1980, Guilford [Guilford1967] formulated what was intended to be a comprehensive theory of the structure of human intellect. He identified three classes of variables:

1. The five activities or operations performed—cognition (immediate awareness of information), memory, convergent (logical) thinking, divergent (creative) thinking, and evaluation
2. The material or content on which the operations are performed—images, symbols, concepts, and nonverbal social perceptions
3. The six products which result from the operations—unitary items, classes, relations, systems, transformations, and implications

Guilford's system results in $5 \times 4 \times 6 = 120$ separate factors or abilities contributing to intelligence. There is no general factor. Guilford and his associates used factor analytic methods on performance tests to prove the existence of many (but not all) of the factors he defined.

Raymond Cattell (no relation to James Cattell, circa 1890), working in the same time frame as that of Guilford, proposed and provided statistical tests for an alternative theory of intelligence in which *g* combined two distinct general abilities: "crystallized" and "fluid" intelligence. Crystallized abilities are based on learned cultural knowledge (vocabulary, numerical skills, mechanical knowledge), while fluid intelligence relates to innate perceptual and reasoning abilities.

Most of the above work used statistical methods to determine how mental processes vary from individual to individual, and to study the relationships among these mental processes in a single individual. One exception to such studies, which are based largely on statistical analysis of performance tests, is the work of Piaget (described in Chapter 5) who proposed a qualitative theory of how intelligence evolves in an individual—from "sensory," "concrete," and "subjective" in the child, to "abstract" and "objective" in the adult. There are also the more cognitive-type theories of G.H. Thomson and E.L. Thorndike (1920–1940) who hypothesized that general intelligence is a function of the number of structural bonds (or stimulus-response connections) that have been formed between specific mental abilities. Performance on any one task would be the result of activation of many of these bonds.

In retrospect, as noted by Butcher [Butcher 73]: "During the first forty years of this century, the idea of intelligence or general mental ability was found to be useful and important by psychologists. . . . Recently, however, the concept has become less generally acceptable and more exposed to various kinds of criticism."

Almost all psychological investigators employ a paradigm based on statistical testing for the existence of presupposed intellectual structures. We believe that the study of the *computational requirements* for intelligent behavior—the underlying theme of work in the field of artificial intelligence—will provide a more productive means for understanding the nature of both human and machine intelligence.

significant practical utility, they offer very little insight into the nature of intelligence. As noted by Butcher [Butcher 73],

> The study of human intelligence has yielded a large accumulation of knowledge about individual differences, but very little about the basic laws of cognitive functioning. . . . For a concept to be valuable it should have more than purely statistical support, and be more than a blind abstraction from a set of correlated performances.

Most of our concern in subsequent chapters will be with what might be called structural/function theories of intelligence. These are theories that propose certain physical or formal structures as the basis for intelligent behavior, and then examine the functionality that results. For example, if we assume that intelligence is a result of formal logical inference, then we might ask if there are human capabilities that could be shown to be unachievable in the formal system because of limitations inherent in logical reasoning. In Chapters 2 and 4 we show that logical systems do indeed have limitations we do not usually ascribe to people.

Finally, there are (largely philosophical) theories about the physical conditions necessary for the mechanization of intelligence; we call these existence theories. For example, there is a school of thought that asserts that intelligence is a nonphysical property of living organisms, and cannot be re-created in a machine. Another school holds that intelligence is an emergent property of organic matter—silicon[2] is inadequate, but when we eventually learn how to build machines out of organic compounds, we might have a chance of inducing intelligent behavior. One other school believes that intelligence is a functional property of formal systems, and is completely independent of any physical embodiment. This latter viewpoint is the one with which we will be primarily concerned.

Theories of Mind

As previously noted, we will extensively discuss the attributes of intelligence and intelligent behavior, describing mechanisms that are capable of achieving such behavior in both living organisms and machines. However, we will not provide a precise definition of intelligence; this book as a whole is our contribution in this regard. Nor will we do much to "explain" or elucidate the conscious awareness that seems to be an essential component of human intelligence. Introspectively, there appears to be an "inner entity," the *mind,* which views the world through the body's sensory organs, "thinks," "understands," and causes the body to react in an appropriate manner.

A primary concern of philosophy is the attempt to understand the relationship between the internal world of our conscious awareness and the external physical world. Plato (c.428–c.348 B.C.) held that the mind (*psyche*) was in charge of the body and directed its movements. In the *Phaedrus* Plato spoke of the mind as having both *appetitive desires* and *higher desires*, and having also a rational capacity to control, direct, and adjudicate between these two types of desires. Later theories held that man was made of two substances, mind and matter. The theory

---

[2] "Silicon" is a shorthand way of referring to silicon-based microcircuits that are used in digital computers.

that the mind and body are distinct, known as "dualism," was given its classical formulation by Descartes in the seventeenth century. In his *Discourse on Method* (1637) he argued that the universe consists of two different substances: mind, or thinking substance, and matter, which can be explained by science and mathematics. Only in man are mind and matter joined together. His concept was that mind was an immaterial nonextended substance that engages in rational thought, feeling, and willing. Matter conforms to the laws of physics with the exception of the human body, which Descartes believed is causally affected by the mind, and which causally produces certain mental events. A basic problem that must be dealt with in this theory is how interaction can occur between the nonphysical and the physical.

The current dominant school of thought[3] regards mind as being a purely physical phenomenon. Sagan [Sagan 78] sums up this view succinctly: "My fundamental premise about the brain is that its workings—what we sometimes call 'mind'—are a consequence of anatomy and physiology and nothing else." A similar view by Restak [Restak 84] is based on a belief that signals from the brain will some day be understood:

> Since the development of appropriate technologies, it has become obvious that thoughts, emotions, and even elementary sensations are accompanied by changes in the state of the brain . . . a thought without a change in brain activity is impossible . . . to understand the "mind," therefore, it is necessary to understand the brain—how concepts are arrived at, the mechanisms underlying perceptions, memory, the neuro-chemistry of our emotions, and so on.

Searle [Searle 84] comments on the mind-body problem: "Mental phenomena, all mental phenomena whether conscious or unconscious, visual or auditory, pains, tickles, itches, thoughts, indeed, all of our mental life are caused by processes going on in the brain."

The *information processing model* is used by Newell and Simon [Newell 72]. They view formal logic as a way of capturing ideas by symbols, and the algorithmic alteration of such symbols as leading to mindlike activity: "The persistence of concern with the mind-body problem can be attributed in part to the apparent radical incongruity and incommensurability of 'ideas'—the material of thought—with the tangible biological substances of the nervous system."

Those who take the above computational point of view feel that the mind-body problem will disappear when we have demonstrated the operation of mind using formalisms and algorithms for manipulating symbols.

One should not think that all modern researchers look at duality with scorn. In his final book, *The Mystery of the Mind*, the famous neurosurgeon Wilder Penfield [Penfield 78] doubts that an understanding of the brain will ever lead to an explanation of the mind: "Consciousness of man, the mind, is something not to be reduced to brain mechanisms." Another example of this point of view is contained in *The Self and the Brain* by Karl Popper

---

[3] Thomas Hobbes (1588–1679), John Locke (1632–1704), and David Hume (1711–1776) originated the idea that thoughts obey physical laws and can be characterized as computational processes.

and John Eccles [Popper 77], an updated plea for dualism, the belief that the brain and the mind are distinct entities.

Until someone provides convincing proof of the physical basis of mind, we can expect the mind-body debate to continue.

HOW CAN INTELLIGENCE BE MEASURED OR EVALUATED?

Assessing Human Intelligence

As noted in the previous section, while an intuitive concept of intelligence exists, there is no formal or scientific definition of intelligence that is widely accepted. If intelligence cannot be defined, then it certainly cannot be measured in any precise or comprehensive manner. If intelligence tests do not measure *intelligence*, what do they measure? The purpose of most of these tests is to predict the future performance of the person being tested with respect to an ability to compete or perform in an academic program or in a skilled work task. Whether or not an "intelligence test" actually does have the required predictive power can only be determined by extensive testing in the specific application area.

There are a number of intelligence tests in widespread use, one of the most popular being the Terman-Merrill revision of the Binet-Simon intelligence scale. The original Binet-Simon work was performed in the period 1905–1911. Binet insisted on three cardinal principles for using his test:

1. The scores are a practical device and are not intended as the basis for a theory of intellect. They do not define anything innate or permanent. What they measure is not "intelligence."
2. The scale is a rough, empirical guide for identifying mildly retarded and learning-disabled children who need special help. It is not a device for ranking normal children.
3. Whatever the cause of difficulty in children identified for help, emphasis should be placed on improvement through special training. Low scores should not be used to mark children as innately incapable.

All of his warnings were disregarded, and his scale was used as a routine device for testing all children [Gould 81]. The Binet-Simon test was superseded by Terman's 1916 standard version, and then by the Terman-Merrill revision of 1937, and by a later revision in 1960. Table 1-3 lists some of the categories of items found in the 1960 revision. It is interesting to note that the procedure for selecting questions for this test was that the questions had to satisfy certain preconceived notions of what results the test should produce. This is standard practice in all intelligence test construction. For example, questions that yield systematically higher scores for either boys or girls are eliminated. By use of question selection and scoring procedures, the test was constructed so that for the white American population, biased somewhat toward urban and above-average socioeconomic level persons, the scores would have a *normal distribution* with an average score of 100, and a standard deviation of 16. This means that 50 percent of the reference group (white Americans) would score under 100; 85 percent would score

TABLE 1-3 ■ Categories of Questions that Appear in the Terman-Merrill Version of Binet-Simon Intelligence Scales

Obey simple commands	Comprehension
Identify object by use	Opposite analogies
Repeat digits	Pictures alike and different
Response to pictures	Memory for sentences
Repeat digits reversed	Vocabulary
Memory for stories	Picture completion (picture of a man)
Find absurdities in pictures	Discriminate animal pictures
Picture vocabulary (recognize pictorial objects)	

under 116; 97.5 percent would score under 132, etc.

Another commonly used intelligence test, the Wechsler intelligence scale, uses separate tests for adults and for children. This test is divided into two main parts, one to test predominantly verbal ability, and a second to test performance (see Table 1-4). Even though the Wechsler and the Binet tests have somewhat different philosophies and different categories of questions, they use similar principles of test construction and produce scores that are in reasonable agreement.

Starting in the 1960s, the role and value of intelligence tests have been seriously challenged. In particular, critics have argued that these tests take too narrow a view of intelligence, and that they are based on such dubious assumptions as: (a) A child is born with a fixed or predetermined level of intelligence; (b) IQ tests can measure this intelligence; (c) IQ scores will show little variation from early childhood to old age; and (d) the tests employed, relatively unchanged since their introduction in the early 1900s, are good predictors of human performance. Not surprisingly, political and social concerns have been intermixed with issues of scien-

TABLE 1-4 ■ Categories of Questions that Appear in the Wechsler Intelligence Scale

I. Verbal Tests

- General information (who is President of the United States?)
- General comprehension (what would you do if? . . . Why do we usually? . . .)
- Arithmetic reasoning (simple mental arithmetic)
- Remember series of digits forward and backward
- Similarities (pairs of words: subject has to tell how they are alike)
- Vocabulary (explain meaning of words)

II. Performance Tests

- Digit symbol coding (subject must assign digits and symbols to pictures)
- Picture completion (subject must detect nose missing from face)
- Block design (construct color-pattern designs in duplication of given patterns)
- Picture arrangement (subject arranges pictures to tell a story)

tific validity in addressing the question of what is reasonable and meaningful in regard to the testing of human intelligence.

Assessing Machine Intelligence

If one were offered a machine purported to be intelligent, what would be an appropriate method of evaluating this claim? The most obvious approach might be to give the machine an IQ test. As will be seen in later chapters, we already know how to build machines that can perform quite well on selected portions of such a test. For example, machines can currently solve high school algebra problems, solve the type of geometric analogy problems used on IQ tests, answer questions about the content of a simple story, parse English sentences, etc. However, none of this would be completely satisfactory because the machine would have to be specially prepared for any specific task that it was asked to perform. The task could not be described to the machine in a normal conversation (verbal or written) if the specific nature of the task was not already programmed into the machine. Such considerations led many people to believe that the ability to communicate freely using some form of natural language is an essential attribute of an intelligent entity.

In 1950, Alan Turing proposed an "imitation game" to provide an operational answer to the question, "Can a machine think?" [Dennett 85, Hodges 83, Turing 50]. The game is played with three people, a man, a woman, and an interrogator who may be of either sex. The interrogator stays in a room apart from the other two, and attempts to determine which of the other two is the man and which is the woman. The man tries to convince the interrogator that he is the woman. Communication between the interrogator and each person is by teleprinter, and the interrogator is free to ask any question of the participants.

Suppose we now ask the question, "What will happen when a machine takes the part of the man in this game?" Turing felt that a machine could be considered "intelligent" when the interrogator decides wrongly as often when the game is played with the machine as when the game is played between a man and woman. It should be noted that to accomplish this, the machine must be able to carry out a dialogue in natural language and reason using an enormous database of "world knowledge." The "man-woman" formulation proposed by Turing is not usually stressed in describing the imitation game. Instead, the theme is usually the idea of a machine convincing an interrogator that it is a person.

The Turing test has more historical and philosophical importance than practical value; Turing did not design the test as a useful tool for psychologists. For example, failing the test does not imply lack of intelligence. The important central idea is that the ability to successfully communicate with a discerning person in a free and unbounded conversation is a better indication of intelligence than any other attribute accessible to measurement.

IS MAN THE ONLY INTELLIGENT ANIMAL?

If we examine the attributes of intelligent behavior that were presented in Table 1-1, we can find examples of superior animal performance in each of the attribute categories. Until recently, however, it was believed that only man, of all animals, could produce (as opposed to understand) structured linguistic phrases to communicate meaning. Experiments more fully described in Chapter 6 (Language and Communication) have demonstrated that chimpanzees can learn American Sign Language (ASL) and can learn to assign word meanings to physical tokens (e.g., small colored plastic disks), and then arrange these tokens into structured sentences to communicate with their trainers. Thus in an objective sense, it appears possible that man differs from the higher mammals mainly in degree of intellectual ability rather than in having some unique and unshared capability.

In a related sense, recent work by Gordon Gallup [Gallup 77] addresses the question: "Do minds exist in species other than our own?" Gallup defines "mind," "consciousness," and "self-awareness" to mean essentially the same thing. His operational test for self-awareness is that an organism can identify itself in a mirror; for example, a child can recognize his reflection at approximately a year and one half to two years of age. Gallup discovered that while humans, chimpanzees, and orangutans can learn to recognize themselves in mirrors, no other primates can! Thus even though gorillas appear to possess some degree of linguistic competence (see Patterson's work with Koko [Patterson 78]), gorillas fail this particular test for self-awareness. Our understanding of the relationship between self-awareness, language, and intelligence is still at a very primitive stage.

THE MACHINERY OF INTELLIGENCE

Reliance on Paradigms

It would appear that we deal with the world by relying on *paradigms*, overall strategies or frameworks that we use as the high-level plan for solving various problems. The use of paradigms allows us to reduce the complexity of our environment by discarding most sensory data and selecting only that which is relevant. Thus, we are usually unconscious of breathing, body support pressures, background hums and noises, but any of these could become important in special situations; e.g., consciousness of breathing could be important to an astronaut in a space suit. If the paradigm for dealing with a situation is not adequate, then performance will be poor: *If the only tool you have is a hammer, you tend to treat everything as if it were a nail.* For example, the city dweller may not have the proper paradigms for dealing with a jungle environment. His "city paradigms" would not help him to focus on the necessary sensory data; he would not be able to properly interpret the jungle environment data being received, and he would not be able to invoke the appropriate actions for survival.

Two Basic Paradigms

There is evidence to show that the two hemispheres of the human brain are specialized to deal with problems in different ways by the use of two distinct types of paradigms. The sequential (or logical) paradigm is based on a problem solving

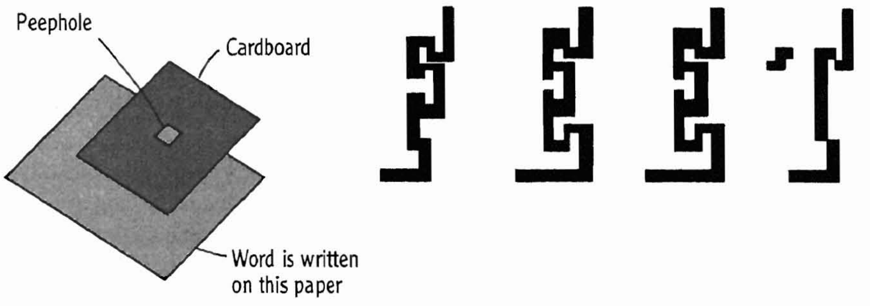


FIGURE 1-1 Experiment Illustrating the Distinction Between the Sequential and Gestalt Paradigms for Problem Solving and Perception.

Subject is asked to read a word. Subject is not permitted to see the word as a whole, but must examine it by looking through a small peephole in a piece of cardboard that he or she can slide over the paper on which the word is written.

approach that considers only a small portion of the available data at any given time, while the parallel (or *gestalt)* paradigm processes data on a global basis, or *all at once.* That these are fundamentally different capabilities can be seen from the experiment offered in Fig. 1-1. A human subject is given an opaque card with a small window in it and asked to explore an English word (printed in a rather unusual type font) by moving the card over it. The subject will not be able to perceive the word because all of the pattern data must be viewed at once to reveal the structure. The important point here is that problems that can be successively decomposed into simple and relatively independent parts can be effectively solved using the sequential/logical paradigm. On the other hand, many problems, especially those of a perceptual nature as in the example, do not permit decomposition, and can be effectively solved only by employing the gestalt paradigm that can deal with global information.

In most normal people, the left hemisphere of the brain is specialized to deal with tasks amenable to a sequential paradigm. These include language understanding and production, logical reasoning, planning, and time sense. The right hemisphere of the brain is more competent to deal with spatial tasks and tasks requiring a global (gestalt) synthesis. These include comparing and identifying visual imagery,[4] visual and analogic reasoning (including,

---

[4] There is evidence to support the surprising discovery that mental images are neither generated nor manipulated by the normal sensory-based visual system; a module in the left hemisphere, but not language-based, appears to provide the necessary competence. There is no similar module in the right hemisphere [Gazzaniga 85, p. 134]

perhaps, dreaming), and body sense and coordination.

Some of the evidence supporting the concept of specialization of the two brain hemispheres with respect to the gestalt and sequential paradigms has come from *split brain* experiments with subjects who have had brain surgery to control epilepsy. The connection between the right and left hemispheres is severed so that signals no longer flow between the hemispheres. By examining the subjects of such experiments, it has been found that the human brain can support two separate and distinct "personalities," one in each hemisphere, as described in Box 1-3. The philosophical implications of this finding are rather staggering and are still being investigated.

ARTIFICIAL INTELLIGENCE (AI)

The Mechanization of Thought

The idea of man converting an inanimate object into a "human-like" thinking entity is an old one. In Greek myth we have the story of Pygmalion, a king of Cyprus who fashions a female figure of ivory that was brought to life by Aphrodite. In the Golem legend of the late sixteenth century, Rabbi Löw of Prague breathes life into a figure of clay. In the nineteenth century there is the story of the scientist Frankenstein, who creates a living creature.

During the seventeenth century, the idea arose of converting thought into a formal notation and using a calculating device to carry out the reasoning operations. In 1650, the English philosopher Thomas Hobbes proposed the idea that thinking is a rule-based computational process, analogous to arithmetic. Gottfried Wilhelm Leibnitz (1646–1716) describes his book *De Arte Combinatorica* (1661) as containing "a general method in which all truths would be reduced to a kind of calculation." Much later, in 1854, George Boole published *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*. In the first chapter he states,"The design of the following treatise is to investigate the fundamental operations of the mind by which *reasoning* is performed."

The dream of devising a formal system that could be a basis for all reasoning seemed to be almost at hand with the publication of Russell and Whitehead's *Principia Mathematica* (1910–1913). The codification of logic and the reduction of significant portions of mathematics to the language of logic appeared to provide the means by which people (or machines) could do mathematics without having to understand what was actually happening; it would be sufficient to manipulate the symbols according to permissible logical transformations. Even the sequencing of the transformations could be done "blindly" (mechanically).

It even seemed possible that all questions of philosophy could be phrased and answered in such a logical language. The logical positivists, extending the empiricism of David Hume, believed that only within the framework of a logical language could philosophical problems be raised with any degree of precision: All problems are either questions of fact or questions of logic; the former are properly relegated to the sciences and philosophy simply becomes a form of logical analysis. Thus,

BOX 1-3 Split-Brain Experiments

If we were to cut in half a complicated mechanism, such as a car, a computer, or a person, we would certainly not expect each of the halves to continue to function. Nevertheless, when the human brain is cut in half by severing the major connecting bundle of nerve fibers linking the two hemispheres, the *corpus callosum* (see Fig. 1-2), the two resulting pieces continue to operate independently, as if two separate personalities now exist in place of the original individual (see [Ornstein 73] and [Gazzaniga 85]).

Since we devote a significant portion of Chapter 2 to a discussion of the structure and functioning of the brain, here we merely note the following facts. The brain, in terms of its outward appearance, is bilaterally symmetrical. The two similar appearing "hemispheres" of brain tissue are spatially separated, and normally communicate through the *corpus callosum.* Each half of the brain controls muscles on only one side of the body, and receives direct sensory inputs from sense organs monitoring only the left or right half of the physical space surrounding the individual. For example, nerve cells in each eye that monitor the left half of the visual field have direct connections only to the right hemisphere. (This right-left "crossover," which also occurs in muscular control, has no special implications for this discussion.)

In a *cerebral commissurotomy* operation, performed to alleviate severe epileptic seizures in some patients who did not respond to medication or other forms of treatment, the *corpus callosum* is cut to prevent

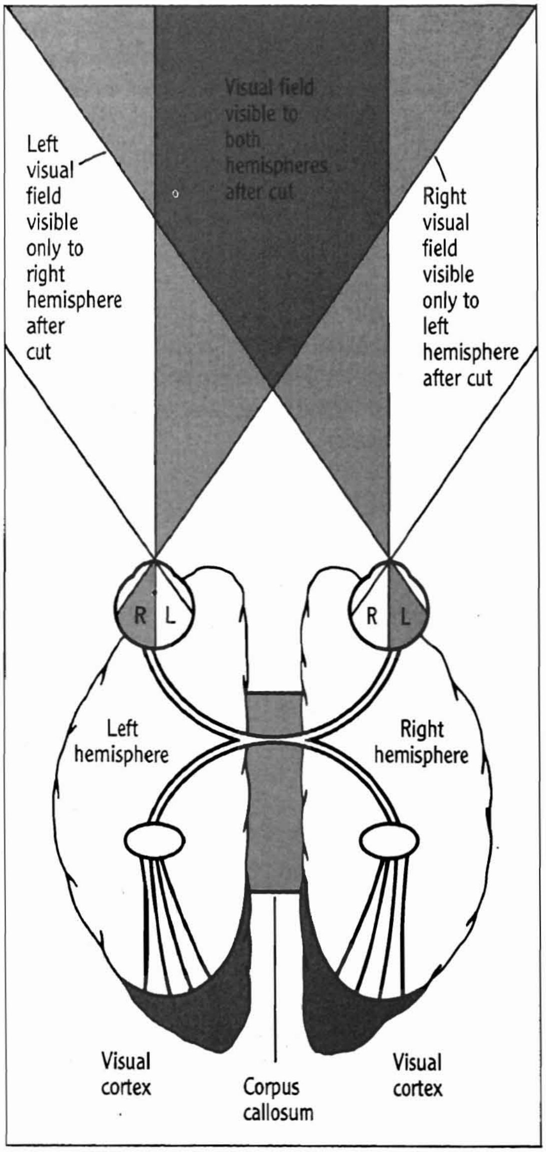


FIGURE 1-2 The Split Brain.

Cutting the corpus callosum effectively separates the two hemispheres of the human brain.
Localized functions relevant to the split-brain experiments: *Left hemisphere*: right visual half-field; right hand; right ear; left nostril; main language center. *Right hemisphere*: left visual half-field; left hand; left ear; right nostril; simple language communication.

as Barrett [Barrett 79] notes, ". . .when Philosophy, which was supposed to question everything, turns to questioning itself, it finds that it has vanished," i.e., it is reduced to physics and logic. However, at least in part for reasons touched on below, the dream of a formal system for reasoning began to fade in the 1930s.

Formal investigation of the limits of mechanical reasoning did not occur until the twentieth century. Alan Turing, a British mathematician, carried out investigations using a conceptual model that he called an automaton (now known as a Turing machine). In the 1950s, Turing was able to prove formally that there is a "universal automaton" that can simulate the performance of any other automaton if it is given an appropriate description of that automaton.[5] In addition, Turing proved that certain types of automata could never be built, e.g., one that could tell whether an arbitrary program run on an arbitrary automaton would ever halt. Results concerning the limitations of automata are described in Chapter 2.

Also in this era, John von Neumann dealt with the questions of how complex a device or construct need be in order to be self-reproductive, i.e., to make a copy of itself. He also investigated the problem of how to design reliable devices that must be made from parts that can malfunction [von Neuman 56a]. He surmised that autom-

---

[5] Simulation of one computer type by another is now quite common. In fact, one often simulates a computer on another type of computer in order to verify the design prior to fabrication.

BOX 1-3 (*continued*)

a seizure starting in one hemisphere from spreading to the other. Although it first appeared that there were no undesirable aftereffects, tests later showed that "split-brain" patients were indeed different after the operation.

In one test situation, the split-brain subject is seated in front of a screen that hides his hands from his direct view. His gaze is fixed at a spot on the center of the screen and the word "nut" is flashed very briefly on the left half of the screen. This image goes to the right hemisphere of his brain which controls the left side of his body. The subject then uses his left hand to pick out (by sense of touch) a nut from a pile of objects hidden from his view. But he cannot verbally report what word was flashed on the screen because the image (of the word "nut") could not reach the left-brain hemisphere where the main centers for language production are located and the left hemisphere receives no direct sensory inputs from the left hand. The language portion of the subject's brain controlling conversation with the experimenter seems unaware of what the subject's left hand is doing. If the word flashed on the screen remains longer than one tenth of a second, the subject can move his eyes so that the word is also projected to the left hemisphere. If the subject can move his eyes freely, information goes to both hemispheres, and this is why the deficiencies caused by severing the hemispheric connections are not readily apparent in daily activities.

Experiments with split-brain patients tend to confirm knowledge obtained through other means in normal human subjects. These results indicate the separation, or at least dominance of skills, to individual hemispheres, based on whether they are sequential/analytic (left hemisphere) or spatial/gestalt (right hemisphere). While there is still some controversy regarding the precise nature of such specialization (Gardner82), there can be little argument with the finding that split-brain patients exhibit two distinct streams of consciousness. It is reasonable to ask whether *cerebral commissurotomy* produces a splitting or doubling of the mind, or whether it exposes a multiplicity previously present.

ata whose "complexity" is below a certain level can only produce less complicated offspring, whereas those above a certain level can reproduce themselves or even construct higher entities.

In recent years, the information processing paradigm has become a popular model for explaining the reasoning ability of the human mind. As stated by Simon [Simon 81], "At the root of intelligence are symbols, with their denotative power and their susceptibility to manipulation . . . and symbols can be manufactured of almost anything that can be arranged and patterned and combined." This view, that intelligence is independent of the mechanisms by which the symbol processing is accomplished, is held by most researchers in the field of artificial intelligence.

The Computer and the Two Paradigms

The digital computer is the only device that has been used to achieve any significant degree of artificial (machine) intelligence. However, the conventional digital computer is a sequential symbol manipulator, and is primarily suitable for tasks that can be broken down into a series of simple steps. Thus, it is only effective for realizing one of the two basic paradigms employed in human intelligence: the sequential paradigm. Attempts to duplicate human abilities involving the global (gestalt) paradigm, such as visual perception, have been strikingly inferior, even for visual tasks that people consider extremely simple.

At the present time there is a vast difference in favor of the human brain, as compared to the computer, with respect to logical complexity, memory characteristics, and learning ability. Computer-based AI must be specialized to very restricted domains to be at all comparable to human performance. For example, games with a limited number of positions and possible moves are well matched to the computer's great search speed and infallible memory.

How can we Distinguish between Mechanical and Intelligent Behavior?

Two basic attributes of intelligence are learning and understanding. One might think that an artificial device possessing these capabilities is indeed intelligent. However, we can illustrate the presence of both of these attributes in the very limited context of a coin-matching game (Box 1-4). In this example, the computer *learns* the playing pattern of its opponent, and in practice will beat almost all human opponents who are not familiar with the details of the program. The computer demonstrates its *understanding* of the game situation by its outstanding ability to predict the opponent's moves. However, the computer starts with the key elements of its later understanding, since the programmer has provided the model of choosing heads or tails based on the statistics of the opponent's previous four-move patterns. The only active role played by the program is to collect the statistics of play, and to make choices based on these statistical data. To the outside observer the program seems intelligent, but once we examine its actual details we see that it is quite simple and mechanical. Some might point out that this same argument can also be applied to human performance; it is conceivable that most of the basic models necessary for intelli-

BOX 1-4 A Coin-Matching Program

The following computer program for playing a coin-matching game seems to the external observer to be intelligent, but turns out to be quite simple and mechanical in design. This illustrates the point that it is difficult to judge intelligence based strictly on observed performance on a specific task.

The computer plays against a single opponent in a game of "matching coins." On each play of the game, the computer makes a choice between heads (H) or tails (T), and indicates its choice by printing H or T. Before looking at the computer's choice the human also decides on H or T, trying to match the choice of the computer. After making a decision, the human opponent pushes either the H or T button on the computer console. If the human matches the computer, he gains one point, if not he loses one point. If the score reaches +25 the human wins, if the score reaches −25, the computer wins. The human is not allowed to flip a coin or use some other random device in making his choice.

Typically, when the game first begins, the score stays close to zero. Then, as the computer observes the behavior of its nonrandom human opponent it finds certain regularities in his play, and is able to predict his moves in advance well enough to beat almost every human player. Even if the human tries to act randomly, he cannot accomplish this well enough to fool the computer.

Thus, this program exhibits the two main attributes of intelligent behavior: (1) it learns, i.e., modifies its own strategy of play to take advantage of the way its opponent is playing, and (2) it understands, i.e., it knows the rules of the game, and after a learning period, it predicts how its opponent will behave, and acts appropriately.

The way the computer accomplishes this apparently sophisticated behavior is actually rather simple: the program forms a table of all possible four-move sequences (there are 16 such sequences), as shown in Fig. 1-3. During the course of play, each time a particular four-move sequence by the opponent is followed by an H, the count in the H column in that row of the table is incremented, and similarly for a T. Thus, after a sufficiently long period of play, the computer can predict the most likely next move of its opponent based on his last four moves.

The basic approach described above can be augmented with a number of additional features designed to keep the human from guessing how the program works and for making it look human in its performance. For example, every once in a while, based on a random process within the program, the computer will make a move which is less likely to win. Obviously, this type of bluffing cannot be done too often.

	Previous four moves of opponent				Relative frequency of H or T chosen by opponent following indicated four-move sequence	
					H Count	T Count
1	T	T	T	T		
2	T	T	T	H		
3	T	T	H	T		
4	T	T	H	H		
•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
16	H	H	H	H		

Number of times that opponent selected T after the sequence TTTT

Number of times that opponent selected H after the sequence TTTT

FIGURE 1-3 Prediction Table Developed by the Program for Playing the Coin-Matching Game.

gent performance are inborn, and all we do is select the proper model and adjust the parameters.

The Role of Representation in Intelligent Behavior

As indicated in the previous section, a paradigm is an overall approach for dealing with a class of problems. One of the most critical elements in the specific realization of a paradigm is the *form* in which the relevant knowledge is encoded; we devote all of Chapter 3 to this important subject. To illustrate the role played by the selected representation in solving a problem, consider the example depicted in Fig. 1-4, which shows a configuration of 17 sticks. The problem is to remove five sticks so as to leave three squares with no extra sticks remaining. You are required to find all such solutions! You might try to find one such solution before you read further.

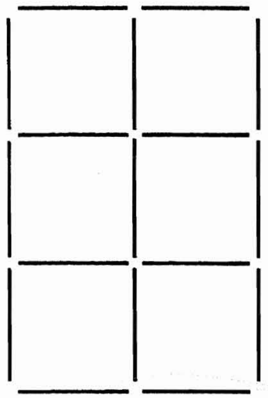


FIGURE 1-4
The Stick Configuration Problem: the Role of Representation in Problem Solving.

The problem is to remove five sticks so as to leave three of the original squares with no extra sticks, and to do this in all possible ways.

If the primitive element you manipulate in searching for a solution is the individual stick, and you remove five sticks at a time and check the result, then even if you are careful not to repeat a particular trial twice, you must make over 6000 trials to be sure that you have found all possible solutions. (There are about 6000 combinations of 17 sticks taken 5 at a time.)

If the primitive element you manipulate is a square, you can select three squares at a time and retain the configuration if there are exactly five sticks left to be removed. Then there are only 20 unique configurations that must be examined to find all solutions, and there is a 300:1 reduction in the number of trials over the approach based on representing the given configuration as a collection of individual sticks. (There are 20 combinations of 6 squares taken 3 at a time.)

Finally, we note that there are 17 sticks, and after removing five, the remaining 12 can form three squares only if these squares are noncontiguous (i.e., have no sides in common). It is easily seen that there are only two configurations of three noncontiguous squares, and both of these are valid solutions. Here, by using a representation that allowed us to employ deductive reasoning, the required effort is reduced by a factor of 3000:1.

SUMMARY AND DISCUSSION

Intelligence is more an open collection of attributes than it is a single well-defined entity. Some of the attributes most closely identified with intelligence are learning, reasoning, understanding, linguistic competence, purposeful behavior, and effective interaction with the environment (including perception). Since intelligence has no clear definition, differing theories of intelligence are not necessarily in conflict, but often differ mainly in the assumed definition of intelligence as either (1) a natural phenomenon appearing in living organisms, especially man, or (2) an arbitrarily specified set of abilities.

Most psychological theories of intelligence are what might be called "performance theories" since they are based on measurements of performance in specified skills, and make assertions about the relationships and correlations between different tests of performance. For example, correlations between tests have been used by investigators attempting to determine if human intelligence is the result of a single coherent mechanism or a collection of loosely integrated independent processes. Such theories are largely empirical and offer very little insight into the nature of intelligence. Most of our concern in the later portions of this book is with understanding how specified abstract structures can produce intelligent behavior.

Intelligence tests, whether for people or machines, have some practical utility, but cannot be expected to accurately measure an undefinable quantity. Another complicating factor in our understanding of intelligence is the role played by consciousness, and the relation between mind and brain.

It is possible to assume that most intelligent behavior arises from one of two distinct paradigms (strategies): In the sequential (or logical) paradigm, a single path is found which links available knowledge and evidence to some desired conclusion; in the parallel (gestalt) paradigm, all connections between evidence and possible conclusions are appraised simultaneously. There is some evidence that the human brain has separate specialized machinery for each of these two paradigms.

A key insight provided by work in artificial intelligence is that intelligent behavior not only requires stored knowledge and methods for manipulating this knowledge, but is critically dependent on the relationship between the specific encoding of the knowledge and the purpose for which this knowledge is used. This concept, the central role of representation in intelligent behavior, is one of our major themes.

The Ultimate Limits of AI. We have briefly sketched the nature of human and machine intelligence. In later chapters we will repeatedly return to the questions, "What can a machine know about the world in which it exists?" and "What are the mechanisms needed to acquire, understand, and employ such knowledge?" We will also address a number of basic questions concerning the limits and ultimate role of machine intelligence:

- Can man create a machine more intelligent than himself?
- Are there components of man's intelligence that cannot be found in any animal or duplicated in a machine?
- Can all intelligent behavior be duplicated by the current approach to AI, namely by decomposing a given problem into a sequence of simple tasks

or subproblems that can be precisely stated and solved?

- Can a machine ever exhibit fully human behavior without having been human and thus properly socialized? In a more limited sense, is human intelligence in some way bound up in the *human experience* or even human heredity?
- Is intelligent behavior realizable, or even conceivable, with the type of computing instruments currently available?
- Is intelligent behavior in some way a property of organic structure, and thus not achievable by nonorganic machinery?

To illustrate how far we still have to go to achieve a human level of performance, consider how much information would have to be stored in a machine to answer random questions of the following type:

If a young man of 20 can gather 10 pounds of blackberries in one day, and a young woman of 18 can gather 9, how many will they gather if they go out in the woods together?