John B. Kioustelidis

THE MECHANISM OF THINKING

Evolutionary Aspects of Thinking and of the Nature of Concepts

An Introductory Note

This is the translation in English of a book originally published in Greek in 2002 (Ιωάννης Β. Κιουστελίδης: Ο Μηχανισμός της Νόησης, Εκδόσεις Παπασωτηρίου, Αθήνα 2002). My aim in writing this book was to show that the mechanism of thinking is essentially very simple, but unformalizable, because it continuously weaves experience with conceptual development.

In a later version of this book I intend to discuss various other views about the mechanism of thinking, which I do not mention or mention only superficially here. These are certain of the views of Piaget, Chomsky, Wittgenstein, as well as the recently very popular theory of conceptual metaphors due to George Lakoff, Mark Johnson, Raphael Nunez and others.

I agree with Piaget about the sensori-motor origin of intelligence, but I don't agree with his claim that an aptitude for symbolic processing, the so called "symbolic function", appears during the second year of life. I think that this function is present already at birth and constitutes an essential part of intelligence.

I very much admire Chomsky's masterly analysis of syntax, but I disagree with him on the innateness of language, (if this is taken to mean the existence of some kind of innate rules for linguistic expression). I think that language is merely due to acquiring names for certain sensori-motor meanings and learning syntax context-dependently. An indication for this, is that, most of everyday speech is ungrammatical, meaning mainly "un-syntactical", in spite of the great importance of syntax for the production of unambiguous linguistic expressions.

I agree with Wittgenstein that language cannot be private, but I think that meanings can, nevertheless, be private (Think of the meaning of a kiss on the cheek or a slap in the face).

I agree with Lakoff and Johnson, on the sensori-motor origin of meaning, but I don't think that conceptual metaphors capture even approximately the meaning of most concepts. They are merely very superficial expressions of far deeper felt and not-at-all linguistically expressible interrelations between concepts. They cannot explain satisfactorily those extremely versatile non-logical aspects of sensori-motor thinking, of which we make constant use in everyday life. They are discussed in Chapter 11 of this book. These aspects of thinking are often misinterpreted as arising from an inborn logic. However, there is no formalizable kind of logic involved in this kind of thinking. In a recent paper about the intelligence of ravens[1], for instance, the results of certain experiments are interpreted as "a sign that they used logic" to solve the problems they faced. However, this is rather an indication of sensori-motor thinking in ravens, which indeed is omnipresent in all animals. After reading through the main part of this book I hope that the reader will be able

[1] Bernd Heinrich-Thomas Bugnyar: Just How Smart are Ravens, SCIENTIFIC AMERICAN, April 2007, pp.46-53

to understand how this kind of thinking comes about. Logic is merely a human invention accompanying language and has nothing to do with sensorimotor thinking, which is far more versatile and subtle.

I also disagree with the views expressed by Lakoff and Nunez on the origin of mathematical concepts and logic. For instance, they reduce logical thinking to visual perception of logical relations by means of a visual representation very similar to Venn-diagrams. No matter how illuminating Venn-diagrams can be, one must remember that John Venn invented them at the end of the 19th century, while logical thinking and logic have existed for thousands of years before that time. It seems strange that all previous logicians never thought of proposing this simple visualization of syllogisms, if it really relies on inborn mechanisms of thinking.

All these differences in the interpretation of the phenomena are mostly implied, but not explicitly discussed in the present book. The aim here was mainly to present a different view about the actual inward mechanisms of thinking. No systematic attempt is made to compare it with all other existing theories.

Those who might communicate their comments and criticism to me are very welcome to do so at jqstel@math.ntua.gr.

The text is in pdf-format since certain chapters contain mathematical symbols.

Athens, 31.3.2009
John B. Kioustelidis

PREFACE

The question, "What is 'thinking' and how does it operate?" sometimes occupies everybody's mind, since thought is the basic instrument through which we achieve the fulfillment of our wishes, by coordinating and directing our actions. Our ability to think is even considered by Descartes as the essence of our being. According to his famous statement: "I think, therefore I exist" ("Cogito ergo sum") the realization that we think is also a realization and confirmation that we exist.

However, this so decisive ability of man is one of the greatest mysteries which we have only minimally been able to elucidate, in spite of our great progress in all fields of Science. There are still many questions concerning it which, although very old, remain unanswered or do not have a complete and satisfactory answer:

Sudden inspirations, inexplicable feelings and impulses which take hold of us in some instances, make us wonder:

Where do such inspirations or impulses come from?

How do we get sudden hunches or reach unexpected conclusions without any obvious trace of rational thinking?

What gives an artist his inspirations?

How does a mathematician discover new mathematical truths?

How does an inventor get the idea for a new invention?

The necessity of communicating and cooperating with other people raises also many other questions:

What are the wishes and intentions of the people who surround us?

How do they look upon the events, how do they evaluate them?

Why are their views often different from our own?

What determines our own thoughts, our own views and our own wishes?

Experiencing how a baby acquires his first language, we ask ourselves:

In what "language" does he translate what he hears?

When we learn a foreign language, we initially learn how to translate it into our own language. But how does a child acquire his first language? In what "language" does he think? What is the relationship between language and thoughts?

There are also other, more disturbing questions:

Are we, in truth, logical animals, as we usually wish to believe?
But then, why do we so often act or react irrationally?
How rational are we when we fall in love?
How rational are the so called, "sport fans"?
Why only very few people are able to discuss and resolve their differences in a rational manner?

Why do we get angry because of someone's behavior instead of reacting with equanimity? When a machine breaks down, we usually don't fly into a rage but simply bring it to the repairman. Why can't we maintain the same passionless attitude with a personal relation that does not develop as we would wish?

Why do we need politicians, instead of logicians, in order to solve the problems arising in public life?

Such questions are puzzling everybody, not only philosophers and scientists, since they are connected with everyday life. They became even more prominent and timely after the invention of the computer, which is a machine that uses logic in order to process data. Since the computer gives the impression that it can replicate thought, many people keep wondering whether mental phenomena can be reduced to mechanical processes or not.

This question however, as well as many of the previous ones, has yet to be answered definitively, in spite of the great progress that has been achieved in the last fifty to hundred years, in many fields of science associated with them.

In fact, in order to answer such questions, a plenitude of disciplines has been created. Developmental Psychology, Cognitive Psychology, Psychoanalysis, Drive Theory, Ethology (Behavior Research), Psycholinguistics, as well as Neurology and theories not directly connected with Psychology or Medicine, like the theories on Discrete State Machines, Automata, and Neural Networks to name a few.

How far have all these disciplines gone in unraveling the mind's mystery?

Not very far apparently. Many scientists working in these fields seem at this time rather pessimistic about prospects of all these attempts.

This impression arises both from their own writings and from interviews given in the 1990s to John Horgan, a science writer who has investigated the achievements of the mind researchers up to now.

According to Gerald Fischbach, the head of Harvard's Department of Neuroscience, although neuroscientists keep finding new types of brain cells, neurotransmitters etc., they have not yet determined how to fit all these findings into a coherent framework.

Another neuroscientist, the Nobel laureate Torsten Wiesel, said: "We need at least a century, maybe a millenium" to comprehend the brain [Horgan, p.18].

The Harvard neuroscientist David Hubel, who won a Nobel prize in 1981 together with Wiesel, at the end of his book *Eye Brain and Vision*, states: "This surprising tendency for attributes such as form, color and movement to be handled by separate structures in the brain, immediately raises the question of how the information is finally assembled […] Where it's assembled, and how, we have no idea."

This fundamental question, sometimes called the binding problem, is a question that torments not only neuroscientists, but also cognitive scientists, psychologists, philosophers and generally scientists of all fields that try to comprehend the mind as a collection of relatively discrete "modules", or "computational units".

As Horgan [p.23] very poignantly expresses it: "Like a precocious eight-year-old tinkering with a radio, mind-scientists excel at taking the brain apart, but they have no idea how to put it back together again".

The philosopher and cognitive scientist, Jerry Fodor of Rutgers University, once a leading proponent of the computational theory of mind, notes in a review of *How the Mind Works* (1998), that certain cognitive tasks such as the ability to detect color or to parse a sentence, can indeed be reduced to computation. However, dividing the mind into many little dedicated computers or modules, still leaves unanswered the question of how the results of all these modular computations become integrated [Horgan, p.197].

Similarly pessimistic views have been expressed by people working in the field of Artificial Intelligence (see the Introduction).

Marvin Minsky, a co-founder of MIT's Artificial Intelligence Laboratory and one of the most famous scientists in this field, is also disappointed from the logical rule-based approach to AI. He observed that the definition of a bird, as a feathered animal that flies, does not apply if the bird is an ostrich or a penguin or is dead or caged or has clipped wings. He also expressed doubts about the effectiveness of neural networks, although he was one of the first to build such a network in the 1950's [Horgan, p. 223].

After such pessimistic assessments of what has been achieved up to now, what are the perspectives of mind research?

In this book, we will try to point out certain of the aspects of thinking that have usually been overlooked, possibly because they are not accessible by simple computational models. We will also give a new interpretation of the facts known about the mental system. This point of view will clarify why the rule-based approach to Artificial Intelligence fails and indicate some alternatives, which may help to develop new approaches to the problem of understanding intelligence.

Our aim is to attempt to present a coherent view of mental phenomena, as they are known today. It may not go as far as some reader would, eventually, wish but it will, hopefully, provide some new insights.

Our guiding principle in this endeavor will be consequent application of integrative, i.e. synthetic reasoning, based on the simple observation that only an evolving conceptual system can be functional already from the moment of its birth and on. We will attempt to find a concept that agrees with all known mental phenomena and does not discard some of them as insignificant, as it has often happened in the past. In this way we will be able to reinterpret the results obtained by many of the disciplines that study the mind and we will be led to a deeper understanding of mental processes. This will also provide some unusual answers to certain philosophical questions.

A final remark is necessary about the title "Evolutionary Aspects of Thinking". It should not give the false impression that the evolutionary processes considered here are directly related to theories of biological evolution. The book puts much stress on evolution, but this is not necessarily understood as Darwinian. In fact, various kinds of evolutionary processes participate in concept formation, our central subject of investigation. Some concepts are acquired by direct adaptation to the social surroundings but some others develop gradually, by a process of slow maturation. This is not a process of "survival of the fittest" form of the concept, since preliminary stages of a concept must not necessarily disappear.

The term "evolutionary" is here preferred instead of "developmental", because the last one might give the impression that the concepts evolve towards a final form following certain developmental stages. However, concept formation can be never-ending. What is more, although there are certainly different functions in the brain, they are not seen here as totally distinct. They have, in our view, a rich common underground, which provides the most prominent feature of the brain: creativity.

BIBLIOGRAPHY

Horgan, John: The Undiscovered Mind, Simon & Schuster, N.Y., 1999.

INTRODUCTION

I.1 The investigations of cognition in antiquity

The early investigations of the human mind reach a first peak with the work of Aristotle (4th century B.C.). He was the first one to give a systematic exposition of Logic and also formulated the laws governing the mechanism of associative memory.

The laws of Logic were discovered gradually by ancient Greek philosophers of the sixth, fifth and fourth centuries B.C., as instruments for resolving philosophical and forensic disputes. Looking back on an already long tradition of philosophical investigations, Aristotle was thus able to formulate precisely the basis of discourse, the laws of Logic.

Meanwhile, his teacher Plato tried to distinguish the elements of thought that are products of our mind, from those that, supposedly, preexist and are not a mere product of our thought. Thus, he created the theory that certain abstract ideas, like the idea of justice or the idea of number are not mere products of our mind, but have an independent existence of their own, consisting a separate immaterial world; the world of ideas. This is a theory that is, even today, very timely, separating scientists and philosophers into Platonists and non-Platonists (see **Appendix 4.5**).

The ability of Logic to validate a chain of syllogisms, i.e. to certify (or reject) the truth of their conclusion, is so impressive that it was thought, for a long time, to be at the core of the mental system. Logic was thus considered as an inborn ability of the human mind, in spite of opposing everyday evidence. How many people behave rationally in their interpersonal relations? How rational is, for instance, our behavior when we fall in love?

In fact, psychological experiments performed in the nineteen seventies have demonstrated beyond any doubt that, when we think, we do not apply directly any rules of logic (see [Riley & Trabasso, 1974] or [Wasson & Johnson-Laird, 1972]). As these experiments have shown, people solve a logical-syllogistic problem far more successfully when it is given in a concrete empirical form, than when the same problem is given in an abstract form, using symbols. This can only mean that the conclusions are not the result of an application of abstract rules of Logic. We would not observe such a difference in performance if people applied the same abstract rules in all cases. Logically valid results seem to be rather due to successful analogical thinking. I.e., a successful modeling of the premises is made and then conclusions are drawn, without reasoning, but by simple observation of the model. At least, this is proposed by various psychologists, e.g. Johnson-Laird.

The second investigation of Aristotle (see Aristotle: "On Memory and Recollection", especially 451b) was just as important. In fact, our ability to memorize and recall whatever is perceived by the senses, or even whole events, depends crucially on the ability of the mind to store related impressions in conjunction with each other. Otherwise, instant recall, e.g. our ability to recognize instantly a known face or landscape, would be impossible.

I.2 19th and 20th century

Further systematic investigations on the mystery of the mechanism of the human mind were not made until the middle of the 19th century. A rising interest in mental processes created, at the time, various branches of psychology, based on systematic use of observation and experimentation.

Neurological investigations also began around the middle of the 19th century, but the first description of an electrical model of the neuron, the basic unit cell of the brain, was given by McCuloch and Pitts only in the early nineteen forties. This model stimulated, from the beginning of the nineteen fifties on, various attempts to give a network model of the human memory, in analogy to the neuronal network in our brain.

Attempts to explain the huge capacity of the human memory by digital storage based on the genetic code, i.e. in a form similar to that by means of which genetic information is stored, were abandoned very early. It became very soon clear that the human mind stores memories in a different way from an electronic computer. The computer writes information in the storage medium successively, in the chronological order in which they arive. In contrast to this, the human brain stores a new memory by distributing it in various regions, connecting it with previous memories that are empirically related to it. I.e., it connects a new memory with older ones, which are somehow similar to it.

In spite of their vast calculation speeds, all serially memorizing electronic machines have a grave disadvantage: they perform miserably in pattern recognition. Neither man, nor animals have difficulty in recognizing known persons or locales instantly, because every reappearance of an old experience instantly activates the corresponding memory centers. This is, however, extremely difficult for a computer.

Thus, we were led to the conclusion that the brain is memorizing by associating new memories with old ones and not merely by storing them in chronological order. I.e., it is a massively parallel processor, that performs the processing of information practically in every storage locality and not in separate processing units.

This impression was strengthened by looking at photographs of sections of the brain's cortex, taken at various ages. The cortex is the outer layer of the brain and consists the basic storage and processing unit for memories. Such photographs show that the number of neurons apparently does not increase after birth, while their interconnections increase vastly.

I.3 Associative Models of Memory and Cognitive Psychology

These first conclusions about the memory mechanism led, in the late nineteen sixties, to the formulation of the first associative network models of memory by Quillian and Collins & Quillian.

Associative network models have been improved ever since, on the basis of psychological evidence. In fact, such models have stimulated a new way of psychological thinking; the view that no psychological theory on mental phenomena is well formulated, unless it can be given in the form of a model, which may be replicated and tested for performance on a computer.

This school of psychological thinking attempts to give ever more detailed models of mental processes, in the form of computer programs replicating the functioning of the model. Various tests can thus be performed on a computer, in order to establish whether the model performs as expected and how well it performs. This trend has also led to more precise and detailed psychological experimentation on human subjects, thus establishing how close the behavior of a given model is to the actual behavior of the mind.

I.4 The Problem of "Elementary Meanings"

Such achievements give the impression of steady progress towards a complete description of the mind mechanism. However, this impression is mistaken. The basic question is still not answered, as we shall see: what are the mental units that are connected with each other in the network model of the memory?

In all associative network models some (more or less supposedly basic and elementary) concepts are connected with each other, providing an increasingly complex network. However, all assumed elementary concepts do not seem to replicate the human storage and processing of thoughts well enough. This has even led some investigators to the reversal of their early views, i.e. to the conclusion that an analysis of meaning in components is not possible.

I.5 Artificial Intelligence

There is also another branch of investigation of human intelligence, which had raised many expectations when it started: the, so-called, study of Artificial Intelligence. A forerunner of it was the eminent mathematician of the 20th century, George Polya, who published several books dealing with the subject of how one should proceed in order to discover some way to solve a (mathematical) problem. His proposals consist of about a dozen basic rules of heuristic investigation, which are really invaluable to anybody dealing with tough mathematical problems ("heuristic" rules are techniques, which help one to guess the answer to some problem, e.g. by extrapolating from more specific cases).

Inspired by this investigation, Allen Newell and Herbert Simon attempted, already in the sixties, to program a computer so that it would try to discover ways to solve a logical or mathematical problem. They also investigated extensively how chess players proceed to find the best move, in every instance of a chess play.

It soon became clear that there are no effective general rules of strategy, enabling a computer to solve, e.g. chess problems. In later times, computers became indeed very capable chess players but this is mainly due, as it is openly confessed, on their immense speed of processing. They simply calculate all possible next moves up to a certain depth and choose the most promising one. In fact, grave doubts about the success of artificial intelligence techniques were expressed even by the members of the team of IBM's computer, Deep Blue, that defeated chess champion Garry Casparov in 1997 (see [Horgan, 1999, p.207]).

However, these psychological investigations provided an important new insight in the functioning of human thinking. It was convincingly demonstrated by Simon and Newell that experts in some field tend to organize things in their memory in "chunks" of interconnected facts and process the chunks rather than the single facts in their memory. No one is able to keep in mind more than five to seven such chunks at the same time. However the complexity of the chunks depends on increasing experience. Thus, an expert chess player sees and keeps in mind groups of interconnected moves rather than the single moves an inexperienced player sees.

This is an interesting and important fact. We can keep in mind, at the same time, either a set of five to seven unrelated letters, a set of five to seven unrelated words, or a set of five to seven meaningful sentences. It is not important that the words or sentences contain much more than seven letters, but that each time we have five to seven meaningful "chunks"[2].

I.6 Expert Systems

---

[2] In fact, this is something that many people discover without realizing its extent. I myself discovered the advantages of chunking many years ago, during my schooldays, when I had to memorize groups of chemical elements with the same valence.

I found very soon that it was much easier to build up words with the chemical symbols of the elements and memorize these words, than to try to remember all the different elements directly. Thus, I built up and memorized the words Hlinakag, Mgcabazn , Bal and Fclibr.

Hlinakag, i.e. H-Li-Na-K-Ag (H: Hydrogen, Li: Lithium, Na: Sodium, K: Potassium, Ag: Silver) are the chemical symbols for some elements with valence +1, Mgcabazn (Emgicabazn), i.e. Mg-Ca-Ba-Zn symbolize elements with valence +2, Bal, B-Al, symbolize elements with valence +3 and Fclibr, i.e., F-Cl-I-Br symbolize elements with valence -1.

Very proud of my discovery, on the next day I asked a school fiend, what "Hlinakag" might mean. Without hesitation, he gave the above explanation. He had used exactly the same method to memorize the element's valences. Many psychological discoveries are based, in fact, on common everyday experiences. These must, nevertheless, be certified by systematic experimentation, which reveals the extent and the significance of the psychological fact under consideration.

An important new effort started soon after people realized that it was very difficult to build automatic problem solvers. In the mid-seventies, they started to develop "Expert Systems" especially for diagnostic purposes.

Expert Systems are computer programs, which map in the computer's memory an expert's knowledge in some field, e.g. diagnostic medicine, in such a way that they might perform similarly well as the expert. There are today, for instance, expert systems able to perform medical diagnosis in restricted areas of medicine as well as expert systems for mechanical failure diagnosis.

However, in spite of intensive efforts to develop universal organizational systems for knowledge, no successful such system was discovered. It seems that each field of knowledge requires its own organizational schema, or rather, that each experts knowledge has its own organizational schema.

In fact, it seems impossible to explicate deeper levels of expertise. Experts are very often unable to explain by which criteria they reach their decisions (e.g. in view of insufficient data). This knowledge is unconscious and often inexplicable.

Similar experiences were reported in 1998 to John Horgan by Hayes-Roth, who works in this field since the early 1980s. According to him, extracting knowledge from human experts and turning it into software that could address real problems, turned out to be an extremely arduous, time-consuming task. What is more, the knowledge accumulated for one project usually has little or no relevance to the next. Thus, the field of Expert Systems and more generally, Artificial Intelligence, has stalled, as he also admitted [Horgan, p.201].

But what about the initial goal of Artificial Intelligence? Can we now tell a computer how to solve problems? The answer should be best given by an expert in this field of research. As Jill Larkin has said during a conference, "It just seems to be very hard to teach people to solve problems" [Morton Hunt, p.249]. If that is the case for people, how much harder is this task in the case of machines?

The difficulty of establishing successful rules for heuristic investigation, is described very nicely in an article by A.H.Schoenfeld [1980], discussing such basic rules of inference. His final remarks are that the students have to be asked to "think" and to create, rather than "recite" a subject matter. This shows that he does not feel that he has in any way reduced this subject to automatically applicable rules.

I.7 Neural Networks

A final field of investigation connected with human intelligence, which has had much acclaim nowadays, is the field of neural networks. In contrast to associative network models of memory, which rather belong to the field of psychological investigation, these are network models of much more elementary recognition capabilities of the brain, aiming towards a direct recognition of simple patterns and not merely towards a simulation of such processes.

It started with Frank Rosenblatt, who studied “perceptrons”, artificial neurons of the McCulloch and Pitts type, trying to find out how they could be used in order to solve pattern recognition and other problems.

However, in the late sixties, it was proved by Minsky and Pappert that simple topological properties, like the closedness of curves, cannot be recognized by a single layer of perceptrons. This stopped this line of research for almost twenty years, until John Hopfield revived it in the early eighties. He discovered how various kinds of artificial neurons may be connected into networks able to perform even computationally very difficult tasks. For instance, they can solve (approximately) the, so called, traveling salesman problem of optimization theory. Such a network may be called “parallel analog computer” (in contrast to digital ones) and solves its tasks by simulation, rather than computation in the strict sense.

Parallel to this, McClelland, Rummelhart and others, proved that multilayer neuronal networks are able to distinguish, i.e. separate, (digitized) patterns of arbitrary complexity, provided that a sufficient number of artificial neurons is available. Actually, a three-layer network proved sufficient for this purpose.

Thus, what seemingly failed in top-down analysis with associative networks of supposedly elementary concepts, was now tried bottom-up, by studying networks that associate, e.g. the most primitive conceivable pattern elements, like horizontal, vertical and skew bars, to form letters.

This, of course, leaves this approach with the opposite open problem; how to build up concepts by combining (organizing) hierarchically simple visual, echoic, tactile etc. criteria. We cannot explain intelligence merely by referring to neural networks, because we need concepts in order to explicate (express explicitly) knowledge.

Thus, presumably we have, at the highest levels of organization of memory, some kind of associative conceptual networks and at the lowest levels, certain specialized neural networks. But what lies between? This is a question we are going to attend to during our further study both of memory organization and the faculty of intelligence.

I.8 The evolutionary consideration of Memory and the phenomenon of Intuition

We will also raise and discuss some important questions, which are often ignored during the study of associative networks. E.g., how does a conceptual network develop into its present form?

Most studies concern themselves only with the question of how associative memory is structured. However, equally important or even more important, is how it develops into its present structure. In the progress of our discussion, we hope to convince the reader that continuing concept evolution is essentially the basis of intelligence.

There is also a prominent mystery with which many memory models are not directly concerned; the mystery of the nature of intuition, i.e. of the ability of the mind to produce sudden insights, that solve problems which long logical investigation could not solve!

Scientists and more specifically, mathematicians, experience again and again this striking phenomenon. After long hours of futile investigation, during which they have tried almost everything logic and experience have taught them, a sudden idea comes, which seems to have nothing common with what preceded it and solves the problem that has so long eluded their efforts. Some people think that these intuitions appear just by chance. However, this is not a convincing explanation, especially to people that have often experienced this phenomenon themselves.

Some of the most prominent mathematicians were so much intrigued by it that they have extensively discussed this phenomenon, describing their own and other scientists' experiences (see, e.g. the books by Henri Poincaré and Jacques Hadamard, who were both great mathematicians). They certainly did not think that they came to their insights by chance. This is not the way they describe it.

Thus, when studying possible memory mechanisms, we also have to answer the questions: "On what kind of mechanism does the faculty of intuition (e.g., mathematical intuition) reside?" "How does the mind produce those inexplicable intuitions that solve problems which long logical investigation could not solve?"

Our investigation of how the known facts may be synthesized, in order to provide an answer to these questions, will lead not only to some new insights about the mechanism of the mind, but also to some unusual philosophical conclusions.

I.9 The basic parts of the book

Our investigation begins with a short exposition of Turing's results on what computers are in principle able to do, and Gödel's results on the impossibility of automating mathematical proofs. Based on these results, we then discuss how far thinking can be automated by means of computers.

Following this, there is a chapter on the human mind's faculty of intuition (**Chapter 2**). A first discussion of this ability is given based on reports of experiences of famous scientists and especially mathematicians. A discussion of the way some mathematical concepts were created, reveals why such a process cannot be automated.

The third chapter gives an account of the basic facts known today about the neural system and especially the brain. The description of discoveries, made especially in the last fifty years, makes clear why most neurologists as well as most cognitive scientists favor today procedural semantics, as a basis for the Long Term Memory system. Procedural semantics is the position that the carrier of the meaning of simple concepts, is the synthesis of various elementary mental procedures and not the analysis of these concepts into simpler ones. In fact, certain recent discoveries seem to support evolutionary procedural semantics (procedural semantics based on continuous evolution of concepts), the central position of this book.

The organization of the neural system is elucidated in **Appendix 3.1** from another point of view, by comparing it to artificial neural networks. After a discussion of the basic features of artificial neural networks, we discuss the associative mechanisms in biological systems and their differences from those of artificial neural networks. This discussion makes clear why biological neural systems are far more effective than artificial ones.

More specific subjects arising during the discussion of a subject are generally referred to an appendix, so that the presentation does not become overburdened with details. This is especially done for various mathematical examples and discussions, which help to clarify and stress certain aspects of mental processing. These discussions may, however, be of some interest independently of the main exposition.

The following two chapters (**Chapters 4** and **5**) consider some basic theories of Cognitive Science and Artificial Intelligence and discuss how far such theories explain mental abilities, as functions of abstract information processing systems of any kind. The role of imagery in mathematical thinking is also discussed here.

The next chapter (**Chapter 6**) comes to the main position of this book, that concepts are results of a process of evolution that begins at the moment we are born. It shows how the infrastructure of concepts is related with what we call Unconscious and how concept evolution is related to the creative powers of the mind. Especially, it discusses how insights are possibly produced. As examples of such intuitive processes, two personal experiences of gaining a mathematical insight by the author (who happens to be a mathematician himself) are extensively discussed in **Appendices 6.2** and **6.3** and tentative explanations of how they may have been reached, are given in terms of the proposed mechanism.

Appendix 6.1 discusses the question, whether "intelligent" machines of sorts can be constructed, referring especially to the interesting research on experimental robots, being done at MIT's Artificial Intelligence Laboratory under R. Brooks.

The relation of language with thinking and with meaning is discussed in **Chapter 7**. On the basis of results of Developmental Psychology, it is shown how linguistic utterances are possibly produced. On the basis of results of Experimental Psychology and Ethology (Behavior Research), it is also made obvious that meaning can be conveyed even without language.

The nature of instincts and drives is discussed in **Chapter 8**, stressing the fact that they play a central role in all creative mental processes. This leads to the philosophical question on the nature of the "Freedom of Will", which is discussed in **Appendix 8.1**.

The ninth chapter discusses the relation between the proposed cognitive mechanisms and some basic assumptions, which underlie all theories referring to an unconscious part of the mind, i.e. Psychoanalysis, Analytic Psychology and other similar ones.

The last chapter (**Chapter 10**) discusses the implications of the position taken here, that certain meanings are incompatible to logical categories for the creativity of thought.

This book is, by necessity, multi-disciplinary. The reader is introduced to results from many disciplines. A host of interesting facts from many disciplines will be presented, which shed light on the functioning of the brain. These facts, as well the viewpoint adopted here, will hopefully, move the reader to think further about mental processes and investigate these phenomena in more depth. The main purpose of this book is not to present results from certain fields of science, but to present a unified, coherent view of the whole mental system, which will make certain aspects of it more understood.

The book is based on a previous paper [Kioustelidis, 1981] and two internal reports written in Brunnel University, during a six month visit to the Cybernetics department in 1984.

A Short Review

In this introduction we have first given a brief exposition of the historical development of the investigations on cognition. Further on, we saw a short description of the current efforts to investigate this phenomenon. Finally, the main parts of this book were presented and its main feature, the evolutionary consideration of the formation of concepts, was stressed.

BIBLIOGRAPHY

Hadamard, Jacques: The Psychology of Invention in the Mathematical Field, Dover, New York.

Horgan, John: The Undiscovered Mind, Simon & Schuster, N.Y., 1999

Hunt, Morton: The Universe Within, Corgi Books, London, 1984.

Kioustelidis, John: Evolutionary Aspects of Thinking, Kybernetes, 1981, vol.10, pp. 201-204.

Riley C.A. & Trabasso, T.: Comparatives, Logical Structures and Encoding in a Transitive Inference Task, Journal of Experimental Child Psychology, vol.17, 1974, pp.187-203.

Wason, P.C. & Johnson Laird, P.N.: Psychology of Reasoning: Structure and Content, Harvard U.P., Cambridge Mass., Batsford, London 1972.

A.H.Schoenfeld: American Mathematical Monthly, December 1980 pp. 794-805.

Chapter 1

CAN A COMPUTER THINK?

1.1 Turing's Contribution

This question has been a subject of hot debate since it was first discussed by Alan Turing in the late nineteen-thirties.

Turing was the first to give a theoretical description of an equivalent of today's digital computer. Roughly speaking, Turing's abstraction of a general computing device, the so-called Turing Machine (TM), may be described as an imaginary digital computer which is not liable to error and has a potentially infinite memory; say, an infinite supply of floppy disks.

Turing's main goal was to describe a machine which would be able to perform all kinds of algorithms, i.e. "mathematical recipes" in order to solve various mathematical problems[3].

On the basis of his theoretical description, Turing was able to research the main features and limitations of algorithms. For instance, as we shall see further on, he was able to show that there are simple mathematical problems whose solution can never be given through a general algorithm.

However, he was also able to show that a Turing Machine is, in a certain sense, not distinguishable from a real person. Carefully analyzing the situation, he argued persuasively that we cannot distinguish an appropriately programmed Turing Machine from a person, through dialogue conducted without direct actual (visual and acoustic) contact. An experimenter communicating through some Input-Output devices – like a computer's keyboard and screen – with an appropriately programmed TM as well as a person, will not be able to determine which is the machine. This means that a TM or a computer can always be programmed so that it can emulate a human and give appropriate answers to all questions [Turing, 1950].

We may think that this cannot be true, since we expect, for instance, a person to have a far deeper understanding of social situations or psychological reactions than a computer, no matter how well it is programmed. However, if a question is beyond the computer's "comprehension" or ability to answer, it can always give some noncommittal reply such as "I never liked ...(e.g. poetry, music)" or reply with another question, just as some people do in order to hide their ignorance. According to Turing then, a carefully programmed computer cannot be outsmarted by an examiner.

[3] More precisely, an algorithm is a sequence of well-defined elementary instructions of mathematics and logic, which provides an "automatic" way for calculating the solution of some typical problem.

This seems to imply that a Turing Machine or equally, a computer, can be in some sense just as smart as a person. However, such a hide-and-seek game is a very narrow indication of intelligence. Intelligence, as we perceive it, is mainly revealed by creativity, and in this issue a person can be far better than a computer. This can not be proved mathematically, but rather on the basis of psychological evidence. As we shall see later, human and even animal intelligence is based on a very different mechanism than digital processing of data.

Many people would be tempted to doubt that Turing's arguments are sound, because they feel that their way of thinking has nothing common with the operation of a machine.

As it happens, a dialogue similar to that proposed by Turing has actually occurred by coincidence some years ago, between a person and a computer. The human partner of this intercourse never suspected that he was communicating with a computer.

The program in the computer, then called ELIZA and later DOCTOR, was created by Joseph Weizenbaum of MIT, in order to simulate the responses of a psychotherapist encouraging people to speak about themselves. Thus, it gave non-committal answers or replied with inquiring questions, in the manner of a therapist listening to a patient account of his troubles. Its whole setting was indeed so successful, that even people knowing that they communicate with a machine tended to get emotionally tied to it.

During one notable incident, the vice-president (VP) of a company in whose computer ELIZA was installed, tried to communicate through it with a therapist who had also access to the computer, through a terminal in his home. Thinking that he was communicating with this therapist, the VP started a conversation with ELIZA, which ended up in an angry telephone call to the intended conversation partner, because the VP thought that the therapist was making fun of him (For a detailed description of the incident see [M.Hunt, 1984, pp.299-300]).

1.2 Can a Computer solve Mathematical problems independently?

1.2.1 The results of Gödel and others on the incompleteness of the theory of Natural Numbers

Since our main concern is how "intelligent" a machine can be, we must certainly consider how far one of the main fields in which human intelligence is exhibited "par excellence", can be automated; the field of mathematics.

The advancements in formal logic in the 20th century, have provided some startling results about the solvability of mathematical problems and the possibility of automating the process of proving or disproving mathematical propositions. However, in order to be more concrete, let us restrict our attention to one of the most fundamental branches of mathematics, the Number Theory, which deals with the properties of integers.

The non-negative integers 0,1,2,3,4,... are often called "Natural Numbers" (see [Stoll, 1979, p.57]), a name that reveals how elementary and fundamental they are considered. Since we learn to count and perform elementary operations with natural numbers very early in life, they seem to us as if given by Nature, as if having an 'a priori' existence, independent of and beyond all abstract mathematical theories and concepts. Therefore, the discovery, in the 20th century, that these simple objects of thought are not so simple and elementary as they seem, came as a great surprise to mathematicians.

In 1931, the Austrian Kurt Gödel produced the startling result that there is no finite axiomatic system for natural numbers. Effectively, there can be no finite number of basic principles, called axioms, which can determine whether any and all propositions concerning natural numbers, are true or false. To put it simply, there will always be an infinite number of what we call "undecidable propositions", which can be covered by the theory only through a new axiom. We call "axioms" of Number Theory a set of seven statements (see appendix 4.7), which are accepted without proof and determine the properties of natural numbers[4]. Up to the beginning of the 20th century, mathematicians believed that on the basis of these axioms all other properties of natural numbers could be derived by means of logical arguments.

Gödel upset this view. More precisely, he provided a method by means of which one can construct a proposition of formal Number Theory, which can neither be proved nor disproved by means of the existing system of axioms. Nevertheless, this proposition must be true if the system of propositions of Number Theory is free of contradictions.

Obviously, a natural reaction would be to incorporate such a true, but formally neither provable nor disprovable proposition (the mathematical terminology is, as stated previously, "undecidable proposition", "Unentscheidbarer Satz") into the already existing axiomatic system, in the hopes of achieving completeness of axioms. However, the process described by Gödel can be repeated for the new, enlarged axiomatic system, providing a new, true but undecidable proposition. This means that the number of such propositions is infinite.

---

[4] For those who are not very familiar with mathematics, I should explain that a (formal) system of mathematical axioms consists of a finite number of basic propositions, the "axioms", which are sometimes considered as self-evident, but in any case are accepted as 'a priori' valid. It also includes a finite number of rules of inference, which allow the production of new propositions by means of the given axioms. The rules of inference are the basic rules of (predicate) logic. In order to avoid the ambiguities of common language, all this system is formalized, i.e. expressed by strings of symbols having a very specific meaning.

No proposition is accepted in mathematics as valid unless it can be derived from the axioms by means of the given rules of inference.

As a matter of fact, Th. Skolem proved in 1933 that the axioms on which Number Theory is founded, are fulfilled not only by natural numbers, but also by other mathematical objects (certain classes of functions of natural numbers [Meschkowski, p.90]). In other words, the natural numbers are not the exclusive objects described by the established axioms of Number Theory. In 1934, Skolem proved that even a system of countably infinite axioms is not enough to describe the natural numbers exclusively. There are still other classes of numerical functions which fulfill the same system of axioms.

What are these propositions obtained through Gödel's method?

Although very complicated, because of the way in which they are constructed, they can be stated simply as propositions about Diophantine Equations, i.e. algebraic equations with integer coefficients, whose integer solutions we are seeking.

A typical example of a Diophantine Equation is, for instance:

$$x^2 + y^2 = z^2$$

whose integer solutions, already known to Plato, are:

$$x = m^2 - n^2,\ y = 2mn,\ z = m^2 + n^2$$

m, n being arbitrary integers.

Another famous equation of this kind is Fermat's equation:

$$x^n + y^n = z^n$$

where n is a positive integer greater than 2. Fermat noted that he had a very beautiful proof that this equation has no integer solutions. However, his proof was lost (provided that he had a valid one) and a complete proof could be given only recently, three centuries later, by Andrew Wiles.

Each proposition constructed through Gödel's method,refers to such an equation. It may state, for instance, that the equation has no integer solutions at all. However, although we can verify that the equation is not fulfilled by as many integers as we care to try, Gödel's result tells us that we cannot prove that no solutions exist, with the means provided by the axiomatic system. Alternatively, the proposition may state that the equation has an infinite number of solutions. Although we may be able to find some of these solutions, e.g. simply by trial and error, Gödel's result tells us that we can never prove that the number of solutions is infinite.

As a counterpart to Gödel's result, Alan Turing and his successors have shown that there are simple arithmetic and combinatorial questions, for which no algorithm can ever decide whether they have an affirmative or negative answer. In terms of Turing Machines, this means that there is no program which will enable such a machine to produce an affirmative or negative answer to these questions. A bit further on, we will provide such an example and explain how it can be proved.

More specifically, it was shown by Yuri Matyasevich in 1971, that no algorithm providing the solutions to all diophantine equations can exist. This last result implies, as Martin Davis [1982, p.229] observes, that in any given axiomatization of Number Theory, there is a corresponding Diophantine Equation which has no positive integer solutions, but is such that this fact cannot be proved within the given axiomatization!

1.2.2 Consequences: The Necessity of Using Heuristic Methods and even Intuition

Before giving a short description of how such surprising results can be reached, we should first discuss their significance for the problem of automating, or mechanizing thinking processes.

Obviously, some would say that there are many, e.g. arithmetic problems (actually an infinite number of such problems), which can never be solved by a straightforward algorithmic process, i.e. a deterministic, mechanical process on a computer.

Still, this does not answer whether a computer is better or worse than the human brain as a "thinking machine" of sorts. It simply tells us that a very important branch of thinking processes cannot be automated. Thus, a computer should be able to do more than apply algorithms, in order to be a successful problem solver.

One may conceive, for instance, of letting a computer "guess" its way to a solution instead of proceeding systematically, i.e. to try a combination of systematic reasoning and chance, similarly to what a human thinker would do. For example, it might guess a possible general solution from the solution of more specific forms of the same problem.

Although no general algorithm for solving diophantine equations can exist, methods for solving many specific equations of this kind have been discovered e.g. by looking at the specific structure (symmetries etc.) of the equation at hand. Thus, a computer using a certain number of heuristic rules of inference, e.g. looking for symmetries, might possibly also be successful.

There is of course, no possibility of proving or disproving that such a procedure would succeed. However, psychological evidence opposes this expectation. In spite of great efforts to elicitate general heuristic principles for solving problems, very meager results have been achieved. Automatic problem solving is still in its infancy with no promising indication that it will ever grow [5].

[5] For instance, in a recent review of a book on automated problem solving by Art Quaife [1992], Desmond Fearnley-Sander of the Department of Mathematics of the University of Tasmania concludes: "Perhaps the book is misnamed. It contains no automated development of theories. On the contrary, the development of the theories that are presented is carried out

Even automatic chess-playing programs, which have achieved a great degree of success against human opponents, have not succeeded mainly by incorporating refined strategies, but by letting ever faster computers try out immense numbers of movements, something that a human chess player could and would never do!

As a matter of fact, it seems that each field of inference requires its own heuristic rules. No general rules of heuristic reasoning, valid for all fields of investigation, have ever been established and the same holds true for the field of mathematics.

Some of the greatest mathematicians of our time have actually tried to provide heuristic rules, by which mathematical problems could be solved (refer to the excellent books of Georg Polya [1954, 1957, 1962]). However, such rules are mainly systematized common sense and not explanations of mathematical insight.

Some other mathematicians of the first order, e.g. Hadamard and Poincaré, puzzled by their own abilities, have also tried to explain the way by which they sometimes had startling insights before they went on to try to prove them by a sound chain of reasoning. It was their failure to do so, which led many of them to conclude that there exists a special faculty of the mind, which we usually call "intuition". Poincaré, for instance, states [1958, p.193]: " Logic remains barren unless it is fertilized by intuition".

Due to its importance, we shall discuss the phenomenon of intuition more extensively in a later chapter. Here we will try to give first a very brief description of how the unusual results of Gödel and his successors were produced.

1.2.3 How were Gödel's results derived?[6]

In order to do so we have to first discuss some "paradoxa" (self-contradictory propositions), many of which were produced around the turn of the 19th to the 20th century, within the realm of logic and set theory.

Possibly, the earliest paradoxon is due to Epimenides the Cretan, a Greek sage of the early 6th century B.C., who supposedly stated that "Cretans always lie" (This phrase is contained in St. Paul's epistle to Titus (1,12) and is believed to refer to a saying of Epimenides). Since this statement is put forward by a Cretan, it claims its own falsity. Thus it cannot be proclaimed to be either true or false.

Does this imply that logic has failed? We do not have to go that far. We simply have to be very careful with "self-referential" statements, i.e. statements, which speak directly or indirectly about themselves. Such statements are, by

---

entirely by the author, with a combination of skill, experience, trial and error and, above all, knowledge." [Fearnley-Sander, 1996, #8].

[6] Those who are not interested in mathematical considerations may skip this section, which gives a description of Gödel's way of reasoning and continue with the next one, which presents the conclusions.

common agreement, excluded from any systematization of mathematics and logic (more specifically, from any formal axiomatic system) in order to avoid dilemmata of the above kind.

One is thus led to distinguish between a language and a "metalanguage". A language produces sentences referring to some field of activity or field of investigation. A metalanguage produces statements concerning the sentences produced by a certain language. It is the "linguistics" of that language.

Similarly, we have to distinguish between Number Theory, a careful axiomatization and formalization of arithmetic and a "meta-theory" of it, whose objects of study are the formulas of the above theory and not the integers themselves.

What can happen when we mix theory and metatheory, is seen very nicely in the case of a variant of the famous paradox of Jules Richard (see [Kac-Ulam, p.137]).

Let us consider functions of one variable f(n) defined on the positive integers, i.e. for n = 1, 2, 3, ..., and whose values are positive integers, such as n^2 or 2n+1. For simplicity's sake we will call them "integer functions". We shall call such a function "computable" if there is an algorithm, a prescription containing a finite sequence of arithmetic and logical operations (additions, subtractions, multiplications, divisions, and e.g. "if ... then" clauses), that allows computing f(m), the value of f for any integer m.

The set of all these algorithms can be arranged according to their length and among algorithms with the same length, according to some lexicographical ordering of the letters or symbols used in the algorithm.

Thus we have the sequence of algorithms A_1,A_2,A_3,... which correspond to the functions f_1, f_2, f_3,..

Consider now the function g given by the formula

$$g(n) = f_n(n) + 1$$

This is clearly a computable function of the above kind: for any specific value, n, we first determine the algorithm A_n in the above sequence and then apply it to n thus obtaining $f_n(n)$. Increasing this value by one we obtain g(n).

Does this algorithm we have just described belong to the above sequence ?

According to our argumentation it should, since all algorithms describing integer functions are contained in it. Let us then say that it has the position m in this catalogue, i.e. that it is A_m.

The function g(n) is then merely the function $f_m(n)$, and its value when n takes on the value m should be $f_m(m)$. But, according to the above definition:

$$g(m) = f_m(m) + 1.$$

This is clearly a contradiction. What went wrong? Careful examination reveals that our argumentation relies basically on the above ordering of all algorithms describing integer functions. However, although the functions are described within the system of arithmetic, their ordering is a "metamathematical" operation.

Looking deeper, we see that everything depends on the assumption that there is a general method, an algorithm, by which we can decide which text is an algorithm for the calculation of functions of the above kind and which is not.

Thus we are led to a first indecidability (see [Martin Davis 1982 ,pp. xvi-xviii] or [Hermes, 1971 p.143-144]: "There is no algorithm that enables one to decide always whether any alleged algorithm, say any arbitrary piece of text, is an algorithm for computing the values of an integer-function."

Note however, that for specific algorithms, this question may be decidable. We may be able to show that they compute a certain function by using arguments which are appropriate for the algorithm under consideration, but cannot be applied in other instances.

In a similar manner one can show that there is no general method, no algorithm, that will tell us when a Turing Machine operating under an arbitrary "program" will stop (see [Hermes, p.145-146].

In order to make the argument easily accessible, we have here followed a line of argumentation using general terms of everyday language, instead of strictly defined mathematical concepts. A more careful way of inference would first define accurately the notion of "algorithm". However, the same argument can be carried through with more strictly defined concepts.

Gödel's derivation of his seminal results was based on a similar line of argumentation as the paradox of Jules Richard. However, in order to be able to prove his results beyond any doubt, he had to first make a careful formalization of the axiomatic system and the derivation rules of arithmetic (number theory), in order to avoid the ambiguities of everyday language. He also had to distinguish very carefully between arithmetic and metaarithmetic.

Since even a popular description of how Gödel derived his seminal results requires a somewhat complex argumentation, we will provide one for the interested reader in Appendix 1.2. Here we will only make some general remarks and we will comment on the significance of these results.

Gödel's ingenious idea was to make arithmetic talk about itself. He achieved this by mirroring metaarithmetical statements into statements of arithmetic. Thus, certain arithmetical formulas could be interpreted metaarithmetically.

How did he do this? Simply by introducing a numerical coding for all arithmetical formulas. This is not difficult to understand and can be done in infinitely many ways. For instance, as we know, all information introduced and processed in a computer, text as well as formulas, is represented by a binary code, i.e. by numbers written in a binary form. Thus, metaarithmetical statements can be reformulated as statements about numbers, the code numbers of the arithmetic formulae under consideration.

Based on this arithmetization, Gödel described a method for constructing an arithmetic formula whose metaarithmetic interpretation is that the formula with its own code number, i.e. itself, is not provable. Then he asked the question: is this formula provable in the strict derivation system of formal Arithmetic, or not?

The only possible answer is that this sentence is neither provable, nor disprovable. It can only be undecidable, since its metamathematical interpretation is that it is not provable.

If a formal proof of this sentence could exist, the fact would belie its metamathematical interpretation. A formal negation of it would also lead to a contradiction because, metamathematically, it would mean that its interpretation is wrong, i.e. that it is provable.

Thus, provided that the formal axiomatic system of Arithmetic is free of contradictions, the formula is undecidable. Nevertheless, we know by metamathematical reasoning that it is true, because it must be formally not provable, as its interpretation claims.

The formula itself is a formal arithmetical proposition referring to some property of the positive integers. Only the corresponding metaarithmetical expression has the strange content that it is formally not provable.

How can a mathematical formula be true but not provable? This can happen even in the basic axiomatic system of formal arithmetic if we ommit one of the axioms, the axiom of "complete induction", as it is called. An example of this kind is given in Appendix 1.2. In an axiom system, which does not include this axiom, we may be able to verify a formula for any natural number and yet not be able to prove it for all natural numbers.

The whole line of Gödel's argumentation does not depend on the specific axioms used. Thus, the construction of an undecidable formula can be repeated even if we incorporate the first undecidable formula as a new axiom, in the previous formal axiomatic system and this process can be repeated indefinitely. Any axiomatic system of Arithmetic is necessarily incomplete, provided that it is free of contradictions.

Gödel's result leads us to distinguish between truth and provability. What is true is not necessarily provable within a formal axiomatic system (actually it was this discovery, which he made heuristically, that led Gödel to his results as Rudy Rucker [1984, p.288] reports).

1.2.4 Indications that Intuition is an Independent Function

The main conclusion drawn from these results is, as said before, that problem solving cannot be reduced to an application of ready made algorithms. A fundamental field of mental activity, the research in mathematics,may not be simply reduced to algorithmic processes and this justifies all subsequent considerations, such as the study of cognitive processes etc. Otherwise, one might be tempted to concentrate merely on developing successful algorithms for everything.

Of course, this conclusion does not completely rule out the possibility of automated heuristic reasoning. If algorithms cannot exist, there may still exist techniques of investigation which often lead to the solution of given problems, though not always. However, as said before, inspiration is more than just systematic reasoning of any sort.

Trying to explain his inspirations, Henri Poincaré, one of the greatest mathematicians of the last two centuries, tries to find recourse in the unconscious processes. Describing how such unconscious processes operate, he abandons heuristics and instead pictures a cloud of ideas, which move about freely and sometimes collide, forming permanent connections [Poincaré, 1956, p.2048-2049].

Carl Friedrich Gauss, generally considered one of the three greatest mathematicians of all times (together with Newton and Archimedes), describing how he came by one of his discoveries states: "...I succeeded not on account of my painful efforts, but by the grace of God. Like a sudden flash of lightning, the riddle was miraculously solved. I myself cannot say what was the connecting thread between what I previously knew and what made my success possible" (see [Hadamard, p.15]). This is hardly the way a person of the highest intelligence would describe his mental processes, if he had any other way to do it.

1.3 Why the process of problem solving in other fields of Mathematics cannot be automated. The nature of many mathematical problems and how some new concepts arise

However, there is more than psychological evidence that computers cannot become good mathematicians, even with the help of heuristic techniques. mathematics is not merely the art for finding solutions to given problems. In many instances it has to invent the problems and in other cases it has to modify the concept of "solution", in order to make a problem solvable.

In the case of research in number theory, considered above, things are in a certain sense far better than in, e.g. the field of Mathematical Analysis because of the objects under consideration, the integers.

If you are investigating the solvability of a diophantine equation, you may at least try to find a solution experimentally, by systematically going through the positive integers and checking whether they fulfill this equation or not. For instance, by trying the integers 1, 2, 3, 4, and 5, one after the other, you find out that the equation: $n^3 = 4n^2 + 25$, has the solution n=5. However, if you are looking for a solution of a differential equation you can do no such thing. Even a slight variation of a function produces a new function, a new candidate for a solution of the equation. So there is no way to systematically go through all functions in order to find a solution[7].

Gödel's incompleteness theorem concerns the theory of integers and thus, indirectly, all other fields in which integers play a part. However, considering functions instead of integers, we easily conclude that we could never hope to reach a finite theory, even without this result. Not only are the objects of investigation infinitely more numerous, in the sense that a one-to-one correspondence between, say, continuous functions and integers is not possible (see Appendix 1.1), but simply because the concept "function" itself keeps on changing.

In the years right after Gödel's result, new classes of functions were invented, the so called "generalized functions", in order to solve differential equations which have no solution in the sense of classical functions, but are interesting for the development of physics and technology. There are at least two classes of generalized functions: Mikusinski's "Convolution Quotients" and Laurent Schwartz's "Distributions".

This is not an exceptional case, but something that happens often in mathematics. When mathematicians reach the conclusion that a problem is not solvable in the ordinary sense, they often invent new mathematical objects that render the problem solvable. Irrational numbers for instance, were invented in this way in order to provide solutions for algebraic equations like $x^2=2$, whose solutions cannot be expressed as quotients of integers. Similarly, the complex numbers were invented in order to make solvable algebraic equations such as $x^2+1=0$, which is not fulfilled by any real number.

Not being content with all these new generalizations of the concept of number, mathematicians have even gone on to generalize the complex numbers, creating the so called "quaternions" and generalized quaternions to the so called "hypercomplex numbers".

So Gödel's incompleteness result may be not so exciting for mathematicians after all. If they are not satisfied with their objects of study, they can invent new ones.

All this manifests what already Poincare [1956, p.2046] has pointed out: mathematics is no mechanical process and cannot bear mechanization (algorithmization). In Poincaré's words, ".. mathematical work is not simply mechanical, ..it could not be done by a machine, however perfect."

For instance, insolvability of certain diophantine equations may simply mean that we have to invent the right kind of generalized integers in order to

[7] Those who know something about set theory may note here that the set of all continuous functions is more numerous (has "greater cardinality" in Cantor's view) than the set of all real numbers, which again is more numerous (has greater cardinality) than the set of the natural numbers (see Appendix 1.1). The set of differentiable functions is at least equally numerous as that of the continuous functions, since to each continuous function we can set in correspondence its indefinite integral, which is a differentiable function.

make them solvable, just as all algebraic equations are solvable in terms of complex numbers. Such a generalization, making many diophantine equations solvable, is due to Ernst Kummer (see [Ian Stewart, 1996, pp.28-34] or [Albert Beiler, pp.276-293]).

How would an algorithm, even with elements of randomness, go about to invent a new class of magnitudes appropriate for solving an equation?

So as you can see, doing mathematics is far more than problem solving. Not only do you have to find the solution, sometimes you have to invent it.

Let us now turn to the way mathematics and physics are actually performed and discuss the startling phenomenon of intuition, as it was experienced by many scientists.

A Short Review

In this chapter we first saw that it is possible to be deceived about whether we are communicating with an appropriately programmed computer or a human being. After that however, we presented the results of Gödel and other investigators, which show that the solution of problems of the theory of Natural Numbers cannot be reduced to strict mathematical algorithms and thus be automated.

In this way, we saw that finding the solution of arithmetical problems by applying a cetain set of rules, is not always possible. For certain problems, a general method which solves the problem cannot exist. This of course, does not preclude the possibility of solving specific instances of the problem by means, e.g. of heuristic methods. Such methods make guesses on the eventual form of the solution, based either on some obvious symmetry in the form of the equations or on some common feature of the solutions of more specific cases of the same problem.

This leads us to the question whether such heuristc methods are always sufficient for finding a solution. Careful study of the nature of certain mathematical concepts, shows that they were incompatible with the concepts known up to then. They did not simply occur as combinations of known concepts, but were incompatible with them. They were actually created in order to make problems, which did not have a solution in terms of the known concepts of the time, solvable. For instance, the concept of the irrational number makes equations like $x^2 = 2$, which have no solution in terms of fractions, solvable. Due to this incompatibility, such concepts cannot result heuristically: heuristic procedures remain within the frame of the already known concepts. They do not create totally new concepts. Thus, we see that mathematicians have another capability, which allows them to create such concepts.

BIBLIOGRAPHY

Beiler, Albert H.: Recreations in the Theory of Numbers, Dover Publications, New York

Davis, Martin: Computability and Unsolvability, Dover, New York, 1982.

Fearnley-Sander, Desmond: Automated Theorem Proving and its Prospects, PSYCHE 2(28), June 1996.

Hadamard, Jacques: The Psychology of Invention in the Mathematical Field, Dover, New York.

Hermes, Hans: Aufzaehlbarkeit, Entscheidbarkeit, Berechenbarkeit, Springer Verlag, Berlin, Heidelberg, New York, 1971

Hunt, Morton: The Universe Within, Corgi Books, London, 1984.

Kac, Mark and Ulam, Stanislaw: Mathematics and Logik, Dover, New York, 1992.

Meschkowski, Herbert: Wandlungen des mathematischen Denkens, Vieweg, Braunschweig, 1964.

Turing, A.M.: Can a Machine Think?, Mind, 1950. Reproduced in J.Newman (ed.): The World of Mathematics, New York, Simon and Schuster, 1956, pp.2099-2123.

Poincaré, Henri: Mathematical Creation. In :J.Newman (ed.): The World of Mathematics, New York, Simon and Schuster, 1956, pp. 2041-2052.

Poincaré, Henri: Science and Method, Dover, 1958.

Polya, George: How to Solve it?, Princeton U.P., 2.ed. 1957

Polya, George: Mathematical Discovery, 2 volumes, John Wiley & Sons, 1962.

Polya, George: Mathematics and Plausible Reasoning, 2 volumes, Princeton, 1954.

Quaife, Art: Automated Development of Fundamental Mathematical Theories, Kluwer Academic Publishers, 1992.

Rucker, Rudy: Infinity and the Mind, Paladin, Granada Publishing, London, 1984

Stewart, Ian: From here to Infinity, Oxford U.P., Oxford, 1996.

Stoll, Robert R.: Set Theory and Logic, Dover, New York, 1979.

Chapter 2
THE PHENOMENON OF INTUITION

2.1 What is Intuition?

What is intuition (if such a faculty of the mind exists) and how can we investigate it?

Actually, this term has various meanings. For instance, we usually call "intuition" the ability to guess somebody's feelings without any verbal communication, or even in spite of contrary verbal communication. This faculty is probably due to the ability to correctly interpret "body language", the usually fleeting nonverbal signals which are sent out unconsciously (see chapter 7), or even slight changes in the intonation of verbal communication.

One may also call "intuition" the ability of a mechanic to correctly guess the cause of a machine failure based on slight, even imperceptible indications or the ability of a doctor to determine the correct cause of a patient's illness, based on equivocal symptoms. Such faculties obviously require a high degree of combinatory ability, concerning the available indications.

However, even these abilities are not what we actually mean here by "intuition". In many instances of scientific investigation, one does not merely have to guess which of a finite set of possibilities represents the correct answer. Very often, the possibilities are actually infinite, so that no answers can be obtained by merely trying them out. Sometimes one even has to invent totally new concepts in order to make an answer tractable.

It is especially this ability that we wish to investigate and we will reserve the name "intuition" for it.

Much has been written about intuition, creativity, inventivity, and other related faculties, especially after the invention of computers and the realization that algorithms have limited possibilities for solving problems, without ever reaching any definitive conclusions about the nature of these properties of the human mind.

Therefore, in order to study this ability, we must clearly define it and explain its relation especially to the term "creativity", which is also widely used.

2.2 Definition of Intuition

We define as "intuition" the ability of the human mind to make logical jumps, i.e. to relate phenomena or concepts which are seemingly logically independent from each other. "Creativity", on the other hand, is usually a more general term, incorporating not only intuition but for instance, also methodological strategies, i.e. procedural techniques, which are helpful for solving problems.

This definition of "intuition" seems to agree with Albert Einstein's and Peter Medawar's views. Albert Einstein says, in his Spencer Lecture in 1933, that the axioms which form the basis of fundamental physical theories are "free inventions of the human intellect", certainly based on empirical observations, but not determined by them. He adds that "experience may suggest the appropriate mathematical concepts, but they most certainly cannot be deduced from it" and that "there is no logical bridge from experience to the basic principles of theory" (see [Roger Newton, p.15]).

The biologist Peter Medawar writes, similarly, that "the process by which we come to form a hypothesis is not illogical but non-logical, i.e., outside logic". However, "once we have formed an opinion we can expose it to criticism, usually by experimentation" [Roger Newton, p.15].

Of course, some people might argue that intuition is not a specific property of the brain but only the result of more refined methodological strategies. However, as of yet, this approach has not been very convincing. Such refined strategies seem to be very difficult to come by, although general methodological rules are usually not difficult to set. They are products of experience and common sense, used more or less consciously by all specialists in a certain field of knowledge. If the specialist's performance relies on further, more refined rules of inference, why should it be so difficult to find them?

We should not forget that great efforts have been made, for instance, to make mathematics more tractable to students, even before people started thinking about how they could make a computer do mathematics. Why didn't all these efforts result in more refined and sophisticated strategies or heuristics? All the known basic mathematical heuristic techniques have been listed, e.g. by A.H.Schoenfeld, in only one page [1980, p.801]. There must certainly be more than that in a mathematician's mind. Why is it then so difficult to make these heuristics conscious?

2.3 Expert Systems

A tentative answer to the question about the possibility of existence of better (more refined) heuristics, can be given on the basis of practical experience. Extensive research from the mid-seventies, in the attempt to develop "Expert Systems", has shown that there are no general rules of expertise (see [Margaret Boden, chapter 7]).

Expert Systems are computer programs meant to model the expert's knowledge in some specific field, so that they could replace a human expert, e.g. in making the diagnosis for a health problem or a machine malfunction.

In spite of great efforts to elicitate general rules of "expert thinking", general heuristics so to speak, nothing of the kind has been established. Partially functional models of expertise could be constructed only by studying the specific field under consideration, with no common general rules appearing in different fields of knowledge. In addition, it proved extremely difficult to make experts explain how they came by their decisions in each particular case. Thus, the models were actually more or less representations of how a certain expert reacts to specific problems, without many generalized features (see [Morton Hunt, p.248])[8].

Along the same lines, Hunt reports [p.253] that most researchers do not currently consider transfer of general problem-solving skills to specific fields automatic, but somehow induced. He also remarks that it isn't clear how this may be achieved. He reports that "Jill Larkin likens this to teaching someone to ride a bike; it helps if you tell and show the learner what to do but most of what has to take place involves the learner trying again and again, until a multitude of experiences coalesce to form the specific skills needed".

Further on, he states: "There is also no convincing evidence that courses in problem-solving can increase one's ability to make the imaginative leaps often crucial to the act itself".

Even if we forget all objections to identifying work in mathematics with problem-solving, when we try to teach people how to solve problems, we often reach Jill Larkin's conclusion: "It just seems to be very hard to teach people to solve problems" ([Hunt, p.249]).

The suddenness and unexpectedness of some new insights are also against the idea of sophisticated strategies and heuristics as a basis for "intuition". They don't look like the result of long processing, especially since they very often occur in the middle of some other activity. One should therefore take into serious consideration the possibility that "intuition" is due to the memory's functions and organization, rather than its contents. This possibility will be discussed in **Chapter 6**.

2.4 The Role of Unconscious Processes

Certainly, as the great mathematician Henri Poincaré had already pointed out, most of the creative mathematical thinking is unconscious (see [Hadamard, 1945]). However, the "rules" on which this thinking is based are not only unconscious but, as mentioned before, also unnaturally difficult to be brought to consciousness. There is no reason to assume that some sort of repressive mechanisms (after Sigmund Freud) are operating in this case, hindering the conscious perception of this kind of knowledge. Mathematical experiences are, at least to good mathematicians, not traumatic.

---

[8] Morton Hunt [p.250] mentions certain general heuristics given in the book "Patterns of Problem Solving" by Prof. Moshe Rubinstein at UCLA. They are of the form:

- "Avoid getting lost in detail",
- "Do not commit yourself too early in a course of action",
- "Change representation"
- "Ask the right question",
- "Work backwards", etc.

Such rules however, are too general to be of any practical use when solving a difficult problem.

The main feature of what we call "intuition" is that it gives us sudden inspirations which often prove very fruitful, yet of inexplicable origin. We don't know whence they come. Remember what Gauss says: "...Like a sudden flash of lightning, the riddle happened to be solved. I myself cannot say what was the conducting thread which connected what I previously knew with what made my success possible" [Hadamard, p.15].

A similar account of personal experiences is given by Henri Poincaré in a lecture given at the Societe de Psychologie in Paris [Poincaré, pp. 2044-2045] (see also [Hadamard, pp.12-13]). He says for instance: "...we entered the omnibus to go to some place or other. At the moment I put my foot on the step, the idea came to me, without anything in my former thoughts seeming to have paved the way for it...". Similar accounts of sudden inspirations are given by many other scientists and artists, as Hadamard reports.

Many people wishing to find a rationale behind this mystery tend to take such descriptions lightly, thinking that the above scientists possibly did not try hard enough to explain how they came by these inspirations. I think that such an appreciation is wrong.

Certainly, subjective reports of psychological experiences are not in general very trustworthy, because people tend to see everything through their own spectacles, to interpret everything on the basis of their own mixture of facts and opinions. Thus, subjective accounts and interpretations of what has possibly happened, like the accounts of my personal experiences which I will give later, may be considered rather as indications of certain thought processes than as exact accounts of what happened. However, in the above cases we have exactly the opposite situation. Gauss and Poincaré do not give any description of what may have happened in their minds, but state straightforwardly that they cannot explain it.

Is it possible that they did not try hard enough? I don't believe so. The mathematician's trade is to look for connecting threads between various mathematical magnitudes and expressions. This is what he is best trained to do. If mathematicians of the highest class say that they cannot find connecting threads leading to their insights, this should be taken seriously into account.

This does not necessarily mean that we should not try to explain this phenomenon, but rather that there IS a specific faculty, a specific phenomenon, which cannot be simply reduced to some, possibly unconscious, logical chain of reasoning.

Poincaré himself attributes this specific ability to unconscious thought processes, but he also stresses that this is not due merely to trying out all possible logical combinations of facts, because the number of such combinations is simply immense and "a whole lifetime would not suffice to examine them" [Poincaré, pp.2043 - 2044]. He also stresses that "a mathematical demonstration is not a simple juxtaposition of syllogisms, but syllogisms placed in a certain order and the order in which these elements are placed is much more important than the elements themselves" [Poincaré, pp.2042-2043].

Trying to explain this phenomenon, Poincaré observes that such an insight is always preceded by extensive conscious work which, however, has been unsuccessful. Then, after a possibly longer period during which there is no conscious occupation with the specific mathematical question, the insight comes suddenly, without any previous indication that the mind was occupied with it. This leads him to the conjecture that this preliminary period of preparation produces certain combinations of facts, which are then, in a period of unconscious incubation, mixed together until a good combination of facts is found, one that leads to an answer to the question.

The guiding instrument in this selection of facts is, according to Poincaré, the mathematician's sense of beauty and harmony. The rules that determine this beauty, this harmony are, however, so fleeting and vague that they cannot be reduced to a part of a mechanical process.

Another interesting observation he makes is that the unconscious work never gives us the result of a somewhat long calculation, where we only have to apply fixed rules. In the unconscious, it seems, there is no strict adherence to rules of any kind; there is no discipline and this disorder permits unexpected combinations [Poincaré, p. 2049].

Now the concept of an unconscious part of our mind and what is more, one that is far better able to guess, to divine, is a concept which many cognitive psychologists mistrust. They have no explanation for the unconscious existence of thoughts, especially ones not related to traumatic experiences, although they also mistrust Freud's theories. For many of them, at the most, lower level mental processes which have not yet reached a conceptual interpretation can be unconscious.

That there are unconscious processes in the brain has been demonstrated experimentally, for instance by Tony Marcel. He observed the performance of people whose sight nerves had been separated at a certain level so, that they could not see with one eye. Covering their healthy eye and letting them guess what the other eye might be seeing he found a surprising percentage of successes, showing that somehow, unconsciously, they were perceiving what was before them [Marcel, 1983]. Thus, cognitive psychologists are recently compelled to accept the existence of high level unconscious mental processes.

That they mistrust Freüd's and his successors' theories is easy to explain. There are simply too many individual interpretations of the facts and even the facts are not easy to establish, since there is a very strong interaction between psychotherapist and patient. Also, not much effort is put into giving an objective description of facts, free from personal interpretations and poetical images (metaphorical expressions).

In defense of psychoanalysts and psychotherapists one can say, however, that the Unconscious is not easy to come by. It is not easy to experiment with and establish clear cut rules for its operation as is shown, e.g. by Poincaré's account of the unquestionably important processing of mathematical questions in the Unconscious.

2.5 What Can the Observation of the Process of Problem Solving Contribute?

Since our main concern here is to investigate mathematical thinking, there is another possibility. We might try to study experimentally, how people go about to solve a certain problem by conscious work. This is actually done by asking people to keep a protocol of their various attempts and give as good an account as they can, of the motivation that leads to these attempts.

Behind such efforts lies the expectation that unconscious processes are not so much different from the conscious ones ([Simon, 1966] see [Arthur Miller, p.240]). Possibly, the Unconscious, simply operates faster and more concentrated than our Conscious.

However, Poincaré does not think so. He correctly observes that the “Unconscious” is not concerned with the application of rules, e.g. of algebraic rules, since it never produces ready made results of lengthy calculations. The application of rules of any kind is seemingly left to the Conscious. "All one may hope from these inspirations, fruits of unconscious work, is a point of departure for such calculations" [Poincaré, p.2049].

Those who are following the problem solving approach to investigate intelligence, also assume that "discovery is a form of problem solving and there are no qualitative differences between […] work of high creativity and journeyman work" ([Simon, 1966], [A.Miller, p.240]). However, I think that only people with no experience in mathematical research can be mislead to make such an assumption. Inventing mathematics has little in common with the problem solving of such experiments.

First of all, problems are based on already existing knowledge and therefore, usually not so difficult as new mathematical insights. As a matter of fact, problems chosen for such experiments have to be reasonably easy to solve, so that they can be approached by conscious effort (no one asks for an answer to a really difficult mathematical question with a yet unknown answer).

Otherwise, the best one could hope for is that such questions would be answered by means of intuition, which is an ability that can not be activated by wilful effort. Thus, the experimenter would have to wait until the resolving insight comes. But even so, the problem solver would probably be unable to give any useful account of how he reached this insight, as we have seen.

This is a serious impediment to any extension of the results obtained by problem solving experiments, in the whole field of mathematical thinking. Unconscious work concentrates mostly on problems of more general nature than simple exercises of any kind.

Besides, even difficult mathematical problems have a basic difference with general mathematical investigation, as it is conducted by most mathematicians. No matter how easy or difficult the problems used in experiments are, they have exactly defined premises and conclusions, e.g. in the form "Prove that if...then...". General mathematical questions on the other hand, are based on research in a general mathematical direction and have no precisely defined premises and conclusions. The mathematician often has no specific goal, but tries to establish the "general structure" of some mathematical field and the relations that may exist between the various magnitudes appearing in it.

2.6 The Necessity of Inventing New Concepts and the Role of Faith

This disparity between problem solving and inventing new mathematical concepts is, I think, the main reason that many efforts to investigate inventivity by observing problem solving performance, has not lead to any conclusive results.

Why is this disparity so important?

In order to understand the reason, let us look at the differences between problem solving and inventing mathematics a little more closely.

A first important difference is that in problem solving one usually knows beforehand the relation to be proved. Even in cases in which a relation is not given, at least he knows that a relation between the expressions under consideration exists.

This is very important as Max Planck explained in an interview [Planck, p. 152]. According to him, scientists have to be believers, not necessarily religious believers, but believers in some universal harmony, some inward symmetry of their special field of study, which they try to reveal. They have to believe that a relation to be discovered exists in order to find it.

In order to clarify this, Planck discusses the discovery of the laws of planet motion by Kepler and compares this achievement to the work of Tycho Brache. As he states: "Brache had the same material under his hands as Kepler and even better opportunities, but he remained only a researcher, because he did not have the same faith in the existence of the eternal laws of creation"[9].

However, this should be obvious even to a layman. If you know that there is something valuable to be found in a house you have a good chance of being able to find it by systematic search as well as by detection, i.e. by considering which would be the most probable hiding places.

Your chances are somewhat worse if you don't know the nature of what you are looking for. You may know that it is valuable, but not know whether it is a diamond, a painting or a very rare postage stamp. However, even in this case you are in a far better position than somebody who lives in the house but does

---

[9] Perhaps it should be mentioned here that Brache was a very precise and meticulous observer of the sky and Kepler, who happened to asist him during the last year of his life, dicovered the laws of planetary motion based mainly on the actual notes kept by Tycho Brache, on his extensive observations about the change of the planet's position, over a period of many years.

not suspect that something valuable is hidden in it. He will never look for it and even if he sees it he may not recognize its value.

In the case of mathematics, things are even worse. Very often you consider a mathematical question, but you don't know whether it has a positive answer, whether there exists a solution and how complex this solution may be. Thus, what you are looking for is not a definite object belonging to a limited set of possible objects. What is more, very often you have to invent the objects you are looking for, as in the case of irrational numbers, complex numbers, generalized functions etc.

As we explained at the end of Chapter 1, creating mathematics involves much more than problem solving. Not only do you have to <u>find</u> the solution, but also the problem. You have to pose the appropriate questions in order to find anything useful. But beyond that, some times instead of finding the solution you have to <u>invent</u> it, to <u>create</u> a new solution-concept because, in the old terms, a solution may not exist.

This is an important observation concerning the whole phenomenon of intelligence. Often, the conceptual system is seen as a fixed set of objects which one has only to order appropriately, establishing the relations between them. In reality, however, this system is open. New concepts, which have never existed before, keep on appearing due to the evolution of the whole conceptual system, of the whole universe of knowledge.

Before discussing possible structures of the mental system and the resulting mechanisms for "intuition" or creative thinking in general, let us first consider how the nervous system, which produces all these phenomena, works.

A Short Review

In this chapter we have first considered various meanings given to the term "intuition" and ended up narrowing it down to the human ability to make logical jumps, i.e. to interrelate phenomena or concepts, which seem to be logically independent from each other. We also saw that the term is used in this sense by many scientists.

Furthermore, we have concluded that such a faculty must exist, since all efforts to eventually establish unconscious rules by means of which we reach such results did not succeed, in spite of intensive investigations in this direction.

This is seen, for instance, in the research for the creation of "Expert Systems", i.e., systems which organize the empirical knowledge in some field so that the system is able to make, e.g. fault diagnosis or medical diagnosis. This research has shown that general rules, general principles, by means of which all experts organize their knowledge, do not seem to exist. Furthermore, it proved to be extremely difficult for experts to explain how they reach their conclusions when insufficient or ambiguous indications are available.

Further on, we have seen how some famous scientists describe personal experiences of intuitive discovery. More specifically, we saw Poincaré's efforts to explain this mystery and his conclusion that there are unconscious procedures totally different from the conscious ones and not relying on logic, because they never contribute to processes which rely on the application of logical rules.

Finally, we have remarked that the attempt to observe efforts to solve, e.g., mathematical problems, cannot reveal such unconscious procedures, because the problems used are never extremely difficult. Otherwise, the problem solvers would probably reach an impasse.

A closer consideration of the way in which scientific investigation is conducted has shown that one of its fundamental elements is the faith in the existence of a new solution or physical law, as Max Planck has stressed. More specifically, in the case of mathematics it was seen that the discovery of new solutions cannot be simply the result of systematic investigation of all possible cases, because the final solution is often a transgression, i.e. a magnitude that is incompatible with the solutions considerd possible up to then.

BIBLIOGRAPHY

Boden, Margaret: Computer Models of Mind, Cambridge U.P., Cambridge, 1988.

Hadamard, Jacques: The Psychology of Invention in the Mathematical Field, Dover, New York.

Hunt, Morton: The Universe Within, Corgi Books, London, 1984.

Marcel, A.J.: Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition, Cognitive Psychology, vol.15, 1983, pp.197-237.

Marcel, A.J.: Conscious and Unconscious Perception: An Approach to the Relations between Phenomenal Experience and Perceptual Processes, Cognitive Psychology, vol.15, 1983, pp.238-300.

Miller, Arthur: Imagery in Scientific Thought, MIT Press, Cambridge Mass., 1986.

Newton, Roger G.: What Makes Nature Tick?, Harvard U.P., Cambridge Mass., 1994.

Poincare, Henri: Mathematical Creation. In :J.Newman (ed.): The World of mathematics, New York, Simon and Schuster, 1956, pp. 2041-2052.

Planck, Max: The Mystery of Our Being, pp.149-156 in Ken Wilber (ed.):Quantum Questions, New Science Library, Shambhala Publ., Boston & London, 1985.

Schoenfeld, A.H.: Teaching Problem Solving Skills, American Mathematical Monthly, 1980, pp.794-805.

Simon, Herbert A.: Scientific Discovery and the Psychology of Problem Solving, pp.22-39 in: H.A.Simon (ed.): Models of Discovery, Reidel, Boston 1977.

Chapter 3
BASIC FACTS ABOUT THE NERVOUS SYSTEM

How does the nervous system produce mental processes? There is still much that we don't know on this subject. However, the experimental research, especially of the last fifty years, has revealed many interesting facts about the mechanism of storing, recalling and processing of information by the nervous system.

These results are very important for the study of mental processes, but also because they unsettle many of the established views about thinking, as well as various mysticist theories about the function of the soul. For instance, the preservation of memories from "earlier lifes", supported by the adherents of reincarnation, is impossible, since the fundamental concepts which are necessary for the storage of information are gradually formed in a child as he grows. Without them, no memory can be recorded. Indicative of this is that, if certain abilities – like linguistic performance – have not been acquired up to approximately the beginning of puberty, it is impossible to acquire them later on.

It is not known exactly how the memories are stored in the brain, because this recording does not happen locally, but is distributed in many regions of the brain in the form of interconnections of thousands of neurons. However, the fact that the recording of new impressions implies change of neuronal connections is indubitable and has been proved by many experiments and observations (see, e.g., Merzenich's experiments, discussed at the end of this chapter).

Thus, we see that in order to understand the functioning of the mind, it is extremely important to understand as much as possible how the neural system functions and more specifically how the brain, its central part, works.

In order to understand this mechanism, first we have to look at some basic facts about the nervous system.

3.1 The Neurons

The neurons or nerve cells are, in some sense, the conductors that transmit incoming information from our sensory organs to the brain. They are also the "conductors" of orders given by the brain to various organs to act according to the goals it sets. The brain itself is nothing but a huge organized set of about 10^{11} (100 billions, i.e., 100,000 millions) nerve cells that perform the final evaluation and storage of incoming information.

However, the neurons are more than conductors. They don't merely transmit information; they also reevaluate it.

Incoming information is received first by a vast number of neurons called "receptors" or "transducers", which convert external stimuli of various kinds into electrical signals. Such receptors are, for instance, the rods and cones of the eyes, the pain-, touch-, hot-, and cold-receptors of the skin, the olfactory receptors of the nose and the stretch-receptors of the muscles. All these receptors convert stimuli coming from the external world or from the body into patterns of electrical impulses that convey information to the neural network.

Each group of electrical impulses interacts with many others already traveling through the "neural net", whose greatest part consists the human brain. This interaction results in the emission of other impulses, which reach the neurons that control our muscles and glands, providing the responses of the body.

Neurons are cells having all the characteristics of a usual cell although they differ from other cells in form, being tree-like. The most important difference is, however, that unlike the other cells of the body, most neurons do not replicate themselves in regular time-intervals. They do not divide themselves and thus can not increase their number or, at least, replace dying cells.

All mammals have almost the whole set of their nerve cells already when they are born (Cell division in most of the brain seems to stop in mice and cats about twenty days after birth, in humans after the third month of age, but in a carp it goes on until its old age. [Jouvet, 1992, chapter 1]).

From that moment on, most dying neurons are not replaced by other neurons, but the neighboring neurons may grow additional connections to compensate for the loss. It was once estimated that over 10,000 nerve cells die every day in a human brain[10].

---

[10]The nerve cells of the hippocampus (one of the central parts of the brain) seem to be an exception, at least in primates (but also in male canary birds, as it was discovered already ten years ago by Fernando Nottenbom of the Rockefeller University). As Eberhard Fuchs of the German Center for Primates in Goetingen reports, recent research of a German - American research team shows that several thousand nerve cells of the gyrus dentatus, which is a structure in the hippocampus, die every day and are continuously replaced by new ones [Der Spiegel, 19,1998, p.228]. Note, however, that the replacement of dying nerve cells is not due to a splitting of already specialised nerve cells of that region. Such a splitting would disrupt already existing nerve-connections and would therefore have rather a damaging than a restoring effect. The cells that are split are unspecialised, so called "stem" cells, which exist at the border of the gyrus dentatus. Some of the cells created in this way then migrate deeper into that region and become specialized, developing the characteristic projections of nerve cells in the same direction as their neighboring cells.

Since the hippocampus controls the storing of new information in the brain, this replacement of old neurons by new ones possibly serves the storing of new information. According to a theory of Fernando Nottenbohm, the hippocampus is an intermediate store for incoming information with limited capacity. It keeps this information either until it is permanently stored in the so called, Long Term Memory storage, or until it is no more interesting and can be discarded. Nottenbohm believes that this process of discarding useless old information and creating storage capacity for new information is connected with the process of neuron elimination and regeneration.

Experimental evidence of the generation of new neurons in the human gyrus dentatus, some of which survived for years, was obtained by Fred H. Gage of the Salk Institute in San Diego and Peter S. Ericson of the Göteborg University Institute of Neuroscience. [Scientific American, Nov.1998, pp.11-12].

An extensive discussion of these and other findings is given the article:

Kempermann, G. – Gage, F.H.: New Nerve Cells for the Adult Brain, Scientific American, May 1999, pp.38-43.

Each nerve cell (neuron) is shaped like a leafless tree, with branches and roots separated by a long trunk, the axon. While some nerve cells have axons that are one millimeter long, others have an axon that reaches from the tip of the toe to the base of the brain. A train of electrical impulses running along this axon tells the brain, e.g. that something has touched the big toe.

These electrical impulses are created electrochemically, i.e., by reversible chemical changes which propagate along the neuron's axon. Each one is only 0.1 Volt strong, lasts only 1/1000 of a second and races along the axon at speeds as high as 500 kph.

Through its "branches", called "dendrites", a neuron receives positive or negative signals (impulse trains) from other neurons, which it sums up. If the result of this evaluation exceeds a certain threshold voltage, the neuron creates a new impulse train, which it sends along the axon to other neurons.

3.2 The Synapses

In order to do so, the neuron's axon splits into a great number of fine extensions (the "roots", referred to above) and approaches the dendrites or other parts of other neurons without actually touching them. These endpoints of the extensions of the axon have the form of small bulbs and are called "synapses". There are about 10^{14} (100,000 billion) synapses in the human brain, with as many as 10^5 (100,000) found on a single neuron at a time.

Impulses reaching a synapse, cause the diffusion of chemicals called "neurotransmitters" across the "synaptic cleft" and these create electrical signals in the dendrites of the neuron on which the synapse impinges .

More precisely there are a number of molecules, called "receptors", on the surface of the next neuron, which can form chemical bonds with the neurotransmitter molecules. When such a complex molecule is created, it causes the membrane of the second neuron to allow certain substances through, such as sodium (Na), potassium (K), or chloride (Cl) ions, which change its voltage. Sodium ions cause the voltage to become positive and thus cause excitation, while potassium (or chloride) ions cause the inside of the cell to acquire more negative charge and thus cause inhibition.

In the synaptic cleft there are also mechanisms for neutralizing neurotransmitter molecules. These "deactivation" mechanisms prevent some of the transmitter molecules from ever reaching the receptors. They also "absorb" the transmitters back out of the receptors after a short time, so that these stop admitting chemicals to the second neuron. This is one way in which the effect of the impulse is limited in time: an impulse may last one millisecond itself, but its effect upon the second neuron, via the neurotransmitter it releases, may either be brief or long-lasting.

However, chemicals are not only important in producing the electric impulses in the neuron. They may also slowly change the strength of the connection between two neurons. Impulses incoming from another neuron may thus, either help or hinder the firing of an impulse by a neuron. The condition for the firing of a neuron is that within a short time period, called "period of latent summation", the sum of the excitatory impulses reaching its dendrites should exceed the sum of the inhibitory impulses by a critical amount called the "threshold of the neuron".

3.3 The Two Kinds of Synapses

Why does the body need inhibitory synapses as well as excitatory ones? Consider for instance the movement of an arm (or a leg). There are two muscles that move it; one that stretches the arm and one that bends it. If both muscles would receive signals to contract at the same time, the arm would be brought to a condition of rigidity, disabling any movement. Thus, the neuron that transmits an excitatory message to one of these muscles also sends an inhibitory to the other one, preventing its operation while the other muscle moves.

The impulses are initiated at the initial segment of the axon, which sets the rate and pattern of the impulse train and since they tend to decay, they are replicated (boosted) at certain locations along the axon (called Ranvier nodes), whenever the message must travel more than about 1 mm. Every 1 mm along the neuron's axon, on average, there is a booster station which reproduces the impulse. Thus, each impulse sets off another impulse 1 mm down the axon, with about a 20 microsecond delay (20 millionths of a second). Then, that impulse sets off yet another impulse at the next relay station, 20 microseconds later and so on.

Therefore, to cover a distance of 1 meter it takes a total of 20 milliseconds (1/50 second), which means that the speed with which the message travels is 50 meters per second (180 kph).

Actually, the distance between booster stations varies, as does the delay before the booster station responds with its own impulse. So speeds vary, between 1 to 150 meters per second.

The transmission of a signal is, in some sense, a relay race, as Calvin and Ojemann [1980, p.15] call it. Not a single neuron, but a whole chain of nerve cells connects the transducer cell in the skin with the brain. However, the message is not simply passed from one neuron to another. Every neuron in this process synthesizes a new view of things. It combines all the thousands of inputs it receives from other neurons, playing off one influence against the other and thus composing "a new story". This modified message is moved to the next neuron in the chain. So the whole process is not so much like the transmission of a message, but rather like the spread of a rumor [Calvin - Ojemann,1980, p.115].

3.4 The Neuronal Signals and their Propagation

A neuron does not merely act as a transmission line, but also as a processor, adding and subtracting influences from many inputs and sending its new message on to many other cells. If we wish to compare the nervous system with a computer we should keep this important difference in mind. The nervous system is in some sense a vastly parallel computing device, since every transmission line is also a processor.

The arrival of impulses on the dendrites of a neuron actually determines the firing of its axon at a slightly later time, since there is a small delay between a period of latent summation and the passage of the corresponding axonal impulse to its endbulbs. Also, after an impulse has traveled along an axon, there is a time called "refractory period" during which the axon is incapable of transmitting an impulse.

We may, therefore, imitate a neuron by dividing its time scale into consecutive intervals of a length equal to one refractory period of the given neuron (which is in the order of a thousandth of a second) and then specifying for each such interval whether or not a voltage spike was generated in it.

Consequently, we usually simulate the neuron by a binary switching device that can be switched on or off in successive time intervals, depending on the states of its inputs in the previous interval. This highly simplified model of the neuron was introduced by Warren McCulloch and Walter Pitts in 1943.

Let us now look at the course followed by incoming signals until they reach the last processing stage in the brain's cortex. The second neuron in the chain from transducer to the brain is typically located in the spinal cord. There are hundreds of transducers in the skin and the muscles converging towards this cell in the spinal cord, plus thousands of inputs from other neurons in the brain and spinal cord. Some produce negative voltage changes ("inhibition"), while others produce positive changes ("excitation").

As said before, if the voltage balance is sufficiently in the positive direction, another impulse is triggered and speeds along the axon of the second nerve cell towards the brain. This balance also determines the rate at which impulses are produced, which varies between 0 and 1,300 impulses per second (a typical value is 300 impulses per second). This rate informs the next neuron in the line what the balance was. The strength of the excitation is thus encoded by the frequency of the impulse train and not by the width (strength) of the individual pulses.

There is a chain of at least four neurons from the transducer neurons of the sensory organs, e.g. of the skin, to the corresponding region of the cerebral cortex, the sensory strip.

After the transducer neuron in the skin, which may sense pressure, the movement of a hair or temperature, across the synapse comes a neuron in the spinal cord or in the brain stem. The third neuron in the chain is often in the thalamus, a structure in the central part of the brain and the axons from the thalamus go up to a fourth neuron in the sensory strip of the cerebral cortex[11] .

In addition to this direct chain, there may also be a relay in the brain stem's reticular activating system, going first to the cerebellum and arriving on the sensory strip[12] from there. Each of the senses has a similar set of direct and indirect pathways from the transducers to the cerebral cortex.

The cerebral cortex is the end in one direction of the relay race of signals, starting at the sensory organs, while its is also a starting place for backward projections transmitting response signals to the muscles.

However, this view is not very accurate. Forward projections are not singularly heading toward the cortex because from the vicinity of each neuron projecting forwards, there are also backward projections. Instead of a stream of signals moving forward there are rather loops of feedforward and feedback projections. Besides, various parts of the left- and right- cortex must cooperate in order to produce a response to an incoming message. The cortex of each hemisphere contains centers that combine incoming information from the senses with memories and already existing thoughts, in order to evaluate a new situation. What is more, incoming signals go first through the thalamus, the central part of the brain and are relayed from there, not only to the cerebral cortex, but often also to interior centers of the brain, which control the basic functions of the body.

There is a good reason for this intervention of lower centers of the brain: certain situations require an immediate reaction, before an extensive evaluation by the two cortices can be performed. Thus, thinking is a highly complex process taking place at the same time (or almost at the same time) in many parts of the brain.

The inputs to the cortex often prefer certain layers. E.g., the messages from the eyes arrive only in layer IVc of the visual cortex. Of these, the most poorly understood are the ones closest to the brain's surface (I, II, and III).

The cortex plays an important role in learning and remembering. So let us see how memories are recorded.

3.5 The Memory

---

[11] Cerebral cortex is called the surface layer of the two hemispheres of the brain, which is no more than 4.5 mm deep (usually about 3mm deep) anywhere and contains about 14 to 17 billion nerve cells of various types, usually divided into six layers numbered I-VI, and some times also in sub-layers, e.g. layer IVc. [Georgieva, 1989, p.334].

[12] Brain stem is called the extension of the spinal cord into the brain, while cerebellum is the separate structure which lies just below the brain and behind the brain stem. One of the basic functions of the cerebellum is to assist the coordination of movements and posture.

Studies of the psychology of memory indicate that its function includes several processes:

1. An "immediate memory", which maintains a sensual input for just a few seconds if it is not reinforced. This memory is usually tested by asking somebody to repeat a long string of numbers, which has been read to him/her. Most people can keep in the immediate memory a string of about seven items, e.g. the digits of a telephone number.

2. Certain inputs of the immediate memory are then passed to the "short term memory", which maintains recent experiences for some hours (e.g. what we had for breakfast) even in the presence of various intervening distractions.

3. "Long term memory", which is used to remember information permanently, e.g. our name, our address, our past or knowledge about the surrounding world. Long term memories are not formed immediately after a new piece of information is acquired. New memories are held in short term memory. The more permanent long term memory is formed gradually over a period of many hours or days, as animal experiments suggest.

The general arousal of our attention, as well as the choise of its target, so that various stimuli may be processed and stored in memory for a shorter or longer period, is controlled by the so called "reticular system", which we will discuss a little later.

It should be noted here that there is some debate about what should be called "immediate" and what "short term" memory. Possibly the mechanism underlying both is the same, with the only difference being that memories are reinforced in order to stay in the short term memory.

However, there are many observations indicating that a different mechanism is used in the brain for long term memory storage. For instance, short term memory can be disrupted by something which interrupts brain function: a blow to the head, an epileptic seizure or a brief interruption of the brain's blood supply. However, this amnesia does not extend to long term memories. Even deep coma, which may silence much of the activity of the brain, does not erase long term memories.

This seems to indicate that there are physiological changes involved in long term memory, for instance the creation of new synaptic connections (due to persistent repetition, i.e. rehearsal), while short term memory and immediate memory may be based merely on electric activation of certain neural connections.

A given neuron cooperates with many other neurons in order to form the trace of a particular memory and storing permanently a particular memory may involve changes in many neurons, possibly involving various specialized regions of the brain.

One should note that although all or almost all neurons of the brain already exist at the time of birth, the mass of the brain at birth is only about one fourth (1/4) that of the adult brain. The brain's size increases because neurons grow considerably in size and the number of axon projections and dendrites increases, as does the number of their connections. This indicates that permanent storing of memories is based on plastic changes, like the growth of dendrites and building up of synapses.

This view is also supported by the observation that the neural connections develop from an immature pattern, only roughly approximating the adult one and need neural function in order to reach the adult form. Children must be stimulated through touch, speech and images to develop fully. Babies, who spend most of their first years of life lying in their cribs, develop abnormally slowly.

A model for long term memory storage has been based upon the observation that, in early infancy, some neurons send out many short branches (called spines) whose number decreases again by half in late infancy. This model assumes that even adult neurons are continually budding and sending new branches out to make random connections with axons in the vicinity. A new synapse is supposed to be then maintained, if its neurons are activated during the mental processing of subsequent events, with it participating in this process. Synapses that are not activated during a short period after their creation, e.g. for a day, are possibly absorbed again. Thus, permanent memory traces are supposed to be built up by conserving those randomly made connections which happen to participate in response to a new event.[13]

On the other hand, short term memory may involve the temporary strengthening of a synapse due to extensive use, which increases the amount of neurotransmitter released by an impulse on a given day. Possibly, this also protects new synapses from being absorbed again.

Sleep seems to play an important role in the processing of the day's experiences, i.e. in reinforcing what should be maintained and eliminating everything else. This is especially true for the periods of dreaming, which are accompanied by rapid eye movements (REM). If a subject is awakened every time he starts a period of REM sleep, there will be more REM sleep the next night. But if such REM sleep deprivation is kept up night after night, the subject's daytime performance will deteriorate much more than if the awakening had occurred during deep sleep.

---

[13] It is worth noting here that the number of synapses reaches a maximum value between the ninth month and the second year of age, depending on the region considered, when the child has 50% more synapses than an adult. The brain’s metabolism reaches the adult’s level about the ninth or tenth month of age and continues increasing until the age of four years. Then it starts decreasing until it reaches again the adult’s level during adolescence [Pinker, 1995, p. 289].

We can recall items from our long term memory astonishingly fast, because the retrieval system uses cues to help us track them down. For example, hearing the name of a place we have visited may bring back many associated memories. However, memory is never an exact recording of pictures or sounds, but rather a selective recording. Different people will remember the same event quite differently.

Part of the human thalamus, which is the innermost part of the brain, seems to help focus the attention, thus allowing access to short term memory and changing the "intensity" with which incoming information will be remembered. Studies in animals and man also indicate that the "hippocampus", a structure in the inner side of each temporal lobe of the brain, plays a central part in the short term memory process .

In contrast to this partial localization of STM (short term memory) mechanisms, it has been quite difficult to associate any particular brain structure with long term memory. There seems to be no area of the brain whose damaging will lead to loss of long term memory, although it is possible to identify major areas of the brain which play a basic role in various mental abilities. Local damage to the cortex, e.g. due to a stroke, can of course be cause of deterioration or elimination of some fundamental ability. For instance, it may impede the ability to remember the name of an object, though we may still know what is its use, or the ability to form syntactically correct sentences.

3.6 The General Structure of the Brain

The most remarkable anatomic feature of the brain is, of course, that it is split into two roughly equal hemispheres, each connected with and controlling the opposite to it side of the body. Each of the eyes is connected to both hemispheres, but in a way that the right visual field of each eye (the right part of the view we see) is connected to the left hemisphere and the left visual field of each eye (the left part of the view we see) to the right hemisphere.

Between the two hemispheres there is a large number of nerve fibres (about 200 million) which connect them, called the corpus callosum, allowing the exchange of information and the coordination of their actions.

Although they are similar, these hemispheres perform very different tasks. Most people have a verbal left brain hemisphere and a visual - spatial and emotional right brain hemisphere.

Another outstanding feature of the brain's surface is that it is deeply lined. Thus, its total area is about 20 times that of the corresponding cranial area that covers it. Each hemisphere is outwardly divided into four parts called "lobes" by two deeper grooves called "central sulcus" and "lateral fissure". The central sulcus runs from the top of the hemisphere downward, while the lateral fissure runs from the bottom of the hemisphere backward and slightly upward. The part in front of the central sulcus is called "frontal lobe", while the part immediately behind it is called "parietal lobe". These parts lie above the lateral fissure, while below it, just on the inside of each one of the temples, is the temporal lobe. Farther back, behind the parietal and temporal lobes is the occipital lobe, not so distinctly separated from them by anatomic features. Just below the temporal and ocipital lobe lies the cerebellum as a separate unit.

The arrangement of all these parts is roughly the following:

Frontal	Parietal Lobe	Ocipital
Lobe	Temporal Lobe	Lobe
	Cerebellum	

3.7 Functions of the Cortex: Localization of Linguistic and other Abilities

We can localize many important mental functions on the brain's cortex through the observation of stroke victims and other brain damaged people, as well as through experimentation with patients during brain operations. Recently we have started using Positron Emission Tomographic (PET) scans, which are obtained by injecting in the bloodstream radioactive water (containing an isotope of Oxygen). Scanning the increase in radioactivity, we can trace the increase of blood flow and therefore of neural activity in various regions of the cortex, during the performance of mental tasks.

This does not mean that we know exactly what function is performed by each part of the cortex. More elementary functions, like the control of various parts of the body or even the production and understanding of language, are easier to locate (although we cannot localize individual concepts or memories) than more synthetic and abstract ones. Thus we know very little about the frontal lobe, because it is concerned mainly with synthetic mental activities, like planning ahead and predicting. However, it has been established that the lower part of the left frontal lobe controls the articulation of speech. We also know that a thin strip of the frontal lobe just in front of the central sulcus controls discrete movements of the body. This strip is called “motor control zone”. Injury to this area of the brain’s surface can cause paralysis of the opposite side of the body. Corresponding to it, just behind the central sulcus, is the sensory zone that receives sensory inputs from all parts of the body. The sensory zone and motoric zone correspond to each other; the various parts of the body are mapped on them in the same order.

These zones are primary regions of the outer brain connected with the inner brain by neuronal projections (dendrites and axon projections). Surrounding each of these zones there are regions primarily concerned with corresponding memories (visual memories just in front of the area controlling vision, acoustic memories around the area that controls hearing etc.). See [Adolf

Faller 1988, p. 351, fig. 222]. These secondary regions are called "association regions" and consist 80% of the surface of the outer brain. They connect the primary regions with "association fibers" and play a central role in information processing. It is the role of especially these regions that is still not clearly understood.

The upper part of the temporal lobe is related to the sense of hearing (damage to it can result in impaired hearing or deafness), while the rest of the temporal lobe integrates multiple sensory functions, like auditory, visual and tactile.

The part of the parietal lobe just behind the central sulcus controls the sensations of pain, temperature, touch and pressure while its lower part, called Wernicke's region, is connected with the reading ability.

Finally, the occipital lobe is mainly dedicated to the processing of visual information. Damage to it results in partial or complete blindness.

Let us look more closely at some of these functions of the cortex and how they are distributed.

As said earlier, all linguistic abilities (using and understanding speech, manipulating grammar etc.), as well as the ability to read, are situated in most people in the left brain hemisphere, which also plays a fundamental part in the ability to regulate the sequence of movements of the face and hands. On the other hand, the right brain hemisphere, in most people, controls the ability to manipulate things in space, follow maps, remember shapes, faces and musical tone sequences. We can also say that in most people the left hemisphere deals with analysis while the right one deals with synthesis.

How do we know all this? From the observation of people with brain damages. Left-brain damage disturbs language in about thirteen people for every one person with language disturbance due to right-brain damage. Right-brain damage usually interferes with the ability to manipulate objects in space or follow routes, but only about twice as often as left-brain damage and more often in males than in females. I.e., males are more likely to have visual-spatial abilities strongly concentrated in the right brain than females.

Roughly 99% of all right-handed people have the control of language in the left hemisphere and the same is true for about two thirds of the left-handed people, who are 15% of the population. This brings the total number of people with left-brain language localization up to 94% of the population. A few left-handers have language on both sides of the brain, which makes them prone to language disorders, particularly stuttering.

This localization of abilities does not already exist at the moment of birth. Young children show less evidence of function localization in the left or right hemisphere and they have the ability to relocate functions to compensate for brain injury, but only in the first years of life (approximately up to the age of five years). However, the ability to use the full range of language seems to rely on neural connections which already exist at the time of birth. The full range of speech and grammatical expression cannot develop if the left hemisphere is removed in the first months of life, while removal of the right hemisphere does not cause any impairment in the use of language.

This removal of one of the brain’s hemispheres becomes necessary by a congenital disorder, in which abnormal blood vessels develop on just one side of the brain and it is usually done before the age of six months. In this case language develops normally only if the right hemisphere is removed. Those who lose the left hemisphere develop speech, but not fully. They are quiet people who speak only when necessary and then with a reduced range of grammatical expression, e.g., in present tense. More elaborate grammatical expressions are beyond their abilities. These people also suffer a deterioration of visual-spatial abilities, usually greater than that of their linguistic abilities.

A surprising result of neurological investigations is that at sites where phoneme recognition can be disrupted, stimulation of the cortex with electrodes will also interfere with the mimicry of sequential oral-facial movements, while the converse is also true: stimulation of brain sites that control a sequence of mouth movements disrupts not only these, but also recognition of speech sounds or phonemes. This observation supports the “motor theory of speech perception”, which claims that understanding of a speech sound has more in common with the mouth movements that produce it than with its acoustical properties, i.e., that understanding the spoken word involves the creation of an internal model of how to speak the heard word [Calvin-Ojerman, 1980, p.37].

Different aspects of language seem to be located in different places of the language cortex and their location relative to each other is generally the same, although the exact location of such subdivisions varies from patient to patient.

However, two different languages spoken by the same patient may not use exactly the same area of the brain. Strokes in bilingual people may impede the ability to use one language without hindering the usage of the other.

The two main areas of the brain that control language are Broca's area, and Wernicke's area .

Broca's area is usually located in the left frontal lobe of the brain, near the region of the motor cortex that controls the movement of the lip muscles. It controls speech, as it was deduced in 1861 by the French surgeon Pierre-Paul Broca. It also handles syntax, because damage to it results in non-grammatical, yet meaningful speech and impairs the understanding of a sentence only if its meaning depends critically on syntax.

Wernicke's area is located in the temporal and parietal lobes of the left hemisphere between Heschl's gyrus, the primary receiver of auditory stimuli and the angular gyrus, which plays a critical role in integrating visual and auditory stimuli, i.e. in the ability to read. Wernicke's area controls language comprehension, i.e. semantics, as it was established in 1874 by the German neurologist Karl Wernicke. He observed that damage to this area resulted in grammatical but nonsensical speech and inability to grasp what other people say (however, this condition tends to be mild unless the damage extends to surrounding areas).

Language comprehension is impaired even in users of sign language who have a stroke in this area. They show impaired signing and also impaired understanding of signing. Part of Wernicke's area seems to deal with naming things, while another part permits us to repeat spoken words and name colors.

Broca's and Wernicke's areas are connected by a nerve bundle called the arcuate fasciculus. If it is damaged (for instance, by a stroke), the patient can understand words but cannot repeat them.

Localization of language areas can differ from person to person, with only the rear third of the frontal lobe - the part controlling motor output of speech - staying constant.

The right-brain plays a major role in music, although musical abilities seem to be distributed on both sides of the brain. If it is put to sleep by barbiturate injections, the subject loses the ability to reproduce pitch correctly and sings in a monotone. Damage, particularly to the right temporal lobe, disturbs the memory for tone sequences and loudness, while rhythm memory is relatively unaffected.

On the other hand, damage to the left temporal lobe may affect rhythm and diminish the emotional response to music, although it does not alter the memory for tones or loudness.

The areas of the left-brain concerning music are separate from those concerning language.

3.8 The Cooperation of the Various Areas of the Brain

Let us see how all these parts of the brain cooperate: sensory inputs are first directed to the sensory area of the brain's cortex. They stay there for about 20 seconds and are then permanently lost, unless they activate the hippocampus in the inner side of the temporal lobe of the brain. The hippocampus activates various parts of the cortex in order to store memories for a longer time period. New sensory inputs can modify the already temporarily stored ones. Stored memories can be recalled by activating the various parts of the cortex that contain partial information and this process is coordinated by the frontal lobe of the brain.

For instance, the mental processes, by which we grasp and lift a glass of water are roughly the following:

First, the visual memory region of the parietal lobe is activated, providing information about the position of the glass relative to the body, i.e. creating some kind of a map indicating the relative positions of hand and glass. Then, the premotoric area, just in front of the motor zone, is activated in order to form a plan of action. This plan is executed by the motoric area of the cortex in cooperation with the cerebellum, which coordinates the finer movements of the body. At the same time, the basal ganglia adjust the position of the body in order to facilitate the movement of the hand.

The basal ganglia are a group of large nuclei lying in the central regions of the brain's hemispheres and partially surrounding the thalamus of each hemisphere. They are involved in the control of movements.

The coordination of sensations and actions when someone plays chess, for instance, is revealed by PET scans of the chess player, produced (as mentioned before) by injecting radioactive sugar in the organism and following up how the blood flow increases in various brain regions, when they are activated. At first, the visual area of the cortex is activated, as the player notes the changes in the positions of the figures. Then, an area of the left temporal lobe is activated, which contains the playing rules. Then, parts of the right frontal lobe are activated, as the player plans his own move.[14]

Similarly, according to Wernicke, when we wish to express a thought, at first the appropriate words are put together in Wernicke's area and then they are sent through the arcuate fasciculus to Broca's area, where the appropriate syntactic – grammatical form is produced and the speech movements of the lips and tongue are determined.

3.9 The Visual System and its Functioning

3.9.1 The Structure of the Visual System

As mentioned before, each of the senses has a four-neuron direct pathway from the transducers to the cerebral cortex, as well as an indirect one going through the reticular activating system. What happens in each of these stages of message transmission and reprocessing is not always well understood. However, it has been established in greater detail in the case of vision.

The first few stages of the neuron chain in the visual pathway are located in the retina at the back of the eye, which has a multilayer structure.

The photoreceptors, i.e. the transducer neurons, are the in this case of two types: rods (110 - 125 millions) and cones (6 -7 millions). The cones provide day vision and color perception and their density decreases towards the periphery of the retina, while the rods provide night vision and their density increases towards the periphery.

---

[14] A more recent technique for localizing brain-functions is to use Magnetic Resonance Imaging (MRI) [Tim Beardsley, 1997]. The MRI machine bombards the brain with radio-frequency waves that excite hydrogen atoms in the blood stream, causing the atoms to emit signals revealing in which parts of the brain the blood concentration is higher. This corresponds to higher neural activity in these parts. While PET scans resolve brain areas lying only about one centimeter apart and are too slow to distinguish changes lasting only a few seconds, MRI can resolve the position of active neurons to about two millimeters and is fast enough to distinguish changes in neural activity which are seconds apart.

The contrast between adjacent regions of the visual world is enhanced by the actions of additional cell types, such as the octopus shaped cells in the retina.

The second and third neurons of the chain are also in the retina, while the axon of the third neuron, called the retinal ganglion cell, runs all the way from the eye to the thalamus. The axon of the fourth neuron runs from the thalamus up to the visual cortex in the back of the brain.

The first neuron, the rod or cone, is very small; it responds, producing an electric signal, only when light shines directly on it. The image of the visual world, projected upside down on the back of the eye, corresponds thus to a mosaic of voltage changes in transducer neurons.

The second and third neurons in the chain start the process of comparing neighboring regions. Such a neuron responds best to a small spot of light, covering a number of adjacent rods and cones. If the light spot moves, the positive voltage changes into a negative one, while if it moves even farther away, the voltage becomes very small. So the transducers in the center cause excitation, while those in the surrounding ring cause inhibition. If both areas are illuminated, as in the case of diffuse light, the positive and negative influences cancel each other, sometimes totally, as if the light were not there at all. In this way, all visual images are somewhat changed by emphasizing the borderlines between darker and lighter areas.

The fourth neuron, the one in the thalamus, discriminates even more against diffuse light, but it may respond equally well to a spot or a line and not be able to distinguish between such symbols as ., 1, /, -, +,etc. This is also true for the fifth neuron, located in the visual cortex.

From the sixth neuron on, however, a voltage change can be evoked only if the borderline has a certain orientation each time. Some sixth order neurons prefer vertical lines, others horizontal, while others some particular angle. If the line is rotated 10° from the preferred orientation, the impulses are likely to cease in one cell and start up in another group of neurons.

The borderline must also have the correct place in space in order to evoke an impulse train in a certain sixth order neuron. In this way there is a correspondence between locations of the visual field and locations of the visual cortex. I.e., there is a map of what we see, spread out across the visual cortex.

The seventh order cells often have even more complex responses. They are also particular about the orientation of a line but not as particular about its position in space. Some of them however, called hypercomplex, are particular about the line's length. They may respond, e.g. only to short, horizontal lines.

Thus, the neurons of the visual system gradually resolve the photographic visual image presented to the mosaic of the retinal transducers into small pieces of information, e.g. by extracting information about the borderlines between lighter and darker areas. When we look at a chessboard, most of the responses of the visual cortex are, for instance, devoted to the lines forming the edges of the black and white squares, while much fewer cells are needed to process the information about the centers of the squares.

Although the visual cortex is wired up from birth to accomplish all this, the wiring is greatly modified by the experiences of infancy. How does this modification happen? Consider, for instance, the perception of depth: the two eyes see slightly different views of the world, since the one is shifted relative to the other and this difference is useful for judging distances. The information from the left and right eye is not combined until the sixth order nerve cells, at least in the visual cortex of primates. At lower levels of the visual system, a cell either responds to light shown in the left eye or light shown in the right eye. But in the cortex there are cells responding to signals from both eyes.

Connections providing this ability exist already at birth in the visual cortex, but they can be lost if the infant does not have an opportunity to use both eyes together. This was proved by experiments with infant monkeys. A frosted contact lens was switched daily from one eye to the other so that the two eyes never had a chance to work together. Then the cells which would ordinarily respond to both eyes specialized only in one eye. Even if the contact lens was then removed and the monkey was given normal visual experience for a long time, the situation did not change. There were very few coordinating nerve cells.

It seems that there is a critical period during infancy (in monkeys, about the first nine months after birth) when experience with the environment makes permanent connection changes like these; this is explained by assuming that during the critical period there is competition among nerve cells in the visual cortex.

Fifth order cells specialize in either the left or the right eye (never both) and normally there are in the visual cortex alternating territories about 0.4 mm wide, each containing cells representing only one of the eyes. However, if the frosted contact lens is left on one eye for a while, during the first six months of life, the usual equal representation of left and right eyes in the narrow band in layer IVc of the visual cortex, where the fifth order cells are located, will change. The territories from the covered eye shrink and those for the active eye expand, as if the active "exercised" cells had won a competition for space in the brain. To understand this, we need only remember that the number of dendritic spines increases greatly during the first few months of life and then drops by half in the following months.

3.9.2 The Visual Perception

We can now understand how the brain “perceives” an object seen by the eyes. This is done by relating the signals of light and movement encoded by neurons at the back of the brain, to those neurons dealing with memory and even emotion. Thus, our perception of what we do depends partly on our experience.

This explains why we are confused when we see a picture of an object, with the dark and light areas reversed. Obviously, what experience tells us should be the figure, now looks like a background. Even more confusing are ambiguous pictures, interpretable in two different ways. Consider, e.g. Esher's picture "Day and Night" or the well known figure, which shows either an old or a young woman.

Figure 3.1: Ambiguous picture: A young girl or an old woman?

Since our minds can attend to only one image at once, our perception shifts back and forth between the two, as we try to make both fit an image existing in our mind. However, this ability of the brain to incorporate incoming information into mental models, is very important. It allows, for instance, the filling in of gaps in what we perceive, so that we can make sense of it.

Some evidence that perception indeed involves matching real objects with mental images is found in a 1980 report of the English neuropsychologist E.T.Rolls. He found three small brain regions connected to the visual cortex, each dealing with a special aspect of perception. In the hypothalamus of a monkey, Rolls located neurons that emitted impulses when the monkey saw an apple, but not when he saw a rubber ball. They seemed to be stimulated only by the sight of something edible and nothing else. Then, near the thalamus and hippocampus of the monkey's brain, Rolls found "recognition" neurons. They fired when the monkey saw a familiar object (a big red ball), but ignored a small black ball, an object the monkey had not seen before. Finally, in the brain's sensory-motor cortex area, Rolls found neurons that could tell one familiar object from another. They disregarded red balls or food, but responded to faces, monkey or human, large or small, any color, upright or inverted [Diagram Group, 1984, p. 259].

3.9.3 On the Unconscious Processing of Visual Data

The ability of the brain to process visual information, even unconsciously, is demonstrated in "split-brain" operations. In these operations, which are performed in order to cure certain types of epilepsy by preventing the spread of abnormal neural excitations, the connections in the corpus callosum are severed.

If, after this operation, objects are shown to the right of the visual field, so that their images are transmitted to the left (the language-) hemisphere, then the person can correctly see them and say what they are. On the other hand, if they are placed to the left of the visual field, then the person will deny having seen them. Amazingly however, he can use his left hand to write a description of the objects, which he still denies having seen! It is almost as if we have two people, perhaps we should say two sorts of consciousness, in one brain.

3.10 The Systems of Emotional Arousal and Control

3.10.1 Sentiments and Personality - The Role of the Frontal Lobes in the Control of Emotions

Comparing the relative sizes of various parts of the brain in primates and man, we find the biggest size difference in the frontal lobes. Yet we know less about the frontal lobes than any other area of the brain.

Damage to both frontal lobes seems to alter two things: the ability to alter behavior with changing experiences and the degree of emotional responsiveness. In a typical experiment, a subject is asked to sort a deck of cards into two piles according to the responses of the examiner. If, halfway through the deck of cards, the examiner changes the piling criterion, a normal person soon catches on that the rules have changed and figures out the new strategy, but a person with frontal lobe damage usually simply continues with the first strategy. With more extensive frontal lobe damage, the ability to abstract is lost.

In addition to this loss of flexibility in dealing with changes in the environment, patients with frontal lobe damage lose emotional responsiveness. They become apathetic, showing neither happiness nor grief and seem not to be concerned with the effect of the environment on them, or with the effect of their own behavior upon others.

The brain areas whose damaging alters the emotional responsiveness, seem to be the same that are concerned with intestinal function, regulation of the heart rate, blood pressure, respiration, digestive activity, and the levels of various hormones. Thus, there seems to be a biological relation between emotions and intestinal reactions, which suggests a biological basis for "psychosomatic" diseases.

More precisely, visceral (= intestinal) regulation in man depends on two systems; the sympathetic and the parasympathetic nervous system. The sympathetic nervous system prepares the body for fighting or fleeing. It raises the heart rate and blood pressure, erects the hair and decreases the digestive activity. The parasympathetic nervous system does the reverse, preparing the body for more vegetative activities.

The central controlling mechanisms for these two systems are located in the hypothalamus, which lies below the thalamus (the relay station of incoming signals) and above the brain stem, the extension of the spinal cord. The hypothalamus also regulates the hormones via the pituitary gland which lies just below it.

Damaging the hypothalamus in animals alters their emotional reactions. When a small area of the outer part of the hypothalamus is destroyed, cats will change from being peaceful and easily handled to being aggressive, responding with blind rage to any contact. But there are also areas in the hypothalamus whose electrical stimulation seems to create pleasure. When animals are given a choice between food, water, sex and activation of these areas by pressing a bar, they press the bar until they are exhausted.

Emotions of rage and pleasure caused at the hypothalamic level are little altered by any environmental stimuli. Interaction between environment and emotion occurs at the cortical level, at those cortex locations where visceral functions can also be altered, such as the inner surfaces of the frontal lobes and the temporal lobes. These areas, called the "limbic system"', surround the brain stem and the corpus callosum like a sickle (limbus = border) and have connections with the hypothalamus.

Extensive damage to the temporal lobe part of the limbic system can make an animal (or man) respond to any environmental input with rage. Damage to the frontal lobe portions of the limbic system can lead to placidity and indifference.

The mixture of emotional character and thought, which is called personality, seems to rely on the interaction of wide areas of the cortex with the limbic system, with the right brain hemisphere generally more strongly involved with emotional experience and perception than the left one.

A related important observation is that affective disorders, like depression or mania, as well as thought disorders, like schizophrenia, have hereditary biological components that are clearly distinct from environmental influences. Studies of early adoption of brothers by different foster parents, show that the tendency to develop a psychosis depends on the biological parents rather than the adoptive parents. Identical twins separated at birth have roughly similar chances of developing a major mental disorder.

3.10.2 The Arousal and Control of Emotions - Amygdalas versus Prefrontal Lobes

As recent research reveals, the center of emotional arousal (excitation) in the brain is the pair of amygdalas (each in one of the two hemispheres). These are almond-shaped clusters of interconnected neurons located above the brain stem. They belong to the limbic system that encircles the brain stem .

The hippocampus and the amygdala are the two key structures of the limbic system, which (together with all other interior parts of the brain) are considered to be the first, primitive brain of mammals, from which the cortex and the neocortex (the evolutionary more recent part of the cortex) gradually evolved. The neocortex consists of the outward layers of the cortex.

Actually, three developmental layers are distinguished in the brain: the brain stem and the surrounding structures are considered to correspond to the first, most primitive brain and are often called the "serpent brain". The layer containing the limbic system, the thalamus and hypothalamus is considered to be the evolutionary extension of the brain in lower mammals and the cortex is the evolutionary addition that characterizes the brain of higher mammals.

According to Paul MacLean, director of the Laboratory of Brain Evolution and Behavior at the USA National Institute of Mental Health, we actually have not one, but three interconnected brains, each having its own way of perceiving the surroundings and responding to them. Each has its own memory and sense of space and time [Restak, pp. 50-52].

The limbic system is close to the rhinal cortex, which is situated on the lower inner surface of the temporal lobes. This is the part of the cortex that evaluates olfactory (= smell) information, but is also important for learning and memory. Damage to the rhinal cortex of both hemispheres causes severe memory loss of events that occur after this damage (amnesia). Amnesics live permanently in the present moment, because they forget information and events within a few minutes. However, they can still acquire skills.

This close connection of smell with memory is not as strange as it may seem at first glance. In lower mammals, the cerebral cortex is small and almost entirely dedicated to the evaluation of olfactory information, which is very important for an animal. It may reveal the proximity to some potential food source, but also the approach of an enemy much earlier than they can be seen or heard. For that same reason, the rhinal cortex is also very close to the center of emotions, the limbic system, since emotions control the behavior (fighting or fleeing) of the animal.

Even in higher mammals, most of the brain's learning and remembering is controlled by the two basic structures of the limbic system. The hippocampus coordinates learning and remembering of the dry facts, while the amygdalas retain the emotional accent that accompanies them. Severing the amygdalas from the rest of the brain results in inability to judge the emotional significance of events.

The amygdala is the center that arouses not only the mild emotions, like affection or dislike, but also all passions like hate or love. If the amygdalas are removed from an animal it exhibits neither fear nor rage and loses all urge to compete or cooperate.

This was first discovered by Joseph LeDoux, a neuroscientist at New York University. His research has shown that sensory signals, which always travel first to the thalamus of the brain, are routed from there for further processing not only to the sensory processing areas of the neocortex, which is the center of thought, but also directly (across a single synapse) to the amygdala. The amygdala (as well as other parts of the limbic system) receives the results of the recognition process through the neocortex and then sends appropriate response orders to other parts of the brain and the body. But the direct connection to the thalamus allows the amygdala to begin to respond before the neocortex has completely evaluated the situation. This is important for an animal living in the wild, because a quick fight – or flight – response is necessary for survival. The same is also true in a civilized surrounding if an unexpected danger appears, e.g. a car turning a corner and rushing towards us. In such an emergency, the amygdala can make us spring aside in a reflex movement, before the mind (the neocortex) has fully appreciated what is happening and decided upon a plan of reaction. In a rat, the amygdala can initiate a response to an outward stimulus in 12 milliseconds (0,012 sec), while the route from the thalamus, which goes first to the neocortex, takes about twice as long.

The existence of a direct route has been proved experimentally with rats whose auditory cortex had been destroyed. These rats received electric shocks paired with an echoic signal. Then, upon the sounding of the signal alone, they still tried to avoid the shock although the sound could no more be processed by their cortex.

3.10.3 The Existence of Unconscious Emotions

The possibility of unconscious emotional reactions in people is demonstrated by another experiment, reported by Goleman [1996, p.60]: when pictures of snakes are shown to people who fear snakes, sensors placed on their skin detect sweat breaking out, indicating anxiety, even though they say that they do not feel any fear. The sweat breaks out even when the picture is presented so rapidly that they are not conscious of what they have seen. If we continue this emotional arousal, feelings of fear become at some instance conscious, even though the person is unaware of what caused them.

There are, therefore, two levels of emotion: conscious emotions and unconscious ones. The moment we begin to be aware of our feelings seems to be the moment when those feelings start to be processed in the prefrontal lobes, the part of the frontal lobes which lies just behind the forehead.

The fact that some emotional reactions can be formed without any conscious participation at all, shows that the amygdala can retain memories and response repertoires that we enact without realizing what we are doing. Our responses are, in this case, activated through the direct route from the thalamus and not the indirect one from the neocortex. This can often lead to inappropriate reactions, since direct responses from the amygdala are based on quick first impressions, pleasant or unpleasant associations of an outward stimulus and not on full consideration. The amygdala does not have a sufficiently large memory to keep all details of a past pleasant or unpleasant situation. It therefore reacts precipitously to superficial resemblances.

Emotionally intense experiences are imprinted in the memory with exceptional strength and vividness, because the brain uses for their storage the same system of neurochemical processes that alerts the body for reaction (see [Goleman, 1996, p.23]). We remember pleasant or unpleasant experiences (e.g., our graduation day or our first date) vividly, while we soon forget neutral ones.

This does not mean, however, that we are always conscious of the reasons for our emotional reactions, because the amygdala can cause them before we can remember with what past experiences they are connected. If the emotional arousal is very strong, then we are unable to think clearly and recall to mind the reason for our reaction. In such instances the amygdala hijacks the whole mental system.

In addition to that, according to LeDoux, emotional experiences and reaction schemata from our very early childhood are necessarily unconscious, since they are stored in the amygdala before the conceptual system has been sufficiently developed to give them precise linguistic expression.

The amygdala of a growing child is sufficiently developed much earlier than the hippocampus and the cortex, which are important for maintaining narrative memories. Thus, very early emotional memories have no matching set of articulate thoughts when they are triggered in later life, because they were established before we had words to express our experience.

This seems to support the psychoanalysts' position that pleasant or unpleasant experiences of early childhood may considerably influence our later behavior without becoming conscious.

Our emotional reactions are not always impulsive. While the amygdala produces impulsive reactions, there is another part of the brain which makes a more careful emotional evaluation of the existing situation and leads to more moderate reactions. This part is the prefrontal cortex, which lies just behind the forehead. It is vital not only to emotional control but also for maintaining the temporary store of information, which is called "working memory" and is needed in order to deal each time with the situation at hand.

The prefrontal lobes are the seat of planning and organizing actions (including emotional reactions) in order to achieve a goal. This is done with the cooperation of other centers of the neocortex that register, analyze and comprehend incoming information. All this activity however, as well as the appropriate reaction to the existing situation, is coordinated by the frontal lobes. More specifically, the prefrontal cortex weighs reactions before acting and dampens the activation signals sent by the amygdala and other limbic centers, if necessary. Rash emotional reactions and accompanying actions are apparently caused by the amygdalas only when extreme urgency of reaction seems necessary. In such cases they hijack the whole reaction mechanism.

3.10.4 The Control of Positive and Negative Emotions

Of the two prefrontal lobes, the left one seems to moderate unpleasant and distressing emotions, while the right one is the seat of negative feelings like fear and aggression. The left lobe probably keeps such emotions in check by inhibiting the right lobe. Thus, stroke patients with damaged left prefrontal cortex are prone to catastrophic worries and fears, while patients with damage to the right prefrontal lobe are "unduly cheerful". The left prefrontal lobe is the seat of optimism, while the right one is the seat of pessimism.

The role of the prefrontal lobes in emotional control was first suspected by neurologists in the 1940s, when misguided prefrontal lobotomy operations "cured" the patients from fits of rage, but also led to the complete vanishing of emotional reactions, a total indifference to what was happening around them.

In order to appreciate the importance of the role of prefrontal lobes, one should note that emotions are not merely important in sentimental matters, but also in judging everyday situations and making appropriate choices and decisions.

3.10.5 The Role of Emotions in Thinking

Without emotional evaluation, no information is useful. Loss of emotional responsiveness is like loss of taste. Just as we are no more able to choose between foods without the sense of taste, when we lose the ability for emotional evaluation, what is happening we are no more able to choose our course of responding.

As Antonio Damasio, a neurologist at the University of Iowa says, feelings are indispensable for rational decisions, because the world often confronts us with a whole host of possible choices. This is the case not only in important matters, like how we should invest our money or whom we should marry, but even in simple ones, like what we should eat today or which place we should choose for vacations or even which route we should choose in order to reach some destination.

Some such problems, e.g. the choice of an optimal route, cannot be solved by mere rational thought, simply because they involve too many unknown parameters. Thus, we have to make our choices based not on concrete logical arguments, but rather on "feelings" (which are possibly based on past positive or negative experiences). Other problems, however, are completely beyond rational thought because they are a matter of taste or personal preferences. Whether we should spend our vacations in the mountains or at the sea, is not a problem we can solve by considering various rational arguments in favor of the one or the other destination, but rather by considering how we feel about these two destinations.

Thus, emotional choices accompany all our thoughts and rational decisions, although we may seldom think about it consciously. On the other hand, rational analysis may play an important role in resolving emotional conflicts.

An indication of the important role of emotions in thinking is the result of observations made by Damasio on a patient, whose prefrontal lobe was damaged due to necessary tumor surgery. The patient was a successful corporate lawyer prior to the operation. After the operation, however, he was unable to hold his job although his intelligence was not influenced by this operation (measured by standard IQ tests). The reason was that he was no longer able to make even the most rudimentary decision and organize his life. He wasted his time considering minor details and seemed to have lost all sense of priority (see [Goleman, 1996, pp. 57-59] or [Damasio, 1994]).

There is also a third center that controls emotions, besides the amygdalas and the prefrontal lobes. This is the somatosensory part of the cortex of the right hemisphere, the location that receives from the body both external touch-, pain- and temperature- sensations, as well as internal sensations of pain, visceral state and position of the joints. Damage to this part may also reduce emotional responsiveness and even the ability to appreciate somebody's present condition. People with stroke damage in this region seem unable to grasp that they have lost control of some part of their body. This phenomenon, called "anosognosia", is characteristic for patients with impairment of the left side of the body, which corresponds to the right brain hemisphere.

3.11 The Activating and Selective Attention System

Mental activity would be impossible without some mechanism that arouses the brain and causes it to give selective attention to some incoming information. These two functions are performed by a mechanism, which is known from animal experiments to reside in the brain stem (the extension of the spinal cord) and is called the reticular (= net-like) activating system.

Arousal and selective attention are two distinct components of the system, which are both essential for higher brain functions. The first regulates the overall arousal and the various parts of the sleep-wakefulness cycle, while the other provides selective attention.

3.11.1 The General Arousal

The arousal part of the system is located in the upper part of the brain stem and acts like an amplifier placed in parallel to sensory inputs. When turned on, this amplifier boosts the sensations in order to "wake up" the cerebral cortex; this is possible because sensory nerves route branch lines through the reticular formation, besides their main paths to the cerebral cortex. Also, many of its cells reach the thalamus and from there, other cells fan out around the brain: to the hypothalamus, the corpus striatum (another name for the basal ganglia), the cerebellum and different regions of the cerebral cortex. Thus, various stimuli directed to the brain make the reticular formation start firing signals at different targets around the brain, sustaining or measurably increasing the electrical activity of the cerebral cortex.

Brain stem injuries may damage this arousing mechanism and cause unconsciousness. Irreversible damage produces permanent coma and sometimes death.

3.11.2 Selective Attention and the Process of becoming Conscious – the "Inner World"

The second component of the activating system plays a more specific part in causing selective attention, instead of generalized arousal. It allows sustained concentration, i.e. focusing our attention on one task at a time, such as reading. Only an unusual or imperative stimulus, such as a loud noise or a telephone call, is then able to distract us. What we do at each moment is determined by this faculty of selective attention. Part of this component of the activating system is located in the thalamus, with the left thalamus concentrating the attention to verbal processes and the right to visual-spatial information.

Since the selective attention mechanism seems to be the system activating or deactivating brain areas in relation to the external or internal world, some neurologists speculate that this system plays a major part in conscious experience. It may determine what we observe in the surrounding environment and the intensity with which it is remembered. For instance, we may drive and think at the same time, being only very hazily aware of the environment, while we are fully aware of the problem we try to solve. At the same time we change gears almost "unconsciously".

This system seems to also play a major part in the creation of a model of the present situation, incorporating both external and internal experience by switching from external stimuli to internal ones, that result from associations already existing in memory. Models of this kind are important in considering possible response strategies, e.g. alternative ways to drive home, which are then often stored and used appropriately when necessary.

Yet we should not overlook that "consciousness" is not primarily a focusing of our attention on a certain task. Its main feature is self-awareness. Consciousness refers to an internal model of ourselves: our past history, what we think we are, our present and ultimate goals, what we wish and what we feel, including instinctive drives and feelings not expressible in words. No matter what we perceive, think or do, this self-reference is a process that accompanies all these activities. For instance, awareness of the surrounding is awareness of our presence in a certain surrounding.

An interesting psychological question which can be resolved in this context is that of the existence of an inner world. Everybody is aware of having an inner world, e.g. "inner speech" and private thoughts. But how could we establish the existence of an inner world in someone else, by direct observation? Such an opportunity is presented if the selective attention system malfunctions. For instance, in patients with hemorrhages in the left thalamus, while attention on the external world is sustained, words from the internal world intrude into external speech in an uncontrollable way, much to the patient's surprise and frustration.

The reticular activating system's selective attention mechanism does not merely switch between external and internal stimuli; it also regulates the flow of information into the brain. This is done by a series of backward connections (towards the sensory organs) which act in a way that adjusts the sensitivity of every sensory system, increasing or decreasing it depending on the novelty or irrelevance of a particular sensation. More precisely, microelectrodes implanted in the brainstem's reticular formation, in the system involved in general arousal, reveal so called "novelty recording cells'" that fire only in response to a new stimulus. If there are ten to fifteen repetitions per second, the response of these neurons stops. This explains why we immediately notice when somebody turns on a radio, but soon lose awareness of this sound. Thus, the brain is prevented from being continuously bombarded by the touch sensations produced by our clothes, but still able to detect the new touch of a mosquito. The same system also regulates pain sensations.

Here we should note that the ability to discern between novel and repetitive sensations is so important that the amplifier controls in the reticular activating system are not the only mechanisms in the nervous system performing this task. Mechanisms which minimize repetitive sensations exist at many levels, beginning with mechanisms existing in single nerve cells. In fact, intensity controls do not exist only within single neurons, but are also the result of neuron systems which regulate the internal settings.

A similar system is supposed to operate during learning. Since synapses (or other parts of neurons) are modified by usage, there may also exist systems of neurons that regulate this plasticity. For example, operative destruction of the hippocampus in both sides of the brain (which was performed in order to relieve epileptic seizures) revealed the central role of this system in forming short term memories. The destruction caused absence of short term memory and thus inability to form new long term memories, although immediate memory and long term memory for events some years prior to the operation were normal. The patient was unable to recognize somebody who had just left the room, when they returned. He could converse normally with the hospital staff, but he did not remember them though he saw them every day.

It is thus hypothesized that one role of the hippocampus in forming short term memories is to regulate the facilitation properties of synapses elsewhere in the brain. Therefore, after the destruction of the hippocampus, the synapses elsewhere may fail to receive the necessary permission to be modified by use and a new memory trace cannot be established.

3.12 The Learning Mechanism

3.12.1 The two kinds of Learning and their Neuronal Mechanisms

Researchers basically distinguish between two types of learning: the explicit or declarative, that requires a conscious record, and the implicit or nondeclarative that has no conscious record.

Explicit learning is fast and may take place even after only one training trial, permitting storage of information about a single event. This learning is severely impaired by damage to the hippocampus but also to the temporal lobes.

In contrast, implicit learning is slow and accumulates through repetition over many trials and does not utilize conscious participation. It is basically the acquirement of skills or reflexes, so that the subject is unable to describe just what has been learned. This kind of learning seems to be a direct formation of associative links between sensory and motor neurons, by plastic changes occurring between them. This view is supported by the fact that implicit learning is not impaired by lesions of the temporal lobe, in contrast to the explicit one.

Donald Hebb proposed that associations, e.g. of conditioned stimuli with reflexes, are formed by coincident neural activity which strengthens the connection between the presynaptic and postsynaptic neuron. Such a mechanism was discovered only in 1973 by Terje Lomo and Timothy Bliss as part of the mechanism of "long term potentiation".

Ladislav Tauc and Eric Kandel discovered another associative mechanism in 1963. They found that the synaptic connection between two neurons (of the marine snail "Aplysia"[15], could be strengthened without activity of the

[15] This snail was selected because it is relatively large, about 15 cm and because it has a relatively simple neural system with only 20000 neurons while, e.g. a bee has more than one

postsynaptic cell, when a third neuron acts on the presynaptic neuron. The third neuron, called a "modulatory neuron", enhances neurotransmitter release from the terminals of the presynaptic neuron, if the electrical impulses (known as action potentials) in the presynaptic cell are coincident with action potentials in the modulatory neuron.

The modulatory neurons act on the sensory neurons, enhancing transmitter release from their terminals. The process, called presynaptic facilitation, induces a non-Hebbian (i.e., not dependent on pairing of stimuli) form of learning, called "sensitization", in which an animal learns, e.g. to enhance a variety of defensive reflex responses after receiving an unpleasant stimulus.

Actually, to effectuate conditioning, it is necessary that the conditioned (not unpleasant) stimulus that is correlated with the unconditioned (unpleasant) stimulus, precedes it by a certain critical period, for instance 0.5 second. If the time interval separating the two stimuli is lengthened, shortened or reversed, conditioning is drastically reduced or does not occur. Obviously, the chemical changes that strengthen a certain connection are effective only during specific time intervals, which must overlap, and recede afterwards.

Conditioning is an implicit mechanism of learning. What about explicit forms of learning? Do they also depend on neural associative mechanisms? If they do, it must be in an indirect way, because explicit learning requires processing and evaluation of the stimuli. Besides, such processes are most successful when the two events that are associated occur simultaneously.

Timothy Bliss and Terje Lomo demonstrated in 1973 that neurons in the hippocampus have remarkable plastic capabilities, capable of providing a mechanism for explicit learning. They found that a brief high-frequency train of action potentials in one of the neural pathways within the hippocampus produces an increase in synaptic strength in that pathway, that lasts for hours in an anesthetized animal and for days and even weeks in an alert animal. This strengthening was called "long term potentiation" (LTP). Experiments suggest that long term potentiation uses a combination of the two independent associative, synaptic learning mechanisms mentioned above.

Particularly important for memory storage is the hippocampus, a structure within the temporal lobe. However, as mentioned before, lesions of the hippocampus interfere only with the storage of new memories, leaving the older ones intact. This suggests that the hippocampus is only a temporary depository for long term memory, processing the newly learned information for a period of days or weeks and then transferring it to relevant areas of the cerebral cortex for more permanent storage.

Both explicit and implicit memory storage proceed in stages. Storage of the initial information is a type of short term memory, lasting minutes to hours and involving changes in the strength of already existing synaptic connections. The long term changes (those that persist for weeks and months) are stored at the

---

million. What is more, the neurons of this snail have a thousand times larger cross sections than human neurons.

same site but they require the activation of genes, the creation of new proteins and the growth of new connections [Kandel and Hawkins p.60].

Since long term memory is based on anatomic changes, does that mean that the brain is constantly changing anatomically as we learn and forget?

This was exactly what Michael Merzenich of the University of California at San Fransisco and his colleagues demonstrated. They examined the representation of the hand in the sensory area of the cerebral cortex. Neuroscientists believed until recently that this representation was stable throughout life. But Merzenich and his colleagues have shown that cortical maps are constantly modified on the basis of the use of the sensory pathways. They encouraged a monkey to touch a rotating disk with only the three middle fingers of its hand. After several thousand disk rotations, the area in the cortex devoted to the three middle fingers was expanded at the expense of that devoted to the other fingers. Practice, therefore, can lead to changes in the cortical representation of the most active fingers. This supports well the view, which we will defend here, that concepts are steadily subject to modification on the basis of new experience.

More details about the underlying mechanisms of explicit and implicit learning will be given in Appendix 3.1, which discusses the properties of artificial neural networks and compares them to biological associative mechanisms.

3.12.2 Learning and Conditioning

An early observation of psychologists was that certain instances of learning depend on the association of a stimulus with a response by the brain. Thus, the Russian physiologist Ivan Pavlov discovered that dogs not only salivated at the sight of food, which is a simple reflex act, but they also learned to salivate at the sound of a bell, if they repeatedly heard it just before the food appeared. Pavlov called this a “conditioned reflex” and, oversimplifying things, he claimed that conditioned reflexes form a basis for all learned behavior.

Another kind of conditioning was described in 1930 by the American physiologist Burrus F. Skinner. Unlike Pavlov's classical conditioning, which involves involuntary, i.e. automatic change in reflex activity, this conditioning, called “operant conditioning”, causes change in voluntary behavior.

Skinner showed that a reward reinforced a response to some stimulus. Thus, rats that accidentally pressed a lever learned to do so deliberately, if that act produced food.

Skinner had also the tendency to oversimplify things. Ignoring the inborn ability and personality differences, which modify all incoming outward experiences, he claimed that the mind is a “tabula rasa”, a neutral memorizing mechanism, which is only formed by education, i.e. by appropriate conditioning.

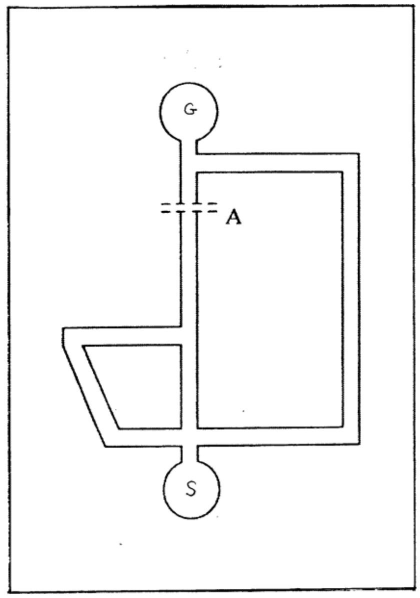


Figure 3.2: Tolman’s maze, which indicated that rats use mental models
S: Starting point, G: Goal

This view was proven incorrect already in 1930 by Edward C. Tolman. He placed rats in a maze with three available ways to a box with food. Soon the rats learned to use the shortest and easiest way to the food box, a straight line. But when Tolman blocked the end of this route, which was also the end of the second in length, the rats did not try first that second route, but took directly the third, the longest one. Obviously, they had built a model of the maze in their minds and knew that the second route would also be blocked. This shows that learning is not merely a stimulus-response association, as Skinner and his followers claimed. The brain is actively participating in the evaluation of experiences.

In any case, learning certainly produces changes in the brain's structure. Neurons can change the strength with which they respond to incoming signals and so modulate their own transmitted impulses, a process that goes on all the time. This also involves growing new neuron branches. Research with octopuses suggests that learning a conditioned reflex strengthens neuron links with small inhibitory neurons, so that when one pathway has been used repeatedly, inhibition blocks other routes. Thus, the same visual signal eventually triggers an unvarying response. Similarly, scientists of the University of California at Berkeley found that rearing rats in a complex, stimulating environment, produced marked increases in the brain's weight and protein content and in the activity of enzymes controlling a neurotransmitter, while understimulated rats showed no such gains. Research also showed that learning coincides with the brain's manufacture of specific compounds known as peptides.

3.13 How are Concepts Represented in the Brain?

A widely held view today is that there are no permanently retained "pictorial" representations of objects or persons, as was traditionally thought, but records of the neural activity that takes place in the sensory and motor cortices during interaction with a given object. This view is called "Procedural Semantics", since the "meaning" of a concept is represented by a procedure rather than some kind of definition. Memories are also seen as records of mental procedures, i.e. as patterns of synaptic connections that can re-create the separate sets of brain activity that define an object or event, while each record can also stimulate related ones.

As A. R. and H. Damasio [1992, p.65] state: "The neural processes that describe the interaction between the individual and the object constitute a rapid sequence of microperceptions and microactions, almost simultaneous as far as consciousness is concerned. They occur in separate functional regions and each region contains additional subdivisions: the visual aspect of perception, for example, is segregated within smaller systems specialized for color, shape and movement".

Where can the records that bind together all these fragmented activities be kept? The Damasios believe that they are embodied in ensembles of neurons within the brain's many "convergence regions". At these sites, the axons of feedforward projecting neurons from one part of the brain, converge and join with reciprocally diverging feedback projections from other regions. When reactivation within the convergence zones stimulates the feedback projections, many anatomically separate and widely distributed neuron ensembles fire simultaneously and reconstruct previous patterns of mental activity.

When storing information, the brain also classifies it so that related events and concepts (shapes, colors, proximities in space and time as well as pertinent body movements and reactions) can be reactivated together. Such associations are, according to the Damasios, recorded in another convergence zone.

A Short Review

In this chapter we have given a brief description of what is known about the brain mechanism from the point of view of Neurology, as well as Experimental Psychology.

At first, we saw the structure and functioning of the main neural cells. After that, we examined the functions of the cerebral cortex and more specifically, how the linguistic and optical regions function to the degree that we know up to date. Then, we considered the role of deeper centers of the brain in connection with corresponding regions of the cerebral cortex. We saw the role of the amygdalas in emotional arousal as well as the role of the prefrontal lobes in the control of these emotions. Here, we also noted how important the emotional evaluation of situations is in making decisions.

Further on, we saw the role of the Reticular Activating System in causing the general arousal of the brain, as well as controlling the selective attention to certain outward stimuli. Finally, we studied the neuronal mechanisms of learning, as well as the related role of acquired reflexes, as it follows from psychological experiments. We ended with a brief consideration of how concepts are, eventually, represented in the brain.

BIBLIOGRAPHY

Beardsley, Tim: The Machinery of Thought, Scientific American, August 1997.

Calvin, William H. - Ojemann, George A.: Inside the Brain, Mentor Book, New American Library, 1980.

Damasio, Antonio: Descarte's Error. Emotion, Reason and the Human Brain, G.P.Putnam's Sohn, New York, 1994.

Damasio, Antonio R. and Hanna: Brain and Language, SCIENTIFIC AMERICAN, Sept.1992, pp. 63 - 71.

Diagram Group: The Brain, a User's Manual, New English Library, London 1984.

Faller, Adolf: Der Koerper des Menschen, dtv 1988

Gellatly, Angus - Zarate, Oscar: Mind & Brain for Beginners, Icon Books, Cambridge, 1998.

Georgieva, S. A.: Essentials of Physiology, Mir Publ., Moscow, 1989.

Goleman, Daniel: Emotional Intelligence, Bantam Books, NY 1996

Jouvet, Michel: Le someil et le rêve, Editions Odile Jacob, 1992.

Kandel, Eric R. and Hawkins, Robert D.: The Biological Basis of Learning and Individuality, SCIENTIFIC AMERICAN, Sept.1992, pp. 53 - 60.

Pinker, Steven: The Language Instinct, Penguin, 1995.

Restak, Richard M.: The Brain. The Last Frontier, Warner Books, New York, 1979.

Thompson, Richard F.: The Brain, a Neuroscience Primer, W.H. Freeman and Company, New York, 2.ed. 1993.

Chapter 4

THE DESCRIPTION OF MENTAL PROCESSES ACCORDING TO COGNITIVE PSYCHOLOGY

Each discipline describes mental processes according to its own point of view. Thus, while neurology is concerned with the physiological changes taking place in neurons during mental processes, cognitive psychology tries to describe such processes from a systemic point of view, without reference to the actual physicochemical and physiological processes. The brain is considered as a mechanism, whose internal way of functioning we are trying to guess on the basis of observations of its output, the various kinds of mental phenomena (linguistic capabilities, pattern recognition, logical reasoning, intuitive thinking, etc.).

The results of such efforts are systemic models describing how mental processes could be reproduced as algorithms on a computer. Inherent in this point of view is the belief that the actual physical mechanisms are not the only ones that can produce mental processing, but a detailed enough model of elementary mental processes and the interaction between them might be equally effective, no matter how we realize these processes technically.

Here we are going to describe the most comprehensive efforts and to what extent they have succeeded. As we shall see, a complete success has been proven extremely difficult or impossible to achieve, although individual mental processes can be partially simulated by such models. For this reason, there are various kinds of proposed models and much dissent between researchers, as to what kind of model would be the most successful one.

In order to see how far such an approach can go, let us first consider how incoming messages can be represented and stored in the brain.

A fact well established experimentally, is that there are at least two distinct memory mechanisms: Short Term Memory (STM) and Long Term Memory (LTM). Actually, there is also an "immediate memory" which holds information, more or less unprocessed, only for 10 to 20 seconds but this information is then lost, unless it is processed, evaluated and reinforced. It then remains in the STM possibly for several hours or even days, if it is considered temporarily important (e.g. an appointment). It may also be stored in the LTM, if it is considered of long lasting importance.

LTM stores information permanently and allows its subsequent retrieval whenever it is necessary for the processing of other incoming information, while information held in the STM is lost when it is no longer chronologically relevant. It is thus generally assumed that LTM is a result of structural changes in the brain, while STM is merely the activation of various locations of the LTM system by incoming signals (visual, echoic, etc.). Therefore, the following discussion basically pertains to the possible organization schemata of Long Term Memory.

4.1 How can messages be represented in the mind?

Most cognitive psychologists and people working in Artificial Intelligence, i.e. people trying to replicate cognitive processes on a computer for the solution of practical problems, believe the answer lies in "Procedural Semantics".

Spoken or written messages are considered as sets of instructions to the receiver to search for or rather, build up certain representations in his mind and perform operations on them.

However, the problem arises when we try to determine how the most elementary such processes, the most elementary instructions, are to be represented. Many researchers seem to understand the basis of these procedures, the elementary instructions, as having a linguistic form or being translatable into such a form. This leads them to insurmountable difficulties and never-ending disputes on how information can be broken into elementary parts and of what nature these parts will be (ɑ discussion of some philosophical and psychological positions on this subject is given in **Appendix 4.1: Is Meaning Logically Decomposable?).**

On the other hand, neurologists also favor procedural semantics, but to them these procedures are of neural nature, i.e. purely analogical and not necessarily translatable into linguistic form. Thus, A.R. and H.Damasio [1992] say: "How are concepts physically represented in the brain? We believe [that] there are not permanently held "pictorial" representations of objects or persons, as was traditionally thought.

Instead, the brain holds in effect a record of the neural activity that takes place in the sensory and motor cortices, during interaction with a given object. The records are patterns of synaptic connections that can re-create the separate sets of activity that define an object or event; each record can also stimulate related ones."

We may thus accept procedural semantics as a basis for representing and storing information, but have yet to answer the basic question, whether "semantic primitives" can exist or not. Extensive and long-lasting efforts of many researchers to formalize meaning, have led to so poor results that many theorists doubt today that a theory of semantics will ever be possible.

For instance, Jerry A. Fodor, once a leading proponent of the computational theory of mind and of the notion that the mind is divided into modules dedicated to specific tasks, acknowledges now that this view has limited applicability. Fodor, who had previously proposed two different theories of cognition, the theory of "Semantic Markers" [Katz & Fodor, 1963] and that of "Meaning Postulates" [Fodor, Fodor, and Garrett, 1975], notes in a review of *How the Mind Works,* presented in 1998, that certain cognitive tasks, such as the ability to detect color or to parse a sentence, can indeed be reduced to computation, but dividing the mind into many little dedicated computers or modules, still leaves unanswered the question of how the results of all these modular computations become integrated. He adds: "As things now stand, we don't have a theory of the psychology of common sense that would survive scrutiny by an intelligent five-year-old" [Horgan, p.197].

Similarly, many linguists doubt that it is in principle possible to analyze meaning into semantic primitives and philosophers disagree on the question, whether a formal account of everyday reasoning about somebody's actions and intentions is possible even in principle [M. Boden, 1988, p.123]. For instance, Hubert Dreyfus [1992] cites arguments by Wittgenstein and Heidegger, that it would be extremely difficult to reproduce human perception and cognition with a formal, rule based model.

According to Wittgenstein, any factual statement about the world has to be explained by other factual statements; there are no fundamental facts or "primitives", as Wittgenstein referred to them, which serve as the foundation for cognition, as quarks and electrons do for the physical realm. Similarly, Heidegger points out that rules rarely apply to all situations; each rule requires extra rules to determine whether the initial rule is relevant for a particular situation, thus creating an infinite regress of rules.

But even if analysis of meaning in basic components is possible, it is very difficult to decide which concepts are basic, either semantically or psychologically, so that they might be used as "atoms of meaning". It has been proposed for instance, by C. Shank, that "ingest" is a semantic primitive, contributing to both "eat" and "drink". However, to many psychologists it seems implausible that an abstract concept like ingestion is psychologically more basic than either eating or drinking.

In spite of these doubts, that analysis into components is in principle possible, there are a number of computer models based on assumed semantic primitives, in terms of which sentence representations are defined. The obvious reason for this persistency of theorists' efforts to define logical or linguistic semantic primitives, is that only by means of such primitives could one reduce the whole mental system to a system of logical algorithms, i.e. design a "thinking machine".

An extensive discussion of many of these attempts can be found in [M. Boden, 1988]. Here we will discuss only some of them, which have been more popular or seem more plausible.

4.2 Semantic Networks and Semantic Primitives

Theories of "Semantic Networks" build up more general concepts by appropriately connecting more specific ones, and so achieve the inheritance of properties from one level to the next.

One of the first semantic networks was that of Quillian [1969]. A more elaborate model, that of Collins and Quillian [1972], assumed that the meaning of a word is its set of associations with other words and considered various sorts of associative links, such as "class inclusion", "part – whole", "property of", and variable relations as specified by a third defining word.

The organization of this network was assumed to reflect the way in which the information was originally acquired. For example, it was seen as more likely that people do not learn directly that poodles are animals, but infer it from the knowledge of the facts that dogs are animals and that poodles are dogs. Hence the representation of "poodle" was not linked directly to the representation of "animal" but indirectly by the chain of links:

poodle -(is a)→ dog -(is a) → animal

Entering the semantic network at a particular point, we can retrieve certain facts directly (e.g. "dogs are animals"), while the retrieval of other facts requires additional links to be followed up (e.g., "poodles are animals").

Several experimental results suggested that semantic memory does consist of a hierarchical network that can be entered with equal ease at any point ([Freedman and Loftus, 1971]; [Loftus and Freedman, 1972]) and supported the notion of a hierarchical retrieval of semantic information ([Collins and Quillian 1969, 1972]). However, it was also found that although the closeness of items in the semantic network facilitates positive judgments about their relation, it impedes negative judgments, a fact that is not explained by this theory. It takes longer to evaluate the truth of the statement "a canary is an ostrich" than that of "a canary is a fish". It seems that the mind has an inherent difficulty in drawing boundary lines between close concepts. The greater the similarity in meaning between words, the easier it is to make a positive judgment on their semantic relation and the harder it is to make a negative one.

A further difficulty presented by the semantic network theory, in its original form, was also that it failed to predict differences in "typicality" within categories. It could not explain why some members of a category are considered more typical representatives of that category than others. E.g. why do people recognize much faster as true that "a canary is a bird" than that "an ostrich is a bird"?

The varying typicality of members of the same category, i.e. the fact that people do not consider all representatives of a category as equally typical members of it, was established by E.Rosh in a series of experimental studies on categorization, first published in 1973 (see, e.g. [Rosh, Mervis, Gray, Johnson, and Boyes-Braem, 1976]).

Note that actually, necessary and sufficient defining properties can be found for very few concepts. Most "natural" categories have no logical properties shared by absolutely all their members (see, e.g. [Smith & Medin, 1981]). The ability to move, for instance, seems to be a characteristic property of animals in contrast to plants or inanimate matter. But then, what are corals or oysters?

The original semantic network theory has no explanation for varying typicality and therefore cannot explain satisfactorily how objects are classified in categories. An attempt to overcome this difficulty was made by Cohen and Murphy [1984]. They amended the network model, in order to account for typicality, by considering the determining, the "defining" properties of a concept, not as necessarily inherited by the subordinate level concepts but only as default features assumed if no contravening information is available. Thus, a bird is by default a flying animal but there may be birds, ostriches for instance, which can not fly. This of course, raises the questions: why should we classify ostriches as birds? Is "bird" a natural category, so to speak, instinctively known, or an artificial concept?

Obviously, "bird" is not such a "natural" category, as we might initially think. Our senses do not tell us immediately whether something is a bird or not. We may recognize immediately that something is a "flying object" but not a "bird". The existence of "natural categories", i.e. categories whose membership is determined by "natural" properties, is possibly an illusion.

Possibly, we could distinguish "primordial categories", i.e. categories directly referring to our senses. But even the "primordial" categories are not immediately given. The ability to perceive and recognize features is a matter of training of the senses developing gradually not only in a child but also in adults.

For instance, the trained eye of an architect recognizes, in a building, features that everybody else perceives only when they are pointed out to him. Similarly, the trained eye of a mechanic recognizes immediately that some part of the engine has some defect, while we perceive this defect only when it is pointed out to us. Of course, a child learns gradually to recognize specific features of an object.

Semantic networks have been very popular as tools for building theories of meaning, since they have the very appealing feature of displaying, in a very simple and direct form, the interconnectedness of concepts within a system of relations. Large numbers of interrelated facts can thus be represented so that they can be easily retrieved by a computer program.

Later semantic network models consider the long term memory as some kind of network with concept-property or action-object pairs, e.g. "John-tall", "see-George" or "cross-street", called "propositions", rather than mere concepts as its knots. Connections between such knots corresponded to relatedness between concepts (or propositions).

The term "proposition" was introduced instead of "concept", in order to take into account the fact that in experiments about language, understanding the simplest contents of memory does not seem to correspond to words, but rather facts representable more appropriately by phrases, i.e. combinations of concepts (see [R.Klatzky 1980, p.196-199]). A proposition was considered to be "something like a sentence, but more abstract, that is, more like the meaning of the sentence than the sentence itself" [Klatzky, 1980, p.197]. However, although it was emphasized that propositions are not actually strings of words, they were conceived as representable by phrases, i.e. somehow corresponding to phrases.

This emphasizes a further difficulty of this kind of model of memory. It cannot resolve the vicious circles appearing in many related groups of definitions, i.e. the fact that often in defining concept B, we use concept A and in defining concept A, we use concept B (see Johnson Laird [1983, pp.230,241]). For instance, when defining a bird, we include the fact that it has wings and feathers but when defining a wing or a feather, we also have to refer to the fact that they a part of a bird.

This raises the question: which concept can we consider as the starting point for the creation of such a cycle? E.g. which is created first; the concept "bird", or the concept "feather"? Feathers are characteristic features of birds and must be mentioned in the definition of "bird". On the other hand, we can hardly define "feather" without mentioning that it is part of a bird.

This dilemma is similar to that posed by the question: "What came first; the hen or the egg?"

A plausible way of resolving the dilemma in both cases is to adopt an evolutionary point of view, as we are going to do here. But let us first look at some other attempts.

4.3 The Problem of Determining Fundamental Meanings

One way to resolve this difficulty is to try and reduce all definitions to a minimal set of supposedly "elementary", i.e. somehow inborn concepts or propositions. This was indeed attempted by R. C. Schank and his co-workers.

Schank's theory of semantic primitives [Schank 1983, 1986]) begins with a set of "primitive actions", which are assumed to represent meanings that underlie all natural language verbs. The set consists of eleven primitive actions and includes items such as PTRANS (physical movement from one location to another), MTRANS (any sort of movement or transformation of a mental representation), ATRANS (transmission of ownership) and INGEST (the concept supposedly basic to eating and drinking). In addition, there are unanalysable semantic relations such as POSS-BY (possession) and CAUSE, as well as quasi-syntactic relations such as "object' and (various) "cases". These primitives are used to build semantic networks called "Conceptual Dependencies" (CDs), which support inferences to concepts and beliefs that are not explicitly mentioned in the text considered.

One might still ask how "primitive actions" are encoded in the mind, but beyond that, there are other problems in this description of meaning. For instance, Shank's translation of the CD-diagram representing the sentence, "Jane punched Mary" is: "At some time past, Jane applied a force to the object "Jane's fist", in the direction from Jane to Mary. She did this by moving her fist in the direction from Jane to Mary. Her action of applying force to the fist caused Jane's fist to be in contact with Mary".

However, this formulation does not make clear the fact that a punch is an intended action and not a coincidental touch. The theory does not provide a precise representation of the exact meaning of the above sentence.

Another way to avoid circular definitions used in later semantic network models, is to include an additional level of representation: the level of sensory encodings that represent concepts, or rather "propositions", in some form related to the sensory inputs, having similar structure with them (See e.g. [Mc Clelland & Rumelhart 1981,1982]). This brings such models closer to Johnson - Laird's theory.

This, in principle correct, attempt to escape the vicious cycle of circular definitions, reveals a weak point of all models of the Long Term Memory system. Their approach is top-down, starting at the level of concepts and trying to reach the level of "meaning-elements", possibly sensory inputs. But it is never very clear what the intermediate levels of such a structure are.

Neurologists on the other hand, adopt a bottom-up approach, starting with sensory inputs, but also encounter the same difficulty: what are the intermediate levels of the organization of the conceptual system?

The basic problem is in all cases the reduction of "macroscopic" objects, like concepts, to "microscopic" ones, i.e. elementary sensory inputs. A resolution of this difficulty, which we will develop later, is that concepts are products of evolution and that the intermediate levels of organization of the conceptual system are its intermediate developmental stages.

Another weak point of many semantic network models is that their nodes are usually thought to represent a network of relations of logical categories. For instance the relation, "a dog is an animal", refers to logical categories. The concept "animal" has to be defined logically in our mind before we can relate it to the concept "dog". However, it is not conceivable how nodes of the LTM network, based on logically well defined categories and relations (e.g. non-overlapping etc.), come into existence. There is no way by which such categories, or basic logic - categorical relations, could be built up in a growing child, at least in the early years of mental development.

According to Developmental Psychology, a child below the seventh year of age does not have a clear conception of numbers like "four". How is he then to understand that a defining feature of a dog is that it has four legs?

No innate logical processing mechanism has been postulated, at least in any of the main models of memory organization. How could then categorical classification come about in early developmental stages?

From a logical point of view, such a classification is also not always possible, without predetermined arbitrary criteria. This is well demonstrated by the often raised question, "Is a tomato a fruit or a vegetable?" The criterion that fruits must be sweet is not valid, for instance for lemons, while there are certain very sweet varieties of beetroots, which are commonly considered to be vegetables.

As a matter of fact, experimental results do not favor classification in logical categories, but classification according to "typicalities", i.e. according to the degree of similarity of an object to a prototype representing a whole category of objects. But even a classification schema according to typicalities does not explain how prototypes are formed, i.e. how they are acquired.

In agreement to the evolutionary view of the conceptual system, which we are going to develop, we will opt here for the encoding of "empirical" categories (directly resulting from experience), instead of logical categories. This will also provide an explanation for the varying typicality of logical concepts, since each one may be based on different empirical conception schemata, even though we place them in the same logical category.

Referring to all these models of LTM, we may also note (without usually being very explicit about it) that they seem to favor sensory, rather than sensori-motor, "encoding" at the lowest memory level. However, the human conceptual system (as well as that of animals) is highly interactive. The brain does not simply receive information, it also acts upon the surrounding through the body or rather, it constantly interacts with the surrounding. Therefore, interaction schemata should be parts of the "encoding" mechanism.

Note, for instance, that an acquired aptitude is extremely difficult to describe in linguistic terms. This is obviously due to the fact that its acquirement is not based on precise sensory descriptions, telling us how we should move our body, but rather on the direct coordination of sensory and motor mechanisms.

Another indication that the conceptual system is in its whole nature interactive, is given by the fact that language cannot be learned merely on the basis of sensory inputs, but only interactively.

As B. A. Moscowitz [1978, p.89] reports, efforts to teach a non-impaired child of deaf-mute parents to speak, by exposing him to television programs, failed, although the child learned to communicate fluently in sign language.

There are also some indications that even supposedly abstract mathematical concepts are based on sensori-motor organization schemata. For instance, the first, primitive notion of "continuity", which is actually used most of the time for doing mathematics, is based on continuous movement. A continuous line is, at the initial, primitive level of mathematics, any line which we can draw without lifting the point of the pen from the paper. However, such a line is by necessity not only continuous, but also piecewise continuously differentiable. I.e., it cannot have sudden changes of direction at <u>every</u> point, but only at certain discrete points.

Therefore, this property is taken, by most people, to be an intrinsic property of any continuous line. They simply cannot imagine the existence of a continuous line without this property. They are, therefore, greatly surprised and even reluctant to believe that there are continuous functions, which are nowhere differentiable. The graphic representation of such a function is difficult to imagine, because it would be a line which has no tangent anywhere, nor direction (the graphic representation of such a function has infinitely many "spikes" at each interval).

Even great mathematicians of the 19th century have been thus misled by the conceptualization mechanism of continuity, which has obviously nothing to do with the abstract definition. Hermite, for instance, wrote in a letter to Stieltjes: "I recoil with dismay and horror at this lamentable plague of functions which do not have derivatives" (see [Doob, 1996]).

Yet even today, mathematicians use in their thinking – most of the time – the primitive notion of continuity described above, rather than the exact definition, because the notion is more "insightful", i.e. easier to handle mentally. The definition is only used when formulating final results, or in cases where the primitive kinesthetic notion would lead to the wrong conclusions.

4.4 The Theory of Mental Models

A different approach is attempted by P.N. Johnson-Laird. He has developed (partially in cooperation with G.A. Miller) a theory of psychological semantics that reduces meaning to semantic primitives, not of lexical nature, but of procedural nature, i.e. relying on biological procedures. This theory does not only provide a way of representing meaning, but also a mechanism for mental processing, a feature that makes it psychologically more plausible. Indeed, in certain aspects it is close to the view of mental processing which we are going to develop here.

The previous theories are based on the assumption that a formal theory of meaning is possible and defines truth on the basis of the validity of logical relations between linguistic expressions. They try to achieve logical coherence of the set of all linguistic expressions, without any reference to the outward reality.

This always leads to a circularity of how meaning is provided for any word. It must refer the meaning of one word to that of others and is thus trapped in a vicious cycle. In contrast, Johnson-Laird and G.A.Miller define truth as a correspondence between language and the real world.

According to this theory, not all words require the same kind of semantic analysis. Some words are analytic, defined on the basis of necessary and sufficient conditions, but others, like the names of natural objects, refer directly to outward experiences and can only have "meaning" based on sensori-motor processes of the body.

Johnson-Laird demonstrated how meaning can be inferred on the basis of a model theory by giving a simple example [Johnson - Laird, 1988, p.339-342]. He developed a procedure for interpreting linguistic expressions, referring to the spatial relations of left and right.

This interpretation is simply a "mental model" that exhibits the spatial relations referred to by the linguistic expressions. It is a "translation" of the linguistic expression into a "mental model" of the situation it describes. I.e., it is an arrangement of "tokens", representing specific objects in an imaginary space, a "mental" space.

Thus, a sentence like, "The dishwasher is on the right of the cupboard", leads to the construction of a mental model of the form:

[cupboard] [dishwasher],

whereas [cupboard] and [dishwasher] are initially only "tokens", mental symbols of some kind, representing the corresponding objects. These tokens are replaced by more detailed descriptions of these objects only if this becomes necessary.

The process of interpreting linguistic expressions by creating mental models is simulated on a computer by defining two "mental" spatial coordinate axes, a horizontal axis and a depth axis. The predicate "on the right" corresponds then to the procedure:

ONRIGHT: Hold the value of the depth axis constant and increment the value of the horizontal axis.

Given the above sentence, this procedure immediately produces the above arrangement of tokens. An additional sentence, "The oven is on the left of the dishwasher", could be represented either by a mental model of the form:

[cupboard] [oven] [dishwasher]

or by one of the form:

[oven] [cupboard] [dishwasher].

Faced with such an equivocal situation, people probably resolve the dilemma by adopting any one of these models, e.g. the first one, as a default model which is valid until contravening information is received. Thus, a further sentence, "The oven is on the left of the cupboard", leads to a revision of this decision and a replacement of the first model by the second one.

4.5 Mental Models and Logic

Such a mental model has the important advantage of allowing valid inferences (in this case, of spatial arrangements of objects) without any use of formal rules of Logic. After the model has been built up, all spatial interrelations can be simply deduced by observing the model. Thus, the sentence, "The dishwasher is on the right of the oven", is verified as "true", since it agrees with the final adopted model.

In fact, this was the initial motivation that led Johnson-Laird and certain others to propose a mechanism of inference based on mental models. It was observed that people perform much better in solving logical problems, when the problems refer to concrete situations, than when they are posed in an abstract form. This led to the conclusion that, in order to make inferences, people do not apply formal rules of a hypothetical "mental logic" but always try to build up a mental model of the given premises and "look" for valid conclusions by observing it.

This idea goes back, at least to K.J.W.Craik [1943, p.51], who suggested that a cerebral model of an outward situation allows verifications of truth or falsity because it "has a similar [physical] relation structure to that of the process it imitates".

4.6 The basic features of Mental Models

The generally true observation, that we use mental models in order to make logical inferences, was extended by Johnson-Laird in order to build up a whole memory system. He proposes that mental models represent objects, situations, sequences of events and in general the surrounding world, as well as social and psychological phenomena. Therefore, he formalizes them by imposing three basic requirements, which seem necessary in order to use them as a basis for a precise theory of meaning:

1. Computability: Mental models and the machinery for constructing and interpreting them should be computable (they should be computationally reproducible).
2. Finiteness: Mental models should be finite in size and cannot directly represent some infinite domain.
3. Constructivism: Mental models should be constructible from "tokens" arranged in a particular structure to represent a state of affairs.

This claim raises two basic questions:

a. How do mental models represent the external world? That is:
   How do tokens function as representatives for entities?
   How do properties of tokens symbolize properties of entities?
   How do relations between tokens represent relations between entities?

b. What kinds of concepts are embodied in mental models? I.e., what kind of concepts rely more on mental models than on definitions?

Let us see how Johnson-Laird answers them: his answer to the first question is that the essential characteristic of a model should be its functional role. A model should be understood as a structural representation of an outward situation in an arbitrary symbolic notation. It is important that the interpretative system should treat an element, A*, in this notation as corresponding to an entity, A, in the outward world, based on a structural resemblance between A* and A, and not on any supposed physical correspondence [Johnson-Laird, 1983, p.403]. They must have similar structure, without necessarily the one being an exact copy of the other.

The second question is answered in connection to a general discussion of the meaning of concepts. Johnson-Laird notes that all concepts are not definable in the same way. Basically, there are simple concepts, like "red" and complex ones, like "vertical". He calls the complex concepts "analytic", because they are the logically definable ones. They relate other concepts to each other in a precise way, described by necessary and sufficient conditions. Such is, for instance, the concept "uncle", which means "a brother of a parent".

Other concepts, however, such as "apple" and "tiger", are terms of a natural sort. They correspond to an outward, objective reality and can be mediated only by experience. The meaning of words corresponding to such concepts cannot be communicated by providing a set of necessary semantic components. It has to be passed on by describing typical instances of the concept and by acquiring immediate experience in connection with it.

Of course, the meaning of the word "lemon" might be decomposed into characteristics like: round, yellow, having peel, having a sour taste and so on. But none of above characteristics is a necessary feature of it. They may all vary more or less. A lemon may be green for instance, if it is not ripe enough. Thus, when we explain the meaning of the word "lemon", we usually describe what a typical lemon is, providing a stereotype for lemons, a schematic model based on default values that will help the listener to recognize a lemon when he sees one. We may note, however, that even so, no description is adequate to make one understand what a lemon is. Unless one sees, feels, smells and tastes a number of lemons he will not really know what a lemon is.

A tomato and a lotus fruit have almost the same outward appearance. However, I don't think that an adequate description of a lotus is that it is "something like a tomato, but with sweet taste".

Secondly, Johnson-Laird also distinguishes some concepts such as "home", "chair" and "melody", which are basically defined by what we intend them to be, the function we wish them to fulfill. He says that these concepts have "constructive semantics". This means that the intended meaning of these words is a mental construction imposed on the world. Their meaning is a matter of convention. Thus, whether we wish to call some piece of furniture "chair", "armchair" or "seat", is a matter of choice and convention while, according to Johnson-Laird, "uncle" is always the brother of a parent. Typical concepts with constructive semantics are predicates like "tall" or "clever", since their meaning is obviously a matter of perspective.

Although the distinction between physical, analytic and constructive terms is certainly useful, we should note here that it is possible to give an intentional meaning to an initially analytic term and an analytic meaning to a term with initially constructive semantics. We may choose to call "uncle" a distant relative or an old neighbor and we may give an analytic definition for the word "chair", if we decide to call "chairs" only such artificial seats that have four legs and a back.

4.7 Procedural Semantics and Language

The procedural primitives considered by this model of LTM have no ready expression in object language. "They are outside of conscious awareness, they cannot be easily described and they cannot be acquired from experience, because the mental representation of experience already demands the ability to construct models of reality on the basis of perception". This means that they are innate and underlie our ability to represent the world and to act based on such representations [Johnson-Laird, 1983, p.413].

This procedural point of view is also extended to abstract concepts. Johnson-Laird assumes that psychological procedures which define abstract meanings are ultimately also based on kinesthetic routines. Thus, for instance, he sees the idea of including certain concepts in a unique class, as being derived from the psychologically more primitive concept of including some things in a region of space. However, the primitive perceptual routines that co-determine meaning are considered as primitive only with respect to semantic procedures, i.e. procedures of semantic analysis. They are not necessarily either computationally unanalysable or inborn.

On the other hand, the acquisition of some such routines may require movement in and action upon the material world, i.e. training. George Miller and Johnson-Laird cite evidence showing that infants learn the word "in" before they learn "on" and both before "under" and "at". They point out that the child's understanding of such terms appears to depend on the availability of "bodily action schemata". They conclude that such a concept is understood as soon as the body has developed the functions necessary for acquiring the concept by interacting with the surrounding.

A confirmation of the necessity of distinguishing between procedurally acquired meanings and meanings acquired by definition is, according to Johnson-Laird, the fact that children up to the age of six years acquire about nine new words per day (see [Templin 1957], [Miller 1977]). This is so great a speed of learning that it seems unlikely they are learning the complete meanings of the words or fixing their extensions by encountering many samples of the actual objects to which they refer. They must be able to grasp the meaning of a word to a great extent solely from the linguistic and interactive context in which it appears.

This conjecture is also supported by a study of Til Wykes and Johnson-Laird, carried out on 3 to 4-year old children. It showed that they can pick up and retain, for at least a week, information about the meaning of an artificially constructed verb, merely from hearing a story like the following:

"John stepped out of the boat and the water "mibbed" his trousers; so he went to change into some dry clothes. The water had "mibbed" his trousers right through; so Simon made him some hot tea. But John dropped his cup and the tea "mibbed" over the floor".

The artificial verb "mib", used here, is not merely another name for an already known verb, since it has both a transitive meaning similar to that of "soak" and an intransitive meaning similar to that of "spill".

The position that certain meanings are acquired procedurally is also supported by experiments which reveal that speakers have systematic gaps in their knowledge of rarely used words and that the pattern of ignorance is predictable.

What we can do with something, as well as its perceptible characteristics, are things more widely known than its other features (see [Keil, 1979]). It is also generally known whether something is consumable or not and whether it is solid or liquid, while it may be unclear whether it is natural or man-made.

4.8 Critical Remarks

Obviously, the hope that we can build up a functional model of the whole mental system, by distinguishing the concepts into "naturally acquired" and "artificial" ones, is inherent in this whole theory. Johnson-Laird hopes that thinking can be reduced to some kind of syntactic processing, which must only take into account that certain concepts are "defined" not logically, but by means of kinesthetic procedures. This hope is misguided. It is the procedural content of the concepts which provides creative insights and not the syntactical structure of statements.

We should note here that the distinction between semantically primitive words (represented by biological procedures) and semantically composite words is problematic and somewhat artificial, because there is no way of discovering such distinct primitive words. All concepts rely much more on a highly complex network of links to other concepts and procedures (or rather, microprocedures) than we wish to admit.

For example, what is the difference between scrutinizing and watching? According to The Oxford Dictionary "to scrutinize" is "to look at or examine carefully", while "to watch" is "to look at, to keep under observation". The difference does not seem very clear.

But even if we are able to fix a word's extension through a dictionary, we can never exactly grasp the speaker's intention, his state of mind when he uses it. This way of modeling concepts has therefore very limited capability of reproducing them, just as approximation of the number $\pi = 3.1415927\ldots$ (the quotient of the circumference to the diameter of a circle) by 3, is not able to reproduce many of the properties of this quantity.

In trying to assess this theory, we should also bear in mind why other theorists tried so persistently to define logical or linguistic semantic primitives. As said before, the obvious reason is that only by means of such primitives could one reduce the whole mental system to a system of algorithms, i.e. design a thinking machine. This obviously is also Johnson-Laird's hope. He hopes that by introducing procedural primitives he can build up a fully functional structural model of the mind. Therefore, we may ask to what extent Johnson-Laird's theory achieves this goal.

The answer is that, although it uses very skillfully many psychological observations, it is obliged to rely on an oversimplification which reduces its effectiveness considerably.

Since semantic primitives are defined on the basis of very extensive procedural semantics, they are not directly constructible. Procedural semantics are not easily analyzable and reproducible (or not at all) in linguistic form. They have to be provided in some other form, for instance as functional values of (yet unknown) neural networks. This limits considerably the effectiveness of this system, especially if, as we have noted here, all concepts have a procedural component.

Certainly, one may try to capture the meaning of natural terms by computational procedures like ONRIGHT, which was used above, and on this fundament build higher levels of synthesis, by means of syntactic rules of interpretation. But this whole construction is too simplistic, just like the one based on "semantic primitives" like INGEST.

There is certainly progress in analyzing and mapping the mental system by means of algorithmic processes, but one should also not overestimate the possibilities of this model, since not only certain concepts, but <u>all</u> concepts have extensive procedural semantic components. These are related to the concept's symbolic content, which is used when we speak in parables and metaphors, when we use the concept figuratively. For instance, an unconscious component of "uncle" is that he is an older man. We would never call a child "Uncle Tom".

Similarly, the word "right" does not merely indicate a relative position, but has also the totally different meaning "correct". How did this second meaning come about? Obviously, since 95% of the people are right-handed, they consider the use of the right hand for performing a task as "correct" and the use of the left hand as "incorrect". Ready-made products are usually constructed only for right-handed people. Thus "right" is associated with "correct". Germans call someone who is clumsy "linkisch", i.e. literally "leftish", since "links" means "left" in German. Such alternative meanings reveal associations which may sometimes be unconscious.

What is more, procedural components do not simply preexist as a priori given entities, but they are created by accumulating experiences. For this reason, they are also deeply interwoven, so that it is extremely difficult or impossible, either to reproduce each meaning separately or to create a whole network of meanings.

Thus, a distinction between primitive and derived concepts is to a great extent artificial. In speaking about primitive concepts we miss the important fact that there are various levels of bodily encoding of a concept, which are built up during the development of the concept, i.e. during its evolution.

We are going to elaborate on this view in the next chapter. Here we will first consider the computational model to which the above theory leads.

4.9 The resulting Computer Model of Semantic Analysis

Johnson-Laird has also implemented his model theory for the creation of a computer model of semantic interpretation and mental processing of language.

The semantic interpretation of sentences is here based on an algorithm that interprets the syntactic structure of a sentence, by proceeding from left to right. It looks at one word at a time and makes a search of the breadth of meaning by building a table of all possible interpretations. Then, it selects one on the basis of general information about the subject discussed.

This program uses syntactic judgments as crucial clues to the final representation of meaning, but it does not build any separate representation of grammatical-syntactic structure as such, as is done in Linguistics. It does not separate the syntactic structure from the semantic structure.

Quite correctly, Johnson-Laird considers syntactic structure as psychologically important, not for its own sake, but only as a way of constraining meaning. He therefore argues that the language user has no reason to build a detailed syntactic parse tree for a sentence, before computing its semantic content.

Thus, the algorithm tests the input words only for the presence of certain syntactic features and uses the existing syntactic structure in order to construct a semantic model or representation of the sentence's meaning.

What sort of model is constructed in this way depends on the meaning of the sentence. An interrogative sentence will lead to a search for the information asked for, whereas a declarative sentence will lead only indirectly to a search, if this is necessary in order to construct the semantic representation.

The way in which the process of comprehension on the basis of mental models proceeds, is similar to the functioning of the hierarchy of bodily motor control. Higher order instructions are given by this mechanism in general terms, without reference to details. These are automatically supplied by the subsequent, more specific levels of control, so that voluntary control of fine movements is only occasionally necessary.

In a similar manner, comprehension usually seems to require conscious attention mostly for high level notions, while the mobilization of the basic procedures for the construction of a mental model is supposed to happen as automatically as the tuning of muscles ([Johnson Lard, 1983, p.406]).

The mental model that derives from visual perception is a single entity, corresponding to a single state of affairs but if it is based on discourse, it may be non-deterministic, compatible with many different states of affairs, since discourse takes some time to carry over a specific meaning. In this case, the simulation procedure is supposed to construct an initial model based on plausible assumptions and to revise it recursively, if any assumption turns out to be wrong.

4.10 Problem Solving by means of Mental Models

What makes Johnson-Laird's theory particularly attractive is that he has extended this model of the conceptual system, in order to explain mental processes of problem solving which are usually attributed to logical processing.

The ability to use logical deductive reasoning is very often considered to be an inborn property of the human mind. For instance, Piaget believed that after acquiring the ability for formal operations in adolescence, adults usually employ abstract deductive reasoning in order to reach logical conclusions in syllogisms.

However, there is extensive evidence suggesting that we do not normally rely on logical deduction, even when we are solving logical problems of deductive reasoning.

A typical finding is that the success rate of problem solving, for problems of identical logical form, can vary between 19% and 98% depending on whether the problem is given in an abstract or a concrete form, referring to everyday situations [Wason, 1977]. Such results are invariably obtained in psychological experiments with groups of people who are even well educated, since they are usually students.

But if problem solving does not involve the use of deductive logic, how is it done? An alternative hypothesis is that people reason by using "mental models" of the situation given in the problem.

As said before, K.J.W.Craik had already suggested in 1943 that a cerebral model of an outward situation allows verifications of truth or falsity because it "has a similar [physical] relation structure to that of the process it imitates".

This idea has since often been generalized and recommended by various computational theorists. However, the most systematic development of it is due to Johnson-Laird [1983].

Johnson-Laird explains "problem solving" not as reasoning by applying logical rules, but by constructing models. He says that when people solve problems, even problems of logic, usually they do not use logic but rather construct mental representations (models) of the premises and transform them, in order to reach conclusions by direct observation of the various aspects of the model.

In support of this view, he observes that logical problem solving on the basis of mental models is computationally far more efficient than the application of formal rules of inference, although the former can sometimes lead to error if the model does not represent correctly all aspects of the premises.

He also notes that while logic is an instrument of reasoning, due to higher cultural evolution, the ability to construct and reason with mental models is a general feature of the human mind, used equally well by primitive people.

Actually, this method of problem solving seems to be also used by mice, as Tolman's maze experiment shows (see the previous chapter). The mice seem to create, on the basis of experience, a mental model of the maze in which we put them and select the appropriate course to the location of food, based on this model.

In this theory, syllogistic problem solving has three stages:

(1) a mental model of the first premise is constructed;

(2) the information in the second premise is added to the mental model of the first premise, taking into account the different ways in which this can be done;

(3) a conclusion is formulated expressing the relation (if any) between the end-terms and which holds in all models of the premises.

Obviously, if the second stage described above has not been thoroughly done, then the third stage is likely to produce invalid conclusions.

The procedures for making inferences on the basis of a mental model are supposed to work in a similar way as the procedures that create models during discourse. In order to verify the universality of a conclusion that is true in a current model, attempts are made to modify the model recursively in any way that maintains the premises but violates the conclusion. If this attempt fails, then the inference is considered to be valid.

However, this view of syllogisms raises an important question: how are predicates, which represent an indefinite number of entities, like “all” in “all cowboys are good riders”, modeled?

Johnson-Laird suggests that this is done by imagining an arbitrary but finite number of cowboys and labeling each of them as a good rider, already knowing that the number of entities depicted is irrelevant to any syllogistic inference that is drawn.

By this he does not mean that the problem-solver must form any mental pictures of cowboys, but rather that some inner symbol is set up to represent each cowboy and each of these symbols is somehow linked to some other symbol representing a good rider. The symbols and links of this kind are supposed to be manipulable by operations of a basically perceptual nature.

This conception of mental models also relies on innate semantic primitives, innate conceptual primitives and operators, which are used for constructing the models.

4.11 Critical Remarks on the possibilities of problem solving and creative thinking provided by mental models

Johnson-Laird's theory is indeed successful in explaining how people solve mainly logical and practical problems, but it cannot explain all creative mental processes.

Considering problems of practical reasoning, we may add here a refinement. The model on which the solution process of such problems is based, is not always a static one, but may instead be dynamic, one that changes continuously. It is a "shifting model".

In solving a logical problem, we add to the premises factual information from our memory. But in many instances, we must keep on bringing new information to the active memory, while previously recalled information is already used and can soon be discarded. For instance, when we drive from one place to another, we first choose a general route which is kept in mind only in its general features, its most salient points. Then, each time we recall the relevant cues for our next movement, as we drive past the previous cue we had in mind. We remember that a certain building or park or crossing will soon appear, just after we see some other salient feature of the road that just precedes it. We do not recall the whole route all at the same time, but successively, one section after the other. That is why we usually find it difficult to describe the whole route, i.e. all these cues, to a third person.

However, there is a more crucial difficulty which simple mental models cannot solve, whether they are static or dynamic.

In order to find our way from one place to another, we only have to recollect all knowledge we have about possible routes. This is not so difficult, because the relevant information is usually finite. We can compare in our mind every route with the next one and decide which seems to be more favorable. There are not infinite possible routes, but usually a small number of them.

However, this is not the case when we solve non-typical mathematical problems. There may actually be infinite possible transformations (of the mental model). Here, plausible reasoning is not enough. We need imagination and a sense of harmony, of symmetry of the expressions involved, which guides the transformations we try.

Although the arguments in favor of the existence of mental models consisting of fixed elements are certainly very convincing, experience shows that mental processing is often, but not always based on such models, at least in Mathematics. The formal models used in mathematics are often the result of unexpected leaps of imagination.

Contrary to what is usually believed, the solution of a mathematical problem is not merely a process of combining appropriately the premises of the given problem, as well as some obvious factual information we keep in mind. This misconception is due to the fact that, when doing mathematical exercises, people usually only have to retrace the line of thought of the creator of the exercise and apply standard theory. Many mathematical problems, however, require a great imaginative leap in order to be solved.

In **Appendix 4.2** we will describe a more demanding mathematical problem that demonstrates this fact. Here we will discuss some simpler examples, more accessible to the general reader who has some knowledge of elementary algebra[16].

Problem 1: Let us first consider the problem of finding two integers, such that two times the first one, added to the second one, equals the product of these integers. How can we go about to find solutions to this problem?

First, we may write it in the symbolic form:

$$2x + y = xy$$

Is that a mental model of the problem? If so, it does not give a direct indication of how the problem can be solved. However, writing the equation in the form:

$$0 = xy - 2x - y,$$

we may note that we can transform the second side into a product, if we increase both sides by 2. Then we have:

$$2 = xy - 2x - y + 2 = (x-1)(y-2)$$

This means that the integers, x-1 and y-2 must be equal to the factors of the first side of the equation, which are ± 1 and ± 2. This leads us to the solutions:

$$x=1\pm 1,\ y=2\pm 2 \text{ or } x=1\pm 2,\ y=2\pm 1, \text{ i.e., } x=2, y=4 \text{ or } x=0, y=0 \text{ or } x=3, y=3, \text{ or } x=-1, y=1.$$

[16] The reader who is not very experienced in algebra may omit the rest of this section. A reader with some algebraic experience should first try to find his own solution, in order to become more conscious of the thought processes involved.

But how are we led to the above transformation of the initial equation? What makes us consider the new expression (x-1)(y-2)?

Obviously, it is by no means merely implied by the premises of the problem. It seems that here we utilize previous experiences that occur in the study of algebraic identities. The general expression, xy-ax-by is associated by an experienced mathematician with the product (x-b)(y-a), because he has often met the identity

$$(x-b)(y-a) = xy - ax - by + ab$$

The expression, xy-ax-by reminds him then of the product (x-b)(y-a).

Here we have plausible reasoning which, however, is not simply based on combining factual information directly arising from the premises. It is rather a utilization of experiences that might be useful.

Thus, formalism may provide visual cues for handling a problem, but immediate visual cues are often not there. Sometimes, it is our wish to reach a certain form that makes us try an unusual transformation, as in the second problem below. The crucial mental models used in these cases, are results of a search process that is guided by considerations of utility. They are by no means models that are contained in the initial data, i.e. resulting automatically from them.

Problem 2: Show that if we multiply numbers, which have the form of sums of two squared integers, then the product is also a sum of two squared integers. I.e., for arbitrary integers a, b, c, d, there are always appropriate integers x and y, so that:

$$(a^2+b^2)(c^2+d^2) = x^2+y^2.$$

For instance, since $5 = 2^2+1^2$ and $13= 3^2+2^2$, it is also $65 = 5*13 = 8^2+1^2=7^2+4^2$. How can we prove such a property?

Our usual knowledge about identities does not provide any useful hints. However, if we know something about complex numbers, we note that:

$$a^2+b^2=|a \pm ib|^2, \text{ and } c^2+d^2= |c \pm id|^2,$$

where i is the imaginary unity, $i = \sqrt{-1}$.

Using these identities with the "+" signs, we have then:

$$(a^2+b^2)(c^2+d^2) = |a + ib|^2|c + id|^2 = |(a + ib)(c + id)|^2=$$

$$= |(ac-bd) + i(ad+bc)|^2= (ac-bd)^2 + (ad+bc)^2$$

Using the same identities with "-" signs, or simply replacing in the above expressions b by –b and d by –d, we also have:

$$(a^2+b^2)(c^2+d^2) = (ac + bd)^2 + (ad - bc)^2$$

There is no way to guess such an answer, merely by using the usual algebraic identities we know. Of course, we may also know the above used relations about complex numbers, but usually we do not involve complex numbers when doing calculations with real numbers. We must have some good reason in order to do so. Thus, we achieve such a derivation only by a great leap of imagination, a search for unusual relations.

How do we make this leap? A possibly unconscious impulse may be to try to write the terms (a^2+b^2) and (c^2+d^2) as products. Products can be combined in various ways, so that another combination of the terms might yield a different expression than the initial one.

This can be done, for instance, with the product of differences of squared integers, $(a^2 - b^2)(c^2 - d^2)$, if we use the identities:

$$a^2 - b^2 = (a+b)(a-b) \text{ and } c^2 - d^2 = (c+d)(c-d).$$

Then we have:

$$(a^2 - b^2)(c^2 - d^2) = [(a+b)(a-b)][(c+d)(c-d)] = [(a+b)(c+d)][(a-b)(c-d)]=$$

$$[(ac+bd)+(ad+bc)][(ac+bd)-(ad+bc)]= (ac+bd)^2 – (ad+bc)^2$$

In fact, since $i^2=-1$, we can derive the previous identity again, if in the above relations we replace b by ib and d by id.

The insight which brings us to the first derivation is no product of rational thinking, although we can subsequently give some rational justification for our proceeding. It is rather the result of a "general feeling" that we can obtain something new by expressing quadratic terms as products of linear terms.

This "feeling" is the result of long experience with algebraic expressions. It is based rather on imaginal schemata of handling algebraic forms, which have acquired the status of automatic reactions.

In this case, looking unconsciously for a reduction of the quadratic terms to linear, the mathematician remembers that:

$$a^2+b^2= (a+ib)(a-ib) =|a + ib|^2.$$

This association may come automatically to the mind of a mathematician, due to long experience with such transformations. Transformations that seem unusual and far fetched to other people may have become reflexes for him, just as the manipulations of a car's equipment for an experienced driver.

4.12 The nature of the process of concept formation in Mathematics

However, leaps of imagination guided by experience, such as the above, are not always enough to answer mathematical problems. In many cases it is necessary to create new concepts, which go beyond the already existing ones and are no mere combinations of them.

This is indeed something that mathematicians have to do very often, in order to be able to solve a problem. They have to create new concepts (e.g., the concept of complex numbers) that reveal new relations and symmetries between the mathematical objects studied initially. Such new concepts are not based on finite mental models that combine already existing concepts. They are totally new inventions. What they express are not already existing relations, but rather intentions of the mathematician, which create quantities with totally new properties.

The simplest example is probably that of the negative numbers. The rules of basic arithmetic operations with negative signs already appear in the "Arithmetika" of Diofantos, the first mathematician who used symbolic notation for equations in his written works, albeit a crude one.

It is therefore surprising that, not only Diofantos, who lived around 250 A.D., but even the creator of the modern algebraic notation, Viète (1540-1603 A.D.) neglects the consideration of negative coefficients and negative roots of equations (see [Boyer-Merzbach, 1989]).

The first one to accept the existence of both negative and complex roots seems to be Girard, in a book published in 1629, while even Descartes (1596-1650 A.D.) calls the positive roots "true" and the negative ones "wrong". The concept of negative numbers apparently required 14 centuries to mature, while all their formal properties were already known.

A typical example of the fact that many new mathematical concepts do not necessarily result from the problem under consideration, but are rather an expression of the mathematician's intentions, is the creation of the set of real numbers.

The set of real numbers is a completion of the set of rational numbers (i.e., of the fractions), that goes beyond them. The non-rational ("irrational") numbers, which constitute this completion, are creations of our imagination and of our intentions. We simply assume that converging sequences of rational numbers have a limit number, even if this "number" cannot be a rational number, the only kind known up to that instance. We invent the irrational numbers just in order to provide limits for all converging sequences (see **Appendix 4.3: Creating Irrational Numbers**).

The fact that some of them – like π or $\sqrt{2}$ – have a geometric meaning might be considered helpful in the creation of the set of irrational numbers. But this did not seem so obvious to mathematicians up to the 18th century. In fact, even Isaac Newton gave geometric proofs for many of his theorems of Analysis, because the concept of building a limit was not very precise in his time.

The absence of the concept of irrational numbers was, according to van der Waerden [1988, ch. V], the reason why the Pythagoreans and their successors gave a geometric form to algebra, replacing in this way the arithmetic form which the Babylonians had given it. That is why the first known mathematician, who uses even a crude algebraic symbolism, is Diofantos of Alexandria who lived in the 3rd century A.D., almost a thousand years after Pythagoras.

For instance, the length of the diagonal of a rectangle, whose sides have a length of 1, is $\sqrt{2}$, an irrational number. For the Greeks, "numbers" were only the integers and the fractions. Irrational numbers had not been defined as independent magnitudes. In contrast, the diagonal was a simple, directly constructible geometric magnitude. So, it was simpler for them to speak about this geometric magnitude than about the corresponding number. The Babylonians gave for such irrational numbers fractional approximations, not caring whether these numbers were fractions or not.

According to van der Waerden [1988, ch. VIII], the Babylonian tradition of numerical methods for solving algebraic problems was possibly continued "unofficially" by the Greeks, without appearing in their official writings. The first Greek who mentions in his work, e.g. the product of two cubes, a magnitude that can not have a geometric interpretation, is Diofantos. But even Diofantos does not consider irrational solutions of algebraic equations.

The fact that the creation of a concept may take centuries to reach maturity indicates that the mental process leading to this concept-creation is not based on a finite mental model, a combination of already existing ones. This is certainly true for all mathematical concepts directly or indirectly connected with infinity.

We may also note that sometimes a new concept is invented in order to solve a problem, but a model for this concept comes much later. Although a complex number can be represented as a two-dimensional vector, initially it was only a computational device for solving quadratic and cubic equations. The vector model of complex numbers was created more than two hundred years after their invention. Originally, the imaginary unity was merely a mathematically meaningless visual symbol, i = $\sqrt{-1}$, which made possible the formal "solution" of all quadratic equations. But even Girolamo Cardano, who introduced it for the first time in his book "Ars Magna" in 1545, in a way did not consider it a legitimate “number”. That is why he called it imaginary (see [Davis-Hersh p.196-198]).

A proof that, sometimes, the mental model comes together with the inspiration or even later than the inspiration, is Kekule's dream that led him to discover the ring structure of the chemical bonds of the aromatic carbohydrate compounds. He had been trying to deduce the molecular structure of benzene for a long time. Then, according to his own words [Morton Hunt, p.257]: [One evening] “I turned myself towards the fire and sank in a reverie. Atoms danced before my eyes. Long chains were firmly joined, all winding and turning with a snakelike motion. Suddenly, one of the serpents caught its own tail and the ring, thus formed, whirled before my eyes. I woke immediately and worked on the consequences all night”.

Here we see something that happens often; a model appears together with the discovery and not before it! The discovery is not a consequence of screening a model, but it comes together with some kind of sensori-motor “pantomime”, simply because it is expressed more concisely by this kind of visual-kinetic modeling!

What was not known before Kekule's discovery was that the properties of a chemical compound are determined, not only by its quantitative composition, but also by the arrangement of the various atoms in space. This is probably what his Unconscious was trying to tell him by the vision of the dancing chains of atoms. It was getting the chain of atoms out of the usual linear arrangement along a straight line.

4.13 The Relation of Subjective and Objective Reality

Since so much importance is accorded by Johnson-Laird's theory to correspondence with an objective reality, we should also note here that we cannot always distinguish between subjective and objective. The objective world is partially something we create ourselves, in the sense that we invent concepts in order to describe situations that are abstractions of reality and not the reality itself.

This is often done in Mathematics and Physics. For instance, objectively there is nothing like a straight line. Rather, we idealize the perception of a stretched string or of a light ray into a straight line, by equipping it with certain basic properties, like the fulfillment of Euclid’s parallel-postulate. But as physicists tell us, even light rays, the prototypes of straight lines, are bent in the neighborhood of matter, due to gravity. Thus, a perfectly straight line exists nowhere in nature.

Strictly speaking, there is also nothing like a continuous line in the observable world, since according to Physics and Chemistry all bodies consist of arrangements of atoms, which have considerable distances between their kernels. Rather, we idealize observed shapes, which have no interruptions and sudden jumps, into continuous curves.

Even a continuous trajectory is nothing we actually perceive, but rather an interpretation of what we see. Our eyes register only changing discrete positions of a body. Of course, we may choose to discard the shape of this body by rather concentrating our attention on some specific point on it, say a color spot. But even then, what we actually perceive are the changing positions of this spot, as the receptor cells of the eyes are successively activated by incoming photons reflected from it and nothing like a continuous curve. Rather, we infer the continuity and even differentiability of a trajectory, because it simplifies the description of motion.

Thus, even seemingly elementary concepts are rather creations of our mind than objectively given.

What is more, concept-creation or eventually concept-reformation is something that happens quite often, at least in Mathematics and Physics. A recent book by Moshé Flato [1990], for example, distinguishes the mathematicians into "problem solvers" and "theory makers". The latter are necessarily also concept makers, because a new theory without some new concept cannot exist.

4.14 Typical Examples of Mathematical and Physical Concepts that Constituted a Transgression of the up to then Accepted ones

As an indication for this, we mention here only a few typical cases in which mathematicians had to surpass, to violate the boundaries of the then existing concepts, thus creating new concepts that did not exist (were meaningless, or even contradictory) in the traditional sense (see also [Davis - Hersh, 1983, pp.152 - 157]):

1. The transition from the rational to real numbers, by postulating the existence of the limits of converging sequences of real numbers.

2. The introduction of complex numbers in order to make solvable in all instances (at least formally), initially quadratic and cubic equations and later all algebraic equations.

3. The transition from figurative descriptions of geometric objects to analytic descriptions by means of algebraic equations. The creation of Analytic Geometry by René Descartes has very much enriched the set of curves considered and thus opened the way for the development of Differential and Integral Calculus, which are necessary for the study of such curves. Classical Euclidian Geometry studies almost exclusively circles, straight edged figures and conic sections. But, e.g. even the relatively simple equation $(x^2+ y^2)^2= x^2 - y^2$ describes a curve of the form ∞, whose two rings are no circles.

4. The transition from sums of quantities tending to become infinitesimals, to integrals and from "ultimate quotients of evanescent quantities", as Newton called them [Hersh, 1998, p.287], to limits of quotients, i.e. derivatives.

5. The transition from ordinary functions to generalized functions (distributions, or convolution quotients), such as Dirac's $\delta(x)$ – "function", which is no function at all in the ordinary sense.

6. The introduction of the idea that there are infinite sets with a different "number of elements", i.e. different kinds of infinity, corresponding to different cardinal numbers.

7. The creation of Nonstandard Analysis, which was done independently by Detlev Laugwitz and Abraham Robinson (see [Medvedev, 1998]). Nonstandard Analysis gives a rigorous foundation for the use of infinitesimals as independent objects and allows the application of ordinary arithmetic or algebraic operations, like multiplication or division, on them[17]. In this way, it is no more necessary to consider limits in order to give precise meaning to derivatives and integrals.

Some recent cases in which physicists also had to transcend the boundaries of the existing concepts are, for instance, the following:

1. Einstein's admission, as a general principle, of the fact indicated by the result of the Michelson-Morley experiment; that the speed of light is the same, no matter whether its source moves or is fixed with respect to the observer. This led him to the derivation, by simple algebraic calculations, of the (already known but not understood) Lorentz Transformations, the foundation of Special Relativity Theory.

2. Einstein's admission of the validity of the fact implied by Planck's formula about black body radiation; that energy does not come in arbitrary quantities, but only in multiples of elementary units called "energy quanta". This led him to the hypothesis that radiation consists of individual light particles, the "photons", which allowed him to explain the photoelectric phenomenon. For this achievement he received the Nobel Prize in 1921.

3. Bohr's admission of the fact suggested by Balmer's formula, that electrons can have only certain orbits around the atom's nucleus, which opened the way for the development of Quantum Mechanics. This hypothesis is in clear contradiction to classical physics. According to the classical theory, no individual orbits could exist. An electron orbiting around a nucleus would constantly emit radiation and therefore spiral down, ending up on the nucleus within a fraction of a second, since it would have lost its energy.

In all three cases, one did not have to develop a complex argumentation, but only to reject the then existing conceptual system and adjust the concepts to experimental results.

4.15 Conclusions

[17] This is achieved by extending the definition of the usual arithmetic operations to composite quantities, consisting of ordinary numbers and infinitesimals.

These observations lead us to the position that, in spite of their significance, mental models do not give a close enough view of mental processes. This view is still static, looking upon concepts as somehow complete entities, having a constant form or at least, a constant final form.

In truth, concepts are not only interdependent but also incomplete, evolving steadily throughout our lifetime; this evolution process is an important part of all creative mental processes, as we shall try to make clear in the next chapters.

In order to see that the concepts of everyday language change i.e., evolve with time, let us consider for instance the concept of "freedom". By this I mean not the lexical entry, but the true significance of the word as we "feel" it inwardly. To a three or four-year-old child, "freedom" roughly means "being unconstrained".

To a citizen between twenty and forty years of age, it means "having one's basic rights respected and being able to develop activities". It also means "being able to exercise political influence directly or indirectly by expressing openly one's political views and participating in elections".

To a person between fifty and sixty years of age, it may primarily mean "being able to control the habits and passions of earlier years, being able to accept the non-fulfillment of some of one's initial goals and wishes (internal freedom)".

To a person over seventy years old, it may primarily mean "being still able to be independent of help by other people" and finally "making peace with one's inescapable destiny, death!", "being able to accept death in peace of mind".

So "freedom" is not merely a lexicon entry, but a concept whose various aspects or facets one has to internalize gradually, as one grows older.

What is more, if we accept the thesis that a concept is defined by the way it helps us grasp our surrounding and make use of it, then even very concrete concepts, like "chair", evolve. To a child, a chair is a big object on which it may climb with some risk.

To an active young person, it is simply a commodity, some piece of furniture on which he can rest for a while.

To an older person, it becomes gradually more and more important; it may mean "aching bones" or "restfulness".

Note that even mathematical concepts evolve, in spite of their precise formulation. This is done by extending their application to new fields of Mathematics, i.e. embedding them in new theories of growing abstraction. The concept "derivative", for instance, was gradually extended from ordinary functions to functionals and even to operators.

The study of another controversy among cognitive scientists is indicative of how many difficulties are met in the attempt to create satisfactory models for mental processes. This is the controversy about whether processing of visual and other information in the brain is done in pictorial form or on the basis of "propositions". Since the results of this controversy do not seem to surpass the degree of penetration of mental processes offered by mental models, we will describe it in a separate appendix, **Appendix 4.4**.

There are also certain important philosophical questions which arise in the course of this investigation. One of them, which arises naturally when we consider the nature of the conceptual system, is the question whether the basic concepts which serve as instruments of thought are constant and permanent, i.e. independent of outward experiences or not.

What many people would like to believe is that the basic concepts are independent of all outward experiences, because otherwise, there is no stability of what we believe and think about the world; there is no absolute truth. If the concepts depend on experiences, they may differ from one person to another, from one generation to the next one. Then, it seems that no general consensus, about what is true and what is false, what is right and what is wrong, can be achieved.

The first one to propose that the main abstract concepts have a somehow objective existence, independent from our minds, was Plato. His ideas will be discussed in **Appendix 4.5**.

An observation that seems to support Plato's views is the ubiquitousness, the omnipresence of mathematical concepts in all physical sciences, the uncanny ability of Mathematics to describe all physical phenomena, although it does not seem to have any immediate connection to the physical world. Since our position here is that the conceptual system is gradually formed by inherited biological traits, but on the basis of experience, we must give some explanation for this unusual success of Mathematics. This subject will be discussed in **Appendix 4.6**.

A Short Review

In this chapter we have discussed the contribution of the Cognitive Psychology in the investigation of the mind. As we saw, the predominant view about the mechanism of Long Term Memory is that new data are stored by Procedural Semantics. This means that new meanings, new experiences, are stored as appropriate connections of procedures which describe these new experiences. However, there is great discord about the nature of the elementary procedures, the elementary meanings that are being so interconnected. While the neurologists see analog neuronal procedures, not necessarily translatable in linguistic form, as being connected, many cognitive scientists favor the existence and interconnection of elementary meanings, which they try to discover. Concentrating our attention on Cognitive Psychology, we have considered mainly two plausible theories about the structure of LTM: the theory of Semantic Networks and the theory of Mental Models.

The theory of Semantic Networks, in its initial form, orders the concepts hierarchically, on the basis of the degree of logical abstraction they have and creates logically more composite concepts by connecting logically simpler ones. In its later forms, instead of simple concepts, it rather interconnects more composite mental expressions translatable into phrases (interpretable by phrases). These are called "propositions". However, in none of its forms can it escape the vicious circle in which many concepts determine each other, like e.g. the concepts hen and egg. So it is not clear which is more elementary, more basic.

A second difficulty faced by this theory is that many objects that belong to the same logical category are not its equally typical representatives. Psychological experiments show that people recognize that a sparrow is a bird, quicker than they recognize that an ostrich is a bird. It can circumvent this difficulty by considering the features of a category as alternatives, which are inherited by its subordinate categories only if no contravening data exist. However, this does not make it more convincing because it raises the question of how a specific concept is included in a more general category, if not on the basis of some common features.

The second theory we have considered, the theory of Mental Models, escapes from the vicious circle of mutual determination of elementary concepts by admitting that certain concepts are not determined by others, but only by sensori-motor procedures. Thus, it distinguishes mainly between two categories of concepts: the linguistically decomposable or composite concepts, which it calls "analytic" and the non-decomposable or simple concepts, which it calls "natural kind concepts"; e.g., the color concepts, the taste concepts etc. To this notion we have countered that all concepts, and not only certain among them, have partially a sensori-motor encoding.

Further on, we have seen that the theory of Mental Models is based on a fundamental result of Experimental Psychology. In order to understand a situation and reach correct conclusions with respect to it, we build up mental models of this situation and we handle them mentally, just as we would do with material models. I.e., we do not use any inborn Categorical Logic, as it was believed earlier. This theory thus attempts to reduce all thought processes to a processing of mental models. It also explains the procedures of grammatical and syntactical analysis and understanding of linguistic expressions, basically as processes of a gradual creation of a mental model.

With respect to the creation and inspection of such models, we have noted that it really is a basic part of thinking, but not sufficient to explain how we reach truly original ideas. Considering elementary algebraic problems we saw that, even in these, the transformations which lead to the solution do not result either by simple inspection of some formal model or by simple use of known relations connected in some obvious way with the problem. They are rather the product of an effort on our part to reduce each problem to a deeper symmetry.

Considering more closely many mathematical concepts, like the concept of negative numbers, irrational numbers and complex numbers, we have established that these concepts were not merely a composition or completion of other, known concepts, but rather came into contradiction, in immediate conflict with up to then accepted concepts. However, they were necessary because the mathematical problem considered did not have a solution of the already acceptable type. A transgression was therefore necessary; it created a new concept of solution, a new kind of "number" which was of course in conflict with the already known ones, but gave a "solution" to the problem under consideration. This new concept was not a synthesis of the old ones, but simply constituted a formal expression of the mathematician's motives. We have also established that similar kinds of new concepts, contradicting the accepted ones, also resulted during the development of Physics.

Finally, we reached the conclusion that the static view about concepts, which considers them as invariable, is wrong. Not only the concepts "number" or "solution" are evolving, but all concepts are subject to a gradual, albeit possibly slow, evolution during our lifetime, as we have tried to show by some examples.

BIBLIOGRAPHY

Badeley, Alan: Your Memory, A user's guide, Penguin, 1983

Boden, Margaret A.: Computer Models of Mind, Cambridge U.P., Cambridge, New York, 1988.

Boyer, Carl B. – Merzbach, Uta C.: A History of Mathematics, Wiley, N.Y., 1989.

Cohen, B. and Murphy, G.L.: Models of Concepts, Cognitive Science, 1984, vol.8, pp.27-58.

Collins, A.M. and Quillian, M.R.: Retrieval Time from semantic Memory, Journal of Verbal Learning and Verbal Behavior, v.8, 1969, pp.240-247.

Collins, A.M. and Quillian, M.R.: How to make a language user. In E. Tulving and W. Donaldson (eds.): Organization and Memory, Academic Press, New York, 1972.

Collins, A.M. & Loftus, E.F.: A spreading-activation theory of semantic memory, Psychological Review, vol.82, 1975, pp.407-428.

Craik, K.J.W.: The Nature of Explanation, Cambridge U.P., Cambridge, 1943.

Damasio, Antonio R. and Hanna: Brain and Language, Scientific American, vol.267, Sept.1992, pp.88-95.

Davis, Philip J. & Hersh Reuben: The Mathematical Experience, Penguin Books, 1983.

Donaldson, Margaret: Children's Minds, Fontana/Collins, Glasgow (9.impr.), 1983.

Doob, Joseph L.: The Development of Rigor in Mathematical Probability (1900-1950), The American Mathematical Monthly, vol.103, 1996, pp.586-595.

Dreyfus, Hubert: What Computers Still Can't Do, MIT Press, Cambridge, 1992.

Flato, Moshé: Le pouvoir des Mathematiques, Hachette, 1990.

Fodor, J.D., Fodor, J.A. and Garrett, M.F.: The psychological unreality of semantic representations, Linguistic Inquiry, vol.4, 1975, pp.515-531.

Fodor, J.A., Garett, M.F., Walker, E.C.T. and Parkes, C.H., Against Definitions. Cognition, vol.8, 1980, pp.263-367.

Freedman, J.L. and Loftus E.F.: Retrieval of words from long-term memory, Journal of Verbal Learning and Verbal Behavior, v.10, 1971, pp.107-115.

Hersh, R.: What is Mathematics Really?, Vintage, London, 1998.

Horgan, John: The Undiscovered Mind, Touchstone, N.Y., 1999.

Hunt Morton: The Universe Within, Corgi Books, London, 1984.

Johnson - Laird, P.N.: Mental Models, Cambridge U.P., Cambridge, London, New York, 1983.

Johnson - Laird, P.N.: The Computer and the Mind, Fontana, London, 1988

Katz, J.J & Fodor, J.A.: The Structure of Semantic Theory. Language, vol.39, 1963, pp.170-210.

Keil, F.C.: Semantic and Conceptual Development: an Ontological Perspective, Harvard U.P., Cambridge, 1979.

Klatzky, Roberta L.: human Memory, Freeman & Co., San Fransisco, 1980.

Loftus, E.F. and Freedman, J.L.: Effect of category-name frequency on the speed of naming an instance of the category, J. of Verbal Learning and Verbal Behavior, v.11, 1972, pp.343-347.

Mc Clelland James L. & Rumelhart David E.: An Interactive Model of Context Effects in Letter Perception - I. An Account of Basic Findings, Psychological Review, v.88, 1981, pp.375-407.

Medvedev, F.A.: Nonstandard Analysis and the History of Classical Analysis, The American Mathematical Monthly, vol. 105, 1998, pp.659-665.

Miller George A.: Spontaneous Apprentices: Children and Language, Seaburg Press, New York, 1977.

Miller, Arthur: Imagery in Scientific Thought, MIT Press, Cambridge, 1986.

Miller, G.A. and Johnson-Laird, P.N.: Language and Perception, Cambridge U.P., Cambridge, 1976.

Moscowitz Breyne Arlene, The Acquisition of Language, Scientific American, Nov.1978, pp.82-96.

Quillian, M.R.: The Teachable language comprehender. A simulation program and theory of language, Communications of the Association for Computing Machinery, v.12, 1969, pp.459-476.

Rock, J.: Orientation and Form, Academic Press, New York, 1973

Rosh, E., Mervis C.B., Gray W.D., Johnson D.M., and Boyes-Braem, P.: Basic Objects in Natural Categories, Cognitive Psychology, 1976, vol.8, pp.382-439.

Rosh, E.: On the internal structure of perceptual and semantic categories. In T.M.Moore (ed.): Cognitive Development and the Acquisition of Language, Academic Press, New York, 1973.

Rumelhart David E. & Mc Clelland James L. : An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model, Psychological Review, v.89, 1982, pp.60-94.

Russell, Bertrand: The Problems of Philosophy, Oxford U.P., 1912.

Russel Bertrand: History of Western Philosophy, Unwin Paperbacks, London, 1982 (first publication in 1946).

Schank R.C.: Dynamic Memory: A Theory of Learning in Computers and People, Cambridge U.P., 1983

Schank R.C.: Explanation Patterns: Understanding Mechanically and Creatively, Erlbaum, Hillsdale, N.J., 1986.

Sells, S.B. and Koob, H.F.: A Classroom Demonstration of "Atmosphere Effect" in Reasoning, Journal of Educational Psychology, vol.28, 1937, reported in: Krech, David & Cruchfield Richard S.: Grundlagen der Psychologie, vol.2, Verlag Julius Beltz, Weinheim, Berlin, Basel, 1971.

Smith, E.E. & Medin, D.L.: Categories and Concepts, Harvard U.P., Cambridge Mass., 1981.

Templin, M.C.: Certain Language Skills in Children: Their Development and Interrelationships, University of Minnesota Press, Minneapolis, 1957.

Van der Waerden, B. L.: Science Awakening I, Kluwer, 1988.

Wason, P.C.: The Theory of Formal Operations: A Critique. In B.A.Gerber (ed.): Piaget and Knowing: Studies in Genetic Epistemology, Routledge & Kegan Paul, London, 1977, pp.119-135.

Chapter 5
HOW DOES THE CONCEPTUAL SYSTEM DEVELOP?

5.1 Basic Facts from Developmental Psychology

Before trying to produce a model (or rather, a modeling frame) of Long Term Memory that fits the previous observations, i.e. takes into account how LTM develops, let us first discuss some findings of Developmental Psychology, which emphasize and characterize these developmental aspects of the concepual system.

It has already been observed by early researchers in this field (see [C. & W. Stern, 1907 and 1965]) that new concepts are not acquired in an integrated form, but develop gradually from preliminary stages, which are called "preconcepts". Jean Piaget especially, has done much work to show that different aspects of a concept are understood and used by a child at different ages, whereas the understanding of a concept, i.e. somebody's competence, comes generally earlier than his ability to use it, i.e. his performance.

Piaget's experimental work is very important, because it made clear that the understanding and usage of concepts by children develop gradually and showed that there are various developmental stages.

However, Piaget's theory relies on an analysis of the concepts into logical parts and on a classification of different stages of concept development into logical categories of growing abstraction. Later researchers have raised objections to such categorizations, because they find that logical categorization does not correspond to the phenomena.

But let us consider in greater detail Piaget's observations, how he interprets them and what objections have occured to such interpretations.

5.2 The Stages of a Child's Mental Development According to Piaget

5.2.1 Sensori-motor Thinking (0-2 years)

According to Piaget, up to the age of two years, a child develops his sensori-motor thinking.

At first, he learns just how to react to outward stimuli in order to fulfill basic needs. But later, acquired sensori-motor reaction schemata are also applied in a more general context than that which initially led to them.

He learns, e.g. to suck on his mothers breast when he is fed, but later he also sucks on various objects of his surrounding, registering their shape, size, taste, hardness, temperature etc. By the end of the first year his actions become predominantly exploratory. They are clearly intentional and experimental.

Sensori-motor action schemata are, in this context, used in order to build up representations of the present situation in the mind, which are then used to solve problems occurring during this exploration.

Piaget notes, e.g. about a girl of sixteen months, that when a drawer was not open enough to look in, she first opened her mouth wider and wider, as if it was representing an opening box, and then she pulled out the drawer. In this way, old sensori-motor representations operate in new situations having some similarity with the past experiences. Thus the child gradually builds up interaction models of the objects in his surrounding, which he uses as bases both of recognition of these objects and of planning actions by first performing mental experiments with them. Thus, sensori-motor thinking is an internalized action.

5.2.2 Symbolic Thinking (1.5-5 years)

a. General Description

As a second step in mental development, lasting from one and a half to five years of age, Piaget sees the emergence of symbolic thought. The built-in action schemata gradually become symbols.

They are thus applied not only to the actual objects that play a crucial role in the child's life, but also on imitations of them. Thus, beads can be used as "fake" food and the child "feeds" with them his dolls during play. He uses such replicas to organize the whole scenery of a certain kind of activity. Leafs can serve as plates etc. This new stage of development is characterized by the dissociation of a certain kind of activity from the objects actually being used or the persons participating in it. Dolls replace people just as beads replace food.

This is already an instance in which there may be objections to Piaget's interpretation. "Symbolic" interpretation of some activity is in some sense present already in much younger children, since a baby sucks not only on a breast or a feeding bottle, but also on other objects that have nothing to do with food.

This kind of activity is restricted in babies merely by the fact that they have not yet developed certain dexterity related skills, which would allow them a larger repertoire of actions.

b. Initial Stages of Linguistic Development

During the same period there is a dramatic increase in the child's use of language. But language is at first just an accompaniment of actions that are based on imagery.

Symbolic representation emerges from sensori-motor representation at about the age of two years. The end product of sensori-motor thought is internalized imitation. This internalized imitation acquires now a symbolic status and can be evoked in absence of the actions that initially created the imitations. Piaget calls this process "deferred imitation". Deferred imitations are connections of images and actions in unique symbols, which the child can evoke in order to conduct preconceptual thinking [Richmond, p.29]. At first, the child's image-action schemata have no corresponding linguistic expression. This happens gradually by invention of appropriate words or by admission of words used by the adults.

Gradually, language ceases to be an accompaniment and starts being used for the reconstruction of some past action, thus providing a medium for non-imaginal, symbolic representation. The word is no longer merely a part of the action, but begins to function as a sign that evokes it.

Soon the words begin to become, not designations of individuals, but class-names. Thus, for a child, at first a dog may only be the puppy with which he plays, but later he learns that other animals are also dogs, when he hears his mother refer to them using the name "dog". His puppy is then differentiated by the additional name "Fido", but the child's images of Fido act both as representations for an animal and for a class of such animals. They are all Fido-like.

Note that adults sometimes also call a certain race of dog "Lassie" instead of "Collie", because they know this race of dogs from the TV series whose "star" is such a dog, called "Lassie".

c. The Preconcepts According to Piaget

At this stage, symbolic representations achieve neither true generality, nor true individuality. The child's representations during this period are uniquely personal. That is why they are called "preconcepts" by Piaget. According to him, a preconcept is an "intermediary between the imaged symbol and the concept proper" [Richmond, 1970, p.20].

Such a preconcept is characterized by direct identification of representatives of a class, one with the other, without the mediation of the membership to the category. In fact, membership to a specific category is not yet well understood. For instance, it is not clear that bees are also animals. However, even adults consider various members of a category as more or less typical members of it. An ostrich is a less typical bird than a canary, as E.Rosh has demonstrated (see **Chapter 4**). So, this is another instance in which Piaget uses a too rigid logical classification, assuming that fuzziness of logical categories is overcome after a certain age and cannot occur in adults.

In fact, the name "preconcepts" will be used here somewhat differently, for the designation of all preliminary stages of a concept which are seen as always having only a gradual transition to a more developed form, without ever totally losing their previous features and without ever reaching a final form (see next chapter). Previous features are never lost; they are merely differentiated by educational training.

According to Piaget, it is typical of this early period that the child is not yet able to represent a complex arrangement of objects by relating the parts to each other. So he may conceive this experience either as an undifferentiated whole or concentrate on parts or details.

For instance, when a child aged four years and three months was shown a row of eggs put in egg-cups and asked if there is the same number of eggs as egg-cups he replied correctly "yes". But when the eggs were taken from the egg-cups and put together he said that there are more egg-cups than eggs, since the row of egg-cups was longer. When the eggs were spaced out in a longer row, he said that there are more eggs.

d. The Egocentricity of Thinking

Another feature of the early childhood is that the child is yet unable to "see" things from another point of view except his own (egocentrism) and believes that everything around him has the same abilities that he has. He therefore credits the inanimate objects around him with feelings like his own and even with intentional actions. A table, on which he has hit his head, is a "bad" table, because it has hit him.

He is also yet unable to distinguish clearly between thought and action. He therefore believes that his thoughts can change events.

e. The Initial Stages of Understanding Space and Time

The understanding of space and time is also very limited. Space is restricted to the objects surrounding the child and embodied in their shape and relative position, their proximity, separation, enclosure and continuity. The child is, e.g. initially unable to conceive how the outlook of an object changes when the observer's position is changed.

Time is also recognized only on the basis of routinely repeated experiences of everyday life (meals, play, sleep, light and darkness etc). The understanding e.g. of the fact that familiar things are subject to changes, even when we are not present to observe them, is also an aspect of the concept of time, which becomes only gradually understood. Because of its complexity, a full understanding of this concept, to the degree that it is used in everyday life, is reached only around the age of twelve years.

5.2.3 Decentralization of thinking (4-8 years)

The third developmental period distinguished by Piaget is between the fourth and eighth year of age. During this period, social interaction based on increasing competence on the use of language has a considerable influence on the child's development. His point of view is decentralized.

He gradually learns to reorient his mental model of the environment to conceive other relative viewpoints, as they are imposed on him through social interaction. He also orders and relates his representations more in accord with the conceptual structure of language.

Space and time are not yet well understood, but towards the end of this period the child begins to understand that space can be empty as well as full and that periods of inactivity, periods between events, have duration.

5.2.4 The Period of Concrete Mental Operations (7-12 years)

The next period of mental development distinguished by Piaget, is the period between seven and twelve years of age, which he calls period of concrete operations.

Before this period, the child is not yet able to recognize, e.g. that two glasses of different width cannot contain the same amount of liquid when its surface is at the same level in both, but rather when it stands higher in the narrower glass.

Concrete mental operations are, according to Piaget, organized groups of related actions that build an integral whole. Such mental operations have evolved from internalization of physical actions and therefore show properties similar to physical activity. They remain attached to empirical reality and can be basically performed for the solution of problems set in a concrete form. A similar problem presented in an abstract verbal form cannot be solved by the child.

Recent and older experimental results show, however, that the same limitation is also pesent in adults (see below), although Piaget does not seem to think so.

In the period between nine and twelve years the child at last reaches, according to Piaget, the ability to perform formal operations not directly connected with concrete physical situations.

5.3 More Recent Interpretations of Piaget's Experiments

In recent years, the work of many other psychologists working in this field, for instance Tom Bower and T. Trabasso (see [Brown & Desforges, 1979]), has resulted into a review of this classification as putting too much stress on distinguishing absolute developmental stages, while the actual phenomena are more gradual and depend very much on familiarity and motivation.

In a certain period, for instance, competence and performance are present or absent depending on other factors. A young child sometimes uses perfectly structured sentences in order to express something that is interesting and important to him but later, when motivation has languished, he is not able to repeat his own sentences in the previous, perfectly structured form.

As Lois Bloom reports, a child aged 32 months produced sentences like “I am trying to get this cow in there” while playing. But when asked to imitate this sentence on the next day, he only said, "Cow in here" (see [M. Donaldson 1983, p.74]).

Although graduations in the assimilation of a concept exist, as Piaget claims, they rather seem to be graduations in the development than in the acquisition of the concept. They rely more on the context of the whole situation rather than on the acquisition of certain logical features of the concept.

In all these instances, it is not the results of Piaget's experiments that have been questioned but their interpretation, because appropriately devised experiments, similar to those of Piaget, produce very different results.

For instance, Piaget concludes, on the basis of his experiments, that a baby less than 8 months old does not yet understand the permanent existence of objects. When we hide a toy under a pillow, even if he can see what we do, he shows a total indifference to it, as if it does not exist anymore.

On the other hand, Tom Bower has shown that much younger babies still try to catch the toy when we hide it. However, he did not hide it by putting it under a pillow, but simply removed all light from it, i.e. he darkened the location where it was. He therefore concludes that the child is probably yet unable to follow movements before the age of eight months [M. Donaldson 1983, p.27]. This is a typical instance which allows for different interpretations.

Does a baby of four months believe in the permanent existence of objects, or not? If he does, and simply is unable to follow mentally their movements, how does he explain their disappearing when we move them away?

In trying to interpret childrens’ behavior one should be very careful. Especially for very early ages, it seems that any description by use of usual concepts is inappropriate.

The baby has probably neither a clear opinion about existence (or not) of objects, nor is he seeking an explanation for their disappearance. His interest in all objects is limited, both in duration and in the features, or aspects, which he considers.

5.4 Functional Internalization versus Logical Contruction. The Role of Logic

The recent findings of Developmental Psychology seem to favor a view of the evolution of the conceptual system, based on distinguishing stages of functional integration of a concept, rather than stages of acquisition of logical aspects (logical components) of the concept. More specifically, there seems to be no qualitative distinction between the way that children think and the way that adults think.

For instance, the distinction of a period of prelogical thinking (according to Piaget, the period before the age of six years) seems inappropriate. Children seem to be able to make valid inferences, but only if the situation encountered is sufficiently familiar to them. They deal much better with logical problems arising in familiar situations than with similar problems put into an abstract, unfamiliar and unmotivated form for them.

As a matter of fact, neither children, nor adults seem to apply rules of logical inference in order to reach conclusions. Early experiments have already shown very poor performance of adults in use of syllogisms, in abstract form. Such experimental results of Sells and Koob [1937] are reported in [Krech & Cruchfield, 1971, p.468-469]. Similar results have been obtained by Wason & Shapiro in 1971, and Wason & Johnson-Laird in 1972 (see [Wason & Johnson-Laird 1972, p.29-34] or [Johnson-Laird,1983, p.52]). Such results of Riley & Trabasso [1974, p.201-202] are also reported in [Brown & Desforges, 1979, p.59].

5.5 The Use of Mental Models Instead of Syllogisms

More recent, careful observations led to the conclusion that children, as well as adults, first develop a model of the whole situation appearing in the problem and then simply give an answer by observing the model, without any use of rules of inference. See e.g. Wason & Johnson Laird [1972].

Brown & Desforges [1979, p.59], reporting work of Trabasso and his colleagues, state: "[Their work] typically involves teaching subjects the relationship between pairs of quantities from a series and then asking them to answer questions about the series. […] Their data suggest that subjects gradually construct an image-like representation of the series presented. When they are asked a question about pairs in the array, they appear not to operate a step by step logical sequence, but to read out from their image. These processes seem to be at work whether the subject is four years or twenty years of age. The developmental difference seems to be merely that the younger child needs more trials to build his representation of the series. The manner of the building, its mode and the success of its use are independent of age".

5.6 The Development of Language

Another important fact that is observed, e.g. in the process of language acquisition, is that cognitive schemata in young children are initially nonspecific, becoming gradually differentiated into specific ones.

Children apply a newly acquired or created word, at first referring to a certain object, but very soon broaden its usage considerably and apply it non-specifically, in reference to very different objects having some (often minor, from the adults point of view) common feature.

For instance, as Eve V. Clark has observed, "bow-wow" was first used by a child in reference to a dog, but soon also in reference to a fur piece with glass eyes, to the father's cufflinks, to pearl buttons on a dress and to a bath thermometer. Similarly "vov-vov" at first also referred to a dog, but soon also to kittens, hens, all the animals at a zoo and a picture of pigs dancing (see B.A.Moscowitz, [1978, p.92, and pp.94-95]).

Note that both words were initially used referring to the same object, a dog. This, and the fact that subsequently they were used in connection with whole classes of other objects, imply that they were not representing those objects as entities, but rather as having some common feature, some common perceptual structure.

Newly acquired rules of grammar, e.g. the building of plurals by the attachment of the suffix "-s", are also applied at first non-specifically in all possible cases, e.g. "toy-s" but also "man-s" (see Moscowitz, [1978, p.93]). Eventually, of course, "meanings" and rules are restricted to their common usage.

Similar observations have been made on even younger children. Thus, Tom Bower [1983, pp.83-84] uses the term "abstract" referring to the perceptual capacities of young children and stresses the fact that the perceptual system becomes more and more specified in the course of development.

With respect to the phenomenon of "non-specificity", described here, the term "overgeneralization" is often used instead. Here however, we avoid this term because it is somewhat misleading. It might be taken to imply that there is some process of generalizing involved. But it is highly improbable that the child actually uses some kind of logical processing in order to "generalize" the meaning of a word. It is much simpler to assume that the child uses any name that comes handy, in order to refer to objects with some perceptual similarity important to him at that moment.

5.7 The Significance of Interaction for Learning

A further important observation is that children's learning is based on interaction with their surrounding.

Language, for instance, cannot be learned merely on the basis of sensory inputs, but only interactionally. As B.A.Moscowitz [1978, p.89] reports, efforts to teach a nonimpaired child of deaf-mute parents to speak, by exposing him regularly to television programs, failed, although the child learned to communicate fluently in sign language.

Similarly, it has been experimentally established by Held and Hein (see the last part of Appendix 3.1) that young kittens learn to interpret what they see only when they are allowed to combine visual and motoric experience actively, through their own movement.

5.8 The Stages of Mental Development

We have thus the following basic picture of human mental development:

a. The first kind of mental activity is simply the coordination of sensations and motoric skills in order to achieve a more efficient use of inborn reflex mechanisms.

b. Further coordination of sensations and motoric skills leads to sensori-motor mental structures, which enable an increasing control of the surrounding.

c. All mental activity is based on interacting with the surrounding (later on, we shall give some more indications for this). This view is also supported by the great influence of familiarity and motivation on the performance of the child in similarly structured tasks.

d. Perceptual structures develop from nonspecific to increasingly specific ones.

e. All these processes are driven by the child's motivation.

Motivation is a very important factor of the development of the memory mechanism, which is strangely often neglected by cognitive psychologists. It starts from the very elementary inborn drives (i.e. elementary control hierarchies of our reactions), getting more refined as the conceptual system develops.

It is just as strongly responsible for any processing in the brain, as the outward stimuli are. Most of the time, i.e. unless the sensory input is very strong, it is our motivation that selects which of the incoming stimuli will be more closely attended to.

Motivation also decides in which direction the further processing of incoming information will go. If we imagine the memory as some kind of network, then motivation is responsible both for keeping the activation in a certain part of the network, when the outward stimulus has vanished, and for spreading the activation even further in this network in a certain direction rather than others.

An indication for this is, for instance, the already mentioned observation that children can perform linguistically much better when they have a strong motivation than when they are not motivated enough (M.Donaldson, [1983, p.74]). But a further discussion of the role of motivation will be given later.

5.9 Are there Discrete Stages of Maturation? Indications from the, so called, "Magical Thinking"

Since the research of Jean Piaget concentrates on distinguishing qualitatively different developmental stages in a child's mental development, there is an important question we have to answer clearly before we can consider possible mechanisms for mental phenomena:

Are these stages of mental development final stages of maturation, once they are reached, or does the previous, less differentiated way of thinking persist even later in some instances?

In the first case, one might reach the conclusion that preliminary developmental stages are simply somehow eliminated from the mind, i.e. that the better operating schema replaces the more primitive one.

In the second case, one must seriously consider the possibility that all mental operation schemata once acquired are never erased from the mind but only covered by later ones.

A first indication pointing to the most likely alternative is provided by the observation mentioned earlier, that adults perform much better when solving logical problems, if the problem is set in a concrete context, just like children do. The precise logical training of later years does not seem to replace the old habits of performing this task.

Another indication is provided, e.g. by the so called, "magical thinking" of a very young child, i.e. his tendency to animate and personify all surrounding things, attributing to them good or bad intentions towards him. For instance, a ball which bounces back and hits him is a "bad ball" that should be punished.

Similarly, at a very young age, a child seems to expect inanimate things to respond to his commands. On the other hand, he would be greatly surprised and frightened if they really did so [Schenk-Danzinger, 1971, p. 67-68]. Nevertheless, he seems to speculate whether such a thing can happen or not. Magical thinking and practical experience about what is possible and what is not possible seem to exist side by side.

Does some leftover of this magical thinking persist in adults? The first reaction would be to say: "No!" But then, why do so many people believe in the validity of horoscopes? How can inanimate things like the stars or planets influence our destiny? Horoscopes tell us, for instance, that on certain days our contacts with other people will be very successful. How can the stars make people like or dislike us? Why do some other people believe that someone has the "evil eye" and can cause harm to others by merely wishing such a thing to happen? Isn't the belief in the "evil eye" similar to our belief that we can make things happen by merely wishing so, which we had at a very early age?

It seems that magical thinking persists even in adults, although practical experience and logical training tell one that certain things cannot happen. Possibly, that is why he attributes the causes of unpredictable events that happen in one's life far away, in the stars, which have nothing to do with the environment from which we usually obtain our practical understanding of the world.

Various observations of the above kind indicate that childish thinking is never really overcome, but merely covered by the strict training in rational thinking, which every child of the Western Society has to undergo while he grows up. More primitive groups of people do not exhibit such features at all. Their thinking during their whole life is more "child-like" than "adult-like".

The fact that "magical thinking" and practical experience can exist side by side, even in adults, is documented in an amusing way by Schenk-Danzinger [1971, p. 68], who recalls an incident with an Albanian bus driver. When the machine of the bus became overheated in a country road, releasing steam, at first he started kicking, boxing and abusing it. This continued for some time so, that it seemed doubtfull whether he knew how to deal with the problem that had occurred. Then, after some time he grew calmer, he gave a can to a young passenger and sent him to fetch water from a nearby farmhouse.

These observations imply that what seem to be natural developmental stages of the mental system, are rather the results of imposing a certain way of thinking on the child.

What kind of models of the mental system support these observations? In the following section we will give some general rules for the structure of such models, without going very far with specific details.

A Short Review

In this chapter we have seen what is known about the mental development of the child on the basis of the findings of Developmental Psychology, which studies how the mental system of a child develops as it grows. In the beginning of the 20th century it was already observed, by the first researchers in this field, that new concepts are not acquired instantly in an integrated form, but develop gradually from preliminary stages, which were called "preconcepts".

A major contribution to this field was the work of the Swiss psychologist Jean Piaget. He distinguished various developmental stages of the basic concepts, like space, time, quantity, number, and determined during which period of the child's life the familiarity with the corresponding developmental stage of the concept is reached. Piaget has analysed the acquirement of concepts in logical stages of growing abstraction and has based his observations and experiments on this categorization. However, more recent researchers think that the developmental stages of the concepts are not subject to an absolute classification in logical categories, nor to a clear distinction of time periods of assimilation of each aspect of a concept. In some intermediate stage of linguistic development, for instance, a child is able or unable to construct a grammatically well structured sentence, depending on the strength of the motives that make him speak.

The basic developmental stages of the mental system are, according to Piaget, four:

a. The period of sensori-motor thought (0–2 years), during which the child learns to react to outward stimuli.

b. The period of symbolic thought (1.5–5 years), during which the acqired action schemata become gradually independent from specific objects and persons and are more widely applied. Various objects are now used as symbols of persons or objects of vital importance to the child and he tries to apply the sensori-motor schemata he knows, to them as well. Thus, e.g. beads may symbolize food and the child feeds a doll with them. However, at this point there is already an objection of more recent researchers, because even a child of a few weeks does not suck only on his mother's breast or on the feeding bottle, but explores all objects by putting them in his mouth. In this way he can feel whether they are hot or cold, hard or soft, whether they have a taste or are tasteless etc.

During the same period, the child begins to use language. We can observe that cognitive schemata are initially used nonspecifically and become only gradually differentiated to specific ones, in this activity as well. Of course, children use a newly acquired or created word at first referring to a certain object. However, they soon use it also in reference to very different objects having some common feature and only gradually restrict its use to the commonly accepted one.

Initially, the linguistic expressions merely accompany the child's actions, but later they are used for the mental reconstruction of past actions. The words do not merely accompany actions, but become self-sufficient symbols. The child's thinking during this period is egocentric. I.e., the child is not yet able to conceive things from a point of view other than his own.

c. The period of decentralized thinking (4–8 years), during which the child acquires increasing dexterity in the use of language and so has also greater flexibility in social contact with other persons, especially children.

d. The period of concrete mental operations (7–12 years), during which the child is gradually able to perform mental operations easily, e.g. comparing numbers etc. In this way, he is able to complete the understanding of various concepts. For instance, now he understands that the quantity of a liquid does not change if we pour it in a receptacle of different shape.

Various other experimental observations in young children, as well as in adults, show that the way of thinking of children is not different from that of the adults. Adults are simply more dexterous in the use of these procedures. Thus, not only adults, but also young children, construct mental models in order to reach logical conclusions. Only young children are unable to construct and study successfully more complex models.

However, the same is also true for features of the early childhood. Even if it may not seem initially plausible, traces of them persist in adults. Consider, for instance, the so called "magical thinking" of very young children, the fact that they attribute good or bad intentions to inanimate objects, as well as the ability to influence their lives. Although adults believe that they are beyond this stage of mental development, there are many of them who seek the help of astrologists, card readers and exorcists, in order to solve their personal problems. There are even more who maintain various superstitions about the magical powers of certain numbers etc.

BIBLIOGRAPHY

Bower, Tom: The Perceptual World of the Child, Fontana/Open Books, 1983.

Brown, Geoffrey & Desforges, Charles: Piaget's Theory. A Psychological Critique, Routlege & Kegan Paul, London, 1979.

Donaldson, Margaret: Children's Minds, Fontana/Collins, Glasgow (9.impr.), 1983.

Johnson Laird, P.N.: Mental Models, Cambridge U.P., Cambridge, 1983.

Moscowitz, Breyne Arlene: The Acquisition of Language, Scientific American, Nov.1978, pp.82-96.

Richmond, P.G.: An Introduction to Piaget, Routlege & Kegan Paul, London, 1970.

Riley C.A. & Trabasso, T.: Comparatives, Logical Structures and Encoding in a Transitive Inference Task, Journal of Experimental Child Psychology, vol.17, 1974, pp.187-203.

Sells, S.B. and Koob, H.F.: A Classroom Demonstration of "Atmosphere Effect" in Reasoning, Journal of Educational Psychology, vol.28, 1937, reported in: Krech, David & Cruchfield Richard S.: Grundlagen der Psychologie, vol.2, Verlag Julius Beltz, Weinheim, Berlin, Basel, 1971.

Schenk-Danzinger, Lotte: Entwicklungspsychologie, Österreichischer Bundesverlag für Unterricht, Wissenschaft und Kunst, Wien, 1971.

Stern, William and Carola: Die Kindersprache, Wissenschaftliche Buchgesellschaft, Darmstadt 1965 (first edition in 1907).

Wason, P.C. & Johnson Laird, P.N.: Psychology of Reasoning: Structure and Content, Harvard U.P., Cambridge Mass., Batsford, London 1972.

Chapter 6

EVOLUTIONARY MODELING OF THE MEMORY.

All previous observations lead to a certain frame for all possible models of Long Term Memory.

We have seen that the so called "sensory encoding" or better yet, the "sensori-motor encoding" of certain basic concepts with physical content, seems indispensable.

We have also seen that the most fundamental concepts (those of space, time, quantity, numerical amount etc.) are not created in the mind as indipendent entities in a short time interval, but they are assimilated gradually in a period of many years.

Furthermore, we have seen that these processes of gradual empirical assimilation do not operate only in children, but also in adults. The difference between the way of thinking of children and adults is not qualitative, but rather quantitative. Adults handle the same processes of cognitive processing more effectively, more dexterously.

What kind of mechanism for long term memory[18] do these observations indicate? In order to discern it, first of all we note that, for a model of the memory system, it is necessary not only to give an account of how this system is structured, but also of how it is gradually built up in the course of its development. This is a question that most theorists overlook or answer superficially.

It is not enough to assume that there are certain concepts with sensori-motor "encoding", or even that all concepts have also some sensori-motor component which is covered by a final superstructure of logical relations to other concepts or logical categories. As the investigations of Piaget and his successors have shown, there are various stages of maturation of the fundamental concepts and not only a dual relation of sensori-motor base and logical superstructure.

6.1 The Empirical Encoding in Interactional Schemata

Assuming that the Long Term Memory system has a network structure, let us therefore discuss how such a semantic network might be built up. Does it possibly start by building up certain central propositions or concepts (with their interconnections) and then proceed to densen this network by building up other, secondary nodes and connections?

[18] With respect to the Short Term Memory we have already seen in **Chapter 3** that it is based rather on an activation of the locations of the brain where the corresponding Long Term memories are stored. Long Term memories, however, result from new neuronal connections.

But then, which are the primary propositions, or concepts? How can they be clearly defined by only very few initial interconnections? How can they be clearly demarcated with respect to other, related concepts?

Obviously, there can be no clear definitions of concepts at the initial stages of the development of the conceptual system. What is then the nature of these initial conceptual structures?

A plausible assumption that could answer this question, is that the (chronologically) first connections (nodes) of the memory network are simply associative connections based on interactive experience, which coordinate sensory and motor mechanisms.

The first preconcepts are thus quite nonspecific sensori-motor structures, becoming increasingly specific by the development of further empirical associations, which differentiate the original memory structures. Of course, certain such connections must be present, i.e. already formed, at the moment of birth. The brain is extensively prestructured at birth as reveals, e.g. the ability of children born blind, to smile or cry.

The whole process is then some kind of evolution, rather than complete restructuring at each stage. It seems reasonable to assume that previous developmental stages of a concept are not replaced by later ones, but remain in the memory as permanent constituents of the concept.

Note that the associations are meant here as empirical associations (based on experiences) and not logical associations (connections of logical categories), as they are in most network models. It is not logical components (categorical constituents) of some concept which are being interconnected, but various sensori-motor schemata that tell us how we can interact with an object, i.e. how we can perceive it, touch it, feel it, handle it or eventually, follow its movements mentally.

The young child distinguishes between coincidental associations and real ones partially on the basis of accumulated experiences and partially on the basis of his efforts to explore the surrounding, efforts that reveal which associations are coincidental and which are real. When two phenomena appear repeatedly close to each other, either in space and time or only in time, then they are connected associatively.

Real, i.e. persistent empirical associations become in this way parts of the system of Long Term Memory. Further associative connections are then gradually built up, improving the coordination of various functions of the organism and thus reducing the haziness of the initial structures.

6.2 The stages of Concept Formation

This leads to the question: how do concepts, as we usually understand them, appear in this system?

This happens when we start developing words and linguistic expression. The words are our marking system for concepts, which makes them discernible, while language is the system for expressing more composite meanings by means of structured sets of words.

This does not stop the process of concept evolution. As we know, children learn to use words around the age of two to four years, but they learn the correct usage of these words, or of the concepts that underlie them, only gradually and for many more years afterwards.

But how do associative connections form the concepts? All associative connections are stages of concept formation, reducing the initial very hazy and nonspecific conceptual patterns to more specific ones.

This is better explained by indicating very roughly how concepts like "cup" and "waterglass" may develop. A preliminary stage of both these concepts, i.e. a common preconcept of them, is a memory structure having roughly the "meaning" (playing the functional role) of [container]. The brackets here […] indicate some meaning, which is not necessarily attached to a certain word. It may not have a name.

[Container], from this point of view, is not a logical category including all vessels but a sensori-motor associative neural connection, showing how we can interact with a vessel-like object. Thus, [container] is also a preconcept of any other perceived object which might be used as a container, for instance an empty coconut or walnut shell or the calyx of a flower (note that in German "Kelch" means both "calyx" and "goblet").

The preconcept [container] is further differentiated by additional structures, telling us how we interact with a cup or a warterglass. Thus, the concept [cup] is determined by the additional features [intransparent], [something we can hold by a handle], etc. and [waterglass] by features like [transparent] etc.

Not all these structures have a name, i.e. not all structures can be attached to a certain word. For instance, the preconcept [handle], appearing in this context, is only one of the initially nameless sensori-motor structures to which the name "handle" is attached. There are also other, different, initially nameless schemata, which acquire later the name "handle". These schemata do not have a common structure. The handle of a bag is different from that of a sliding door. The first is a strap, while the second is inflexible. The name "handle" is imposed on them simply because we can use that attachment of each of these objects in order to hold them using our hand and "handle" them in general.

The concept "handle" is thus determined by various preconceptual sensori-motor schemata, associated with any salient protuberance or projection of an object, which we might grip, push or pull.

Words are not absolutely necessary for thinking, although they facilitate it by allowing the construction of more complex models of the surrounding. As we very well know, a young child for instance, starts using names for objects quite some time after he has developed many abilities of orientation and interaction with the surrounding. Young children start using words around the age of two years, well after they have developed considerable abilities to control their surrounding.

Consider, for instance, the following incident: my daughter, then aged 18 months, wanted me to go with her through an illustrated book, naming the objects shown in it. She could not yet tell me what she wanted by speech. What did she do? She simply took the book and my finger and pointed my finger to the illustrations, thus indicating perfectly clearly what she wanted, without using words.

Another good indication that words are not necessary prerequisites of meaning, is that, as we shall discuss later, we also communicate by gestures and facial expressions, without having any adequate linguistic description for them (see [D.Lewis, 1980]). All these gestures and expressions must be stored in the LTM in a non-linguistic form.

One might think here of rejecting the term "concept" and using some other term, not so strongly associated with words, at least for the first stages of mental development. However, since words are related to more developed stages of the memory system, it seems more appropriate to adopt the old term "preconcepts", in order to describe the various developmental stages of a mental structure until it reaches a very specific form, referring to a certain word-concept.

6.3 The Main Features of Semantic Networks of this Kind

The main features of this sketchy model (or, if you like, modeling schema) of the LTM mechanism can thus be outlined in the following way:

1. For the encoding of an object are not used any of its outward, supposedly "objective" characteristics, but the way we can handle it (touch it, feel it, move it, etc.), look at it, or react to its possible changes.

The "model" that corresponds to a certain concrete object has no "similar structure" to some alleged "objective form" of the object, but is simply an encoding of the way we can interact with it by using our senses and motor skills.

Similarly, the encoding of a concept referring to some action is a description of how we can perform or mentally follow (e.g. as in the case of flying objects) that action and the encoding of a property is a description of how we perceive that property.

2. If two objects can be partially handled or perceived in the same way, then this partial mechanism (structure) is a common "preconcept" of them. The encoding does not start all over again each time, but uses already developed structures as much as possible.

3. Words, i.e. names for things, actions, etc. are simply tags placed on some nodes of this network, when it has already been developed to some extent. They are necessary for communicating and for thinking in logical categories, but they are not necessary for the storage (encoding) of meanings in their initial form. They rather serve for the refinement of meanings by

participating in the construction of more composite, more complex semantic structures (see next chapter).

The fact that words are, at least initially, not closely connected with meanings is obvious from the fact that young children often invent their own words as names for objects or actions, which they gradually abandon when they adopt the corresponding words used by the adults. What is more, the words do not have a clearly determined meaning from the beginning, since the fundamental concepts develop gradually in a child's mind, during an interval of twelve or more years.

Even abstract concepts may not have a full linguistic encoding. Although we use definitions for them, we also "encode" some of their properties merely as processes without a name, e.g. as in the case of the concept "continuous function", as we have seen in the chapter on Cognitive Psychology. Riding a bicycle can certainly be described in linguistic terms, but the aptitude of keeping balance on two wheels is not linguistically encoded.

A further indication for the lack of full linguistic (or propositional) encodings for all concepts is that we use the same name for very different concepts (e.g. "foot" of a man, "foot" of a chair, or "foot" of a hill). This shows that the name is not as important for meaning as the underlying structure.

Certainly, not every partial memory structure has a name. Some connections exist only "imaginally", without having names, e.g. known faces or landscapes. This is why we are able to recognize faces, landscapesor familiar patterns of action (movements etc.) without being able to describe them. Obviously, they are not stored in the memory in a linguistic form or a form related to it.

The lack of names is also the reason why many contents of the memory are not easy to trace. They are simply not addressable linguistically. They have no linguistic "adress". They are not placed in a certain "adress" of a network of linguistic structures.

4. The "meaning" of preconcepts has increasing haziness, increasing non-specificity as we go to deeper (earlier) stages of concept formation. Thus, very different concepts may have some common preconcept, or rather, preconceptual structure. An indication for this is the way, in which young children initially "overgeneralize" the meaning of words.

Eve V. Clark of Stanford University has observed that, for instance, the lexical item "mooi" was first used by a child in reference to the moon, but soon also in reference to a cake, round marks on a window, round shapes in books etc. (see [Moscowitz, 1978, p.94-95]). Similarly, as mentioned in the previous chapter, the lexical item "bow-wow" first referred to a dog but later also to a fur piece with glass eyes, father's cufflinks, pearl buttons on a dress etc., while the lexical item "vov-vov" also first referred to a dog but later to kittens, hens, all animals at a zoo, etc.

The meaning is then gradually narrowed down until the word refers to a particular object. The word itself, which is often created by the child on the basis, e.g. of echoic impressions, is then also replaced by the one used by adults in everyday language.

6.4 A Comparison of the Expressions "Over-generalized Usage of Words" and "Non-specific Usage of Words"

As we have noted in the previous chapter, the term "overgeneralization", commonly used instead of "non-specificity", is somewhat misleading because there is no process of generalization involved. The child merely attaches any name coming handy to already more or less developed concepts, guided in his choice of names by common preconceptual content. Even different names, such as "bow-wow" and "vov-vov", may be attached to the same object, depending each time on the motivation, the point of view, i.e. on the prevailing features in the child's mind, at a given moment.

It should also be noted that non-specificity, lack of differentiation, often also appears in adult language, although we usually do not realize it. The connection between names (words) and concepts is very loose. For instance, the term "feet" is used not only in connection with animals, but also in connection with furniture ("feet" of tables, chairs etc.) and hills, or mountains. The meaning, in each of these cases, is conveyed not simply by the term iself, but also by the context in which it is used, i.e. the "model" of the general situation we have in mind.

But why is the same word used in connection with so diverse objects? The reason is, obviously, that all these objects have some common preconceptual structures. The feet of an animal, as well as those of a chair, are the parts on which the main body rests and with the help of which it rests on the ground (for the first meaning, [the parts which support the main body], sometimes the word "legs" is more accurately used). Similarly, the feet (or foot) of a hill or mountain is the place where its surface touches the plane.

It is also interesting to note here, that the image which underlies this naming is totally incorrect and subjective, rather than objective. Mountains are not objects (with a more or less flat bottom) that have been somehow placed on the plane. Yet the schema used for this description is the one which is more common, the one we usually use when we act on an object altering, its position, i.e. the place on which it rests.

In general, multiple meanings of a word seem to be an indication of common preconceptual structure of the concepts designated by it. We could also speak of common features among the "models" (in Johnson-Laird's sense), which underlie these concepts, but this description does not make very clear the existence of common encoding structures. Such terminology might give rise to the idea that models of the two concepts are somehow compared, until we find common features. However, the opposite is rather the case.

The same low level memory structure is used for the perception of both objects. That is why the similarity is often very hazy, as in the case of the word "eye", used in connection both to a man and a cyclone. A human eye and the eye of a cyclone are not merely more or less round objects but more importantly, although it is not apparent unless we take time to consider it, they are focal points of interest. Very early, even before we learn to speak, we learn that by "catching their eyes", as we say, we may hold the attention of other people focused on us (see [D.Lewis, 1980]). The eye is thus a central point of our interest. Similarly, the eye of a cyclone is its central point, the region around which the whole phenomenon develops. The human eye is also able to express powerful feelings, while the eye of a cyclone seems to be the center of action of natural powers.

The following observation is characteristic of how much the eye is a focal point of our interest. In a recent study of portrait paintings of 265 famous painters, who lived from the renaissance up to the present time, it was discovered that in almost all cases the painter had placed one eye at the vertical middle line, i.e. at the center of the painting [DER SPIEGEL, 20/1998, p. 181].

6.5 The Role of Logical Categories

A final remark is required on the role of logical concepts, i.e. of logical categories. If the storage system in LTM is based on empirical associations, then what is the reason for the existence of logical categories? Why should they be created?

Obviously, one of the basic reasons is the necessity to communicate and cooperate with other people. Unless we undergo a process of "standardization" of our concepts, reducing them to some more or less common usage, there is no way of communicating with other people. Thus, logical classification and conceptualization serves the formation of language, our main communication system.

In order to be able to describe a certain situation to other people, we need not only attach "names", i.e. words, to various nodes of the empirical association network, but we must also adjust the meaning of these "names" to a common understanding, avoiding ambiguity as far as possible.

By necessity,the standardization of the meaning of words involves categorization into nonambiguous categories, i.e. logical categories.

6.6 Empirical Associations versus Logical Categories

What are the arguments that support the above kind of modeling of the LTM? First of all, it is not faced with the dilemma, which of two related concepts is formed first in the LTM network; e.g. the concept "tree" or the concept "leaf".

This dilemma is not insignificant for any model characterizing a concept by the use of logical relations, since it cannot be resolved as here, by allowing the existence of initially incomplete "definitions". This would make the whole system inoperative, since in a logically structured system, action descriptions do not constitute the basis of encoding, but have to be inferred on the basis of the logical structural descriptions.

In other words, we should have "evolving procedural semantics", if the system is expected to operate before it has reached a highly developed form. Categorical networks are descriptions of abstract structure and must be complete in order to be useful. They are like computer programs, which are uninterpretable by the computer, if even a comma is missing.

For example, an incomplete concept such as "bird", which lacks the categorical feature, "has wings", is useless because we cannot deduce that a bird is able to fly. On the other hand, if "bird" is encoded by perceptual schemata, e.g. the perceptual schema which allows us to follow mentally the movements of flying objects, we still have a reasonable understanding of what a bird is, even if we have no precise knowledge of its anatomy.

If the encoding itself refers to interaction, lack of some features may result in incomplete control of the situation, but not total inoperativness.

Can we use words without knowing exactly what they mean? Yes! We do not need to have an exact definition of a concept in our mind in order to use it. For instance, everybody speaks about love or justice but very few, if any, would know exactly what these terms mean (see also the experimental results of Keil [1979] reported in **Chapter 4**).

Thus, a modeling of the above kind seems to fit fairly well what we know about the first stages of mental development. Additionally, without providing any specific clues on the mechanism of mental imagery, this model does not use an encoding totally incompatible with an imaginal one. However, it is not only imaginal or generally sensory, but rather sensori-motor (kinesthetic).

We should not forget that, as it is experimentally established, we do not acquire aptitudes by merely looking at other people's performance. We have to actively try to imitate them.

However, the idea that meaning is intimately connected with words is widely spread. So let us discuss more closely how meaning is encoded. Let us consider, for example, the concept "doctor (of medicine)", which seems fairly easy to define. Do we really encode the concept "doctor" by associating this word with the logical categories "professional", "has a university degree in medicine" and so on?

This is not at all the way in which a mother explains to a child what a doctor is. What she is likely to say is, e.g. "When we are ill, we go to the doctor and he gives us medicine to make us well". She is telling the child how he can interact with a doctor and not what a doctor is in abstract terms. In fact, this is the kind of answer that most people give, when asked to explain the meaning of some word. They do not define the underlying concept by linking it to certain abstract categories, but try to describe what kind of activities are connected with this word.

When asked, "what is an egg?", they are more likely to say that it is edible (most probably referring to a chicken egg), than that it is initially an unicellular organism developing to multicellular etc., although the latter is a defining feature of all eggs, even those laid by fish, while the former is superficial and unimportant for a strict, "objective" definition.

Let us consider another example. In **Chapter 4** we have discussed the procedure ONRIGHT, used by Johnson-Laird in order to build up a mental model referring to the relative position of various objects. This computational procedure is based on using a coordinate system. Every object having a greater x-coordinate than another one, is located on the right of this second object. But is that really the way we learn to distinguish right from left?

At least in my case, this is not so. At school I was taught to use my right hand for writing, which was very difficult for me at the beginning, because by nature I am left-handed. So, by necessity, I became partially ambidexterous. For instance, I can also write passably with my left hand, although I was never allowed to do so when I was young. This however, led to a confusion in my mind. I still find it difficult to distinguish right from left. I may instruct somebody to turn to the right, although I actually mean that he should turn to the left. So I truly needed some marking system in order to distinguish right from left, while this may come natural to right-handed people. The marking system I used when I was young was to try and make the sign of the cross, since I was taught very early to use my right hand to do so. Later, it was also possible to consider which hand I use for writing, since I got used to using my right hand for writing. But I still get mixed up sometimes.

What is the significance of these observations? Obviously, there is no mental coordinate system. Instead, first I identify my right hand by making, e.g. the mental preparations for writing, and then I consider as lying on my right everything that seems to be closer to my right hand. This whole procedure utilizes a sensori-motor schema, used for performing some action ,and not some abstract computational procedure of any kind in order to define ONRIGHT.

6.7 On the Visual-Kinetic Encoding of Abstract Mathematical Concepts

There are also many indications that the way we use even abstract, e.g. mathematical, concepts is based more on sensori-motor organization schemata than on strict definitions.

As we have already pointed out in the chapter on Cognitive Psychology, for instance the first, primitive notion of continuity is based on continuous movement, e.g. on the way in which we draw a continuous line on paper. In fact, Leonhard Euler defined a continuous function as a "curve scribed by freely leading the hand" [Stewart, 1995, p.237].

But such a line is by necessity not merely continuous, but also piecewise continuously differentiable (it cannot change direction appreciably from one point to the next, except for isolated points). This makes it very difficult to understand that there are continuous and nowhere differentiable lines, i.e. having no tangent anywhere.

Only after the study of examples of continuous, but nowhere differentiable curves, does a student believe that continuity does not imply continuous differentiability almost everywhere. Such examples are given, for instance, by Vilenkin [1968, pp. 100-104] (see also [Ian Stewart, 1996, pp. 238-241]) but are not easy to visualize. They are conceived as results of limiting processes that introduce ever more breaks in an initially smooth line.

In fact, the concept of continuity usually used during mathematical work is still the imprecise, visual-kinetic one. The strict definition of continuity is applied only in order to precisely formulate a proof already discovered by using the visual-kinetic concept or when one suspects that he might be led to the wrong results otherwise.

Another abstract mathematical concept which has a procedural "definition" parallel to its abstract, linguistic one, is the concept "integral". During computations, it is usually understood procedurally rather than on the basis of its formal definition as a limit. Since the definition is used only for computing certain basic integrals and all others are reduced to such basic integrals by appropriate transformations, the mathematician usually recalls how he can "handle" (transform) the integral he encounters, rather than the abstract definition.

This visual-kinetic, i.e. sensori-motor, encoding actually happens with many concepts in mathematics. The first, abstract definitions may have a linguistic form based on strict logical analysis. But soon the mathematician learns to ignore the strict definition, or the actual rules of symbolical processing, and replaces them with more handy visual-kinetic schemata.

Thus, algebraic transformations are initially based on precise application of certain proof techniques. Soon, however, the proof techniques are simply replaced by visual-kinetic rules saying how we can transform a certain formula, without any formal justification.

The equation: $x + b = a$, for instance, is proved to be equivalent to $x = a - b$, by subtracting from each side the same number b. Soon, however, the pupil learns that he can always move an added quantity from one side of the equation to the other by changing its sign.

Similarly, he learns that he can move a factor from one side of an equation to the denominator of the other side. Thus, $a*x = b$ is transformed to $x = b/a$, usually without mentioning that these equations are equivalent, because both sides of an equation can always be divided by the same nonzero number.

In order to perform the division $Q = x^2/(x+b)$ I find it easier, instead of performing the rules of division, to extend the expression appropriately, e.g. by writing:

$$Q = (x^2 - b^2 + b^2)/(x+b) = x - b + b^2/(x+b),$$

i.e. to perform visual transformations of the expression, since I know the identity:

$$x^2 - b^2 = (x - b)(x + b)$$

This formula may also serve to indicate how imaginal transformations, i.e. visual-kinetic schemata, are used when doing mathematics. The student may initially keep this identity in mind in a formal form like:

$$A^2 - B^2 = (A+B)(A-B)$$

However, after he realizes that he can substitute any possible expressions for A and B in it and acquires some experience with such substitutions, he rather keeps in mind a sensori-motor schema that we might express as:

"$(\ldots)^2 - (\cdots)^2$ is equivalent to $\{(\ldots)+(\cdots)\}\{(\ldots)-(\cdots)\}$"

where within the parentheses () can be anything.

This is psychologically not the same as saying that we can replace anything for A and B in the initial formula. In order to simplify a more composite expression, like $(a+b)^2 - (a-b)^2$, by writing:

$$(a+b)^2 - (a-b)^2 = \{(a+b) + (a-b)\}\{(a+b) - (a-b)\} = 2a*2b = 4ab$$

we must first recognize that here we have a difference of two squares. How does a mathematician do so? Certainly, he does not first replace a+b by A and a-b by B to reach the general form $A^2 - B^2$, in order to recognize the general structure, the general form of this expression. This is done instantly, without a thought.

It rather seems that he temporarily considers every pair of parentheses (), together with the expressions they contain, <u>as unique symbols</u>. These symbols are subsequently recognized as composite expressions which are added and subtracted.

6.8 The relation of this Modeling to other LTM models

In the above description we have used the term "sketchy model" or "modeling schema of a semantic network" instead of the term "model", because the proposed mechanism does not have an easily replicable structure, in contrast to most connectionist models, which have categorical connections.

We might ask, for instance, whether preconcepts can be given names, so that we can place them at the nodes of a connectionist network in the usual manner. This, however, does not seem to be appropriate for their characterization, since they are preliminary structures of many different concepts, thus having a (linguistically) fuzzy meaning.

Besides, any names we might give to preconcepts would have to be temporary, being eliminated or transferred to higher developmental stages, as it really happens during the evolution of the conceptual system. We continuously specialize or generalize the meaning of words. For instance, at first we may call everything made of flour "bread". Later we maintain this name only for certain kinds of bakery and learn that some others are cakes, pies, etc. Similarly, initially we consider only paper notes and coins as "money". Later we learn that there is also "plastic money" (credit cards), "electronic money" etc.

In the first case, the meaning of "bread" is specialized, while in the second case, the meaning of "money" is generalized. What is more, these processes happen by convention and not as results of some kind of intricate logical processing. The generalization is simply based on attaching the handy name "money" to the general meaning [means for easy transfer of value], which is in our mind in all the above cases. The specialization is necessary when we have new kinds of bakery that deviate too much from the usual bread.

Because of their intrinsic fuzziness, we also cannot replace preconcepts by frames containing some linguistic description of their meaning. This view also differs from that suggested by "natural category" studies: according to these studies, objects are not perceived as sets of distinct (logical) attributes, but as wholes.

Therefore, based on such observations, certain theorists group the objects not in categories with sharp boundaries, but in "fuzzy categories", clusters of objects with a dense center of representative members of the group and fuzzy boundaries.

Such "fuzzy categories" or "fuzzy sets" contain objects which are only occasionally grouped within this category, while in other instances they are perceived as members of other categories. They are said to have "varying typicality" as representatives of a certain category[19].

An experiment illustrating these points was devised by the linguist William Labov. He showed his subjects pictures of receptacles of varying form and proportions, all equipped with a handle, asking them to imagine the object in each picture in a person's hand and to name it (see figure 6.1 [Morton Hunt, pp. 153-154]). When the receptacle was about as wide as deep, it was called a "cup". If it was more deep than wide, it was more likely to be called a "vase", while if it was more wide than deep, it was usually called a "bowl".

But these classifications varied also according to the use Labov attributed to each receptacle. If he said that it was filled with mashed potatoes, it was more likely to be called a "bowl" even if it was not so wide. If he said that it was filled with cut flowers, it was called a "vase", even if it was not very deep. (see [M.Hunt, pp. 153-154]).

[19] Mathematically, this is expressed by attaching to each object a certain probability, or rather indicator of certainty (between 0% and 100%) that it belongs to this category or is considered as a member of it.

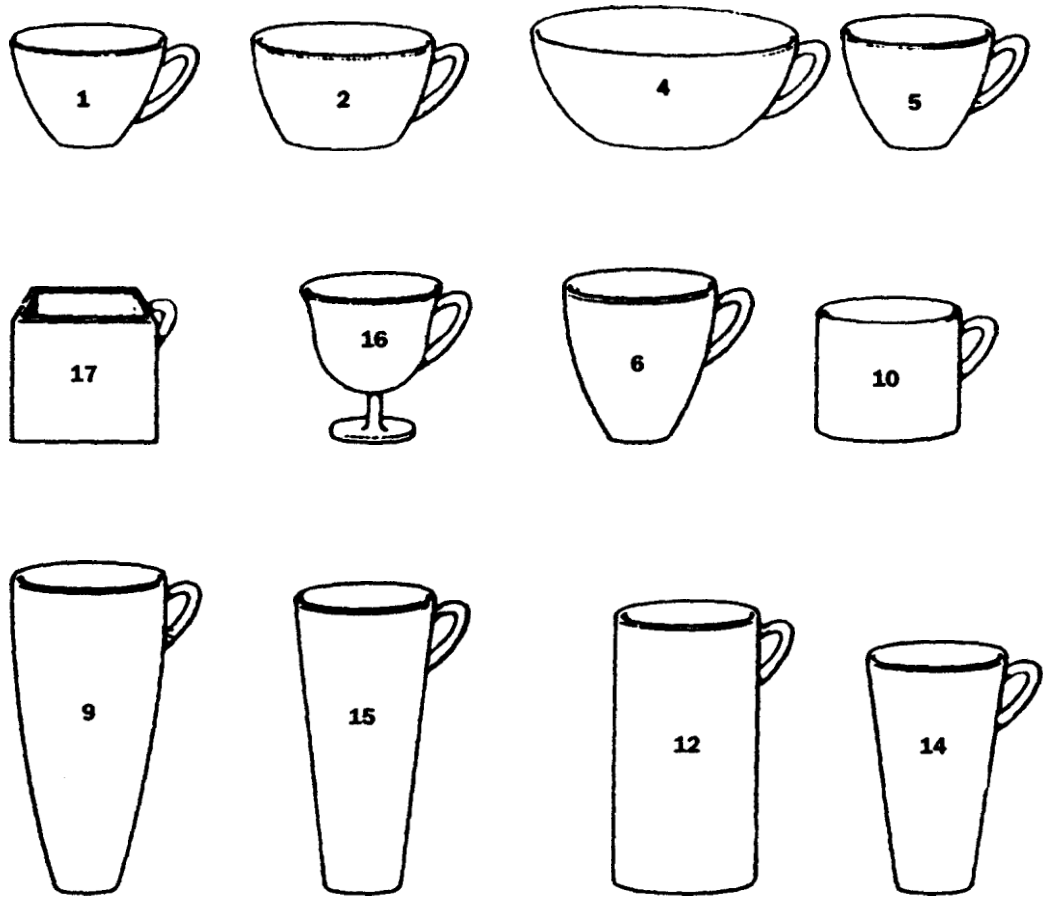


Fig. 6.1: Which container is a cup and which is a vase?

These findings seem to suggest that each natural category is made up of a grouping of certain "best" or more typical representatives. However, when we consider these "natural categories" as concepts, it is not clear how these most typical representatives of each one are kept in the mind; obviously not as mere visual images.

The position taken here is that there is no clustering of more or less typical objects which defines a concept, i.e. a natural category, but a clustering of preconceptual substructures or features. There is nothing like "the most typical cup" in the mind, but varying networks of empirically developed preconcepts, i.e. partial preconceptual substructures.

If an object has more features which advice its use as a cup, it will be called "cup". If it has more features appropriate for its use as a vase, it will be called "a vase". If a use is suggested, which is typical for a vase, the object may be called a vase, even if it does not seem very appropriate for such a use.

It may be thought that fuzzy categories, fuzzy sets, can at least be good approximations for preconceptual content. However, such an "approximation" does not take us very far. How typical a receptacle is, e.g. a wallnut shell?

The preconceptual depth of concepts is important in order to allow us to grasp all the possibilities that are intrinsic in a concept. By restricting a "natural category", i.e. fuzzy category, to only certain objects and to a certain degree of typicality of each one, as representative of this category, we exclude infinite other possibilities. We completely eliminate the possibility of some other objects, which do not seem "at first sight" to have something common with the category, to be considered as belonging to it.

It must also be noted here, that the modeling discussed is not a two-stage model with a higher level of concepts building the knots of the network and a lower level of imaginal or combined kinetic-imaginal "definitions" building the concepts.

The sensori-motor organisational structures which define a concept are, to a great extent, also parts of other concepts. Thus, activation of a concept also spreads to its preconcepts and may activate various other concepts. This is what makes human thought creative.

But what is the use of preconcepts, if we cannot define them precisely? Preconcepts are identifiable, since they consist the developmental history of a concept, but they have no simple form. They might be considered procedures, in the sense of procedural semantics, but only the most elementary preconcepts should have a simple procedural description.

Simplicistic descriptions of the conceptual system may be more attractive, but it now becomes more and more obvious that they cannot account for highly complex mental phenomena like creativity and intuition.

Besides, we are led to the present view of the conceptual system not by certain artificial and arbitrary assumptions, but simply by considering a concept's developmental history, not as a thing of the past, but as its permanent constituent. This is an assumption which is not so far-fetched.

6.9 The Relation of Empirical Associative Networks to Mental Models

Comparing this kind of modeling to the theory of mental models, we should note that mental models cannot represent the meaning of simpler and more fundamental concepts. An empty nutshell may eventually be interpreted as a receptacle and used as such. How could such a variation of meaning, i.e. reinterpretation, be explained by mental models?

Note that such a process of reinterpretation is usually automatic, impulsive, like the reflexes and not a product of remodeling, of reforming a certain model. It is far more plausible that a procedural mechanism, [receptacle], is part of the preconceptual interactive mechanism representing "nutshell". The same kinesthetic mechanism is, of course, part of any other concept that could be interpreted as [receptacle].

The separate processing mechanism for solving logical problems, which is proposed by the theory of mental models, seems to actually exist in the mental system. But there are also deeper processing levels, which are very important for creative reinterpretation and understanding.

Difficult mathematical problems, for instance, have no obvious models from which one could "read" the answer! The apropriate mathematical model is not a mere synthesis of the given premises. Very often, the necessary "model" for solving the problem has no direct connection to the premises, as we saw in the cases discussed in **Chapter 4** and **Appendix 4.3**.

Indeed, in order to create or understand a mathematical, physical or even interpretative theory of historical processes, we always need to build up an appropriate model, which we gradually enrich as we acquire a deeper understanding of this theory. But the process of model-formation itself is important in such cases and not mainly how we "observe" or "examine" the model implied by the premises.

This model formation can only be based on deeper semantic evaluation of the existing cues. This happens by scanning their preconceptual levels, guided by motivation. The motives select both the appropriate cues and the direction of investigation.

6.10 The Significance of Motivation

All this discussion about LTM organization should not make us overlook the very important factor of the development of the mental system, which is often neglected in such studies: the motivation.

Motivation begins at the very elementary, inborn drives (control hierarchies of our responses) and gets ever more refined as the conceptual system develops. Motives are just as strongly responsible for any processing in the brain, as the outward stimuli.

Most of the time, i.e. unless the sensory input is very strong, it is our motivation that selects which of the incoming stimuli will be more closely attended to. Motivation also decides in which direction the further processing of incoming information will go.

If we imagine the memory as some kind of network, then motivation is responsible for keeping the activation in a certain part of the network, when the outward stimulus has vanished, as well as for spreading the activation even further in this network in a certain direction, rather than others.

An indication for this is, for instance, the already mentioned observation that children can perform linguistically much better when they have a strong motivation than when they are not motivated enough ([M.Donaldson, 1983, p.74]).

A model of LTM that incorporates motivation should also include drive mechanisms that provide initial interest, initial activation of certain parts of the network, as well as other mechanisms maintaining activation until curiosity or other, e.g. material biological needs have been satisfied.

However, the inborn drives of an organism, the instincts[20], are not simply parts of an information processing mechanism. The "encoding" of many basic concepts must include a mechanism for immediate activation of drives. The drives consist a part of the concept and not only mechanisms for data evaluation, because in many instances the reaction of the organism to outward stimuli is of vital importance.

If we meet a free lion on the street, what will we do? Will we first make some syllogism like, "this is a dangerous animal, it could hurt me or kill me" or will we start running without any thought? Thought is needed here rather in order to moderate the first, instinctive reaction to flee, so that we do not immediately draw the attention of the lion.

If fear is not part of the concept "snake" for some people, why do they start sweating from fear, as it is observed by Goleman (see **Chapter 3**), even if pictures of snakes are presented to them for so short time intervals (fractions of a second) that they don't realize what they saw?

This experiment also shows how closely concepts are related to sensori-motor mechanisms. Obviously, the activation of at least a part of the conceptual network happens here instantly, without any search for and evaluation of logical categorical properties.

Because of its importance, the role of motivation will be discussed later, in a separate chapter.

6.11 Possibilities for the Study of Preconceptual Structure by Means of Experiments or Observations

How and to what extent can a mechanism, like the one considered here, be revealed by appropriate experimentation? The method that suggests itself immediately is to observe very young children.

However, one must know what to look for. Empirical categories do not have unique hierarchical structure like logical ones. The encoding of a concept in two different persons may rely on partially differing substructures (preconcepts), although a certain degree of coincidence can be expected. For instance, the concept "hollow" may be encoded either by experiencing hollow natural objects (a calyx of a flower, a cavity in a rock or in a tree etc.) or by experiencing artificial hollow objects (cups, vases etc.). In any case, what is important is the interactive experience with hollow objects and not the particular selection of objects.

If preconceptual content is to be revealed, we have to start the observation, not at the linguistic level, but already when the child builds up his first sensori-motor experiences with his body. We must start the observations when he first explores hollow objects and not when he first pronounces the word

[20] The greek word "ἐνστικτον" means "punctured in", "engraved" so to speak.

"hollow". However, the study of "overgeneralizations" at the linguistic level could also reveal some preconceptual structures.

The study of how children born blind acquire their concepts might simplify the study of the preconceptual structure of these concepts. In this case, all experience is collected by touching and handling the objects, which is far more precisely observable than a visual scanning, a visual examination of them.

However, there are also some other fields of study which may help the understanding of how concepts evolve, but not in a direct way, because they do not exactly study concept formation. Such a field of observation is how the meaning of foreign words is learned by experiencing their usage. The meaning is at first hazy and general, containing rather the "feeling" of the word than its actual meaning and only gradually becoming narrowed down to a precise form.

Here there is no exact parallel to the acquirement of new concepts by young children, since learning a foreign language requires in most cases (but not always) the identification of already known concepts with a new name, a new word.

Similarly, etymological studies reveal indirect connections between various concepts. In ancient Greek, for instance, the concepts "to have" and "to hold" are both expressed by the same word: "ἐχω". This implies that they may be intrinsically connected. Also, in German, the verbs "to hold" and "to stop" are very related. The first is translated as "halten" and the second as "an-halten" (hold-by/on/at).

Since creating artificial intelligence is a basic motivation in many studies about the functioning of the mind, we might ask here what the implications of this modeling are, for such a goal. Such a task is indeed very formidable and only a very vaguely general answer can be given at present, so we will defer it for **Appendix 6.1**. Here, let us discuss another important question. Does this model account for the basic mental faculty of intuition?

6.12 A Mechanism for Intuition

How does procedural semantics account for intuition? The process of concept evolution outlined here provides a very different answer to the question of how intuition works, than the usual recourse to unconscious heuristic rules.

There is no necessity to assume the existence of unconscious heuristics. A far simpler mechanism is possibly operating here: intuition may be a tracking back, i.e. in depth, of some common preconceptual structure of the problem's factors, which we try to interconnect. It can also be the localization, the tracing back of factors that are important for the solution of a problem, through their preconceptual content.

Deeper, earlier layers of the conceptual network are considered here as increasingly non-specific. Thus, they embrace various aspects of superficially very different logical categorical concepts, which are usually represented by words. This makes it possible to discover common features without logical reasoning or logical heuristic search.

The basic difficulty encountered here is not finding deduction steps, but rather becoming aware of some common preconceptual feature and "explicating" it, making it explicit. We have not only to locate it, but also to model it in terms of higher developed concepts, in order to give it linguistic expression. This view also answers the question of why so many creative mental processes are unconscious, without being traumatic.

The very early memory structures do not only lack a direct memory address, since they have no names, but it is also increasingly difficult and soon impossible to describe them, because of their increasing haziness (non-specificity) as we move to earlier levels.

Lower levels of the memory network are thus not merely usually unconscious for some reason, but in principle unconsciousable, not able to be made conscious.

This is also a possible explanation of the phenomenon that the part of the mind called "the Unconscious" uses symbols and allegories, not only in the case where it refers to traumatic experiences, when they may be used as disguises, but also when it refers to creative processes, like creative dreams.

Such was, for instance, Kekule's dream of snakes biting their tails, which led him to the discovery of the ring structure of the benzene molecule (see **Chapter 4.12**).

It seems that the most direct way to point to some preconceptual content is "theatrically", by "enacting" it, i.e. by using some pantomime that stresses that particular point.

However, the process of intuition is not merely an automatic perception of existing preconceptual structure. The existence of common preconceptual structure does not mean that we are instantly able to "perceive" it. It is hidden in a vast network of other preconceptual connections.

First of all, we have to focus our attention on it, activating this part of the network enough to raise it to "semiconsciousness". Then we have to explicate this common preconceptual structure by describing it in terms of other concepts, adjusting it to logical categories (categories built by logical classification) and logical relations. This may even lead to the formation of new concepts, comprising some features that for some reason have to be separated from already existing concepts and considered as independent entities.

Thus, thinking seems to be an interplay between two processes:

a. Tracing back of empirical associations.

b. Logical processing, logical explication (in which mental models may also play an important role).

How can we be certain that this is the basic mechanism behind the phenomenon of intuition? There can be no certainty in this context, since intuition is not replicable to command and cannot be directly experimentally investigated.

The best possibility, which seems to remain open for a practical study of intuition, is to consider personal accounts of such instances, although self-observation is not always a very reliable source of information because it can be influenced by wishful thinking.

The lack of any other source of direct information about this mental phenomenon has led many scientists interested in studying it, like Hadamard and Poincaré (both famous mathematicians), to give personal accounts of such instances of inspiration in order to describe its basic features.

Following their lead, I will therefore describe here two personal experiences of gaining new mathematical insight and try to interpret them in terms of the model proposed here.

6.13 Two Personal Experiences

The following descriptions are non-technical and the mathematical terms, occasionally appearing in them, are just used to make certain magnitudes identifiable. They are not important for understanding the phenomena. However, in order to avoid overburdening not mathematically versed readers, I will give the full description in **Appendix 6.2** and **Appendix 6.3** and present here only the main conclusions of these discussions.

In the first case under consideration (**Appendix 6.2**), the concept "error" or "accuracy", used in one mathematical field, led to the modification of a very different error-concept (accuracy-concept), used in another mathematical field.

This happened seemingly through a common preconcept, a very primitive notion, which lies behind any developed and specialized concept of "accuracy". Certain aspects of the first error-concept were thus transferred to the second one, leading to a very different appreciation of it. This allowed the creation of a new error-estimation technique for the second mathematical field.

The important steps in this case were not the ensuing complex mathematical transformations, but the subliminal, the subconscious modification of the concept itself. This concept modification led, in some sense, to a modification of the motivation, i.e. of the guiding target. Instead of seeking directly a satisfactory estimation of the actual error, I tried out a gradual bracketing of it in intervals, which were rapidly growing narrower. This was something that could be achieved much easier.

A second personal experience of discovering a mathematical relation, discussed in **Appendix 6.3**, involves more directly a visual modeling of the premises, but it also uses a kinetic modification of this model [Kioustelidis, 1980].

The whole process was not merely a matter of putting together all relevant information into a model and then scrutinizing it to find the answer, but rather a matter of composing and trying out various models until a model appeared, which yielded the answer. This answer was obtained by means of a mental sensori-motor evaluation of the situation represented by the model. The "sketching pad", on which I performed my modifications, was in my mind.

The final model that I was studying transformed the initial problem into the simple geometric question of how I could reduce a certain area. Considering this visual representation in my mind, I suddenly realized that to do so, I only needed to extend a certain curve. Its extension would "cut away" some of the area. This realization occurred by the change of the initially static image into a kinetic one. Only when I tried to shift one of the two endpoints of the curve did I see that this would reduce the area. In this case, a sensori-motor action schema was applied on the visual representation of the problem. The mental procedure that was used seemed like those we apply in order to prepare the action of peeling an orange or cutting a loaf of bread into pieces.

Another important factor involved in the process was the motivation. I was fairly confident that a simple, directly observable answer to the problem did exist, so that I persistently tried to find it. This conviction was due to the fact that one year earlier I had derived by calculations a very simple answer to a very specific instance of this problem.

This is a case that reveals how important can be the sensori-motor schemata that are attached to some abstract concepts. It also shows how important is the role of motivation in making some discovery.

6.14 The Deeper Role of Motivation in Mathematics

However, the role of motivation can be even more fundamental than the one observed here. Many discoveries in Mathematics are in fact mainly "formalizations", i.e. the expression with formulae of the answer implied, required by the motives of the investigation. They give a formal expression to a recurrent motive, a wish to have an answer to a certain type of problem by all means.

Important new concepts are often abstract mental constructs that give a formal expression to certain motives, certain wishes of the investigating mind. Thus, a derivative was initially an analytic formalization of the geometric concept of building a "tangent" to a given curve. Also, complex numbers are an artificial construction that allows the solvability of algebraic equations like $x^2+4=0$ and $x^4+16=0$, while negative numbers are an artificial construction that allows the solution of equations like $x+1=0$ or $x+5=2$. That is, these constructions express formally "solutions", which up to this moment did not exist, did not have a meaning. Similarly, "Spline Approximations" are a means of constructing good approximations to a given curve, by piecing together pieces of simple curves, which can thus be adjusted as well as we wish to the given one.

In all these cases, a great number of difficult theorems must be proved in order to make these mathematical instruments safely applicable. But the most important part in all these discoveries was the initial motivation, which created a new concept, a new tool, and not the subsequent formal investigations.

Progress in a scientific field may stall for decades or centuries, until an appropriate new concept appears; one that allows new outlooks. Then, the proof of the pertinent results and formulas is completed in a few years.

6.15 The Main Stages of a Mathematical Investigation

Concluding, we might say that mathematical work seems to be based on developing a model of the field we wish to study, but this model is not simply a synthesis of the premises of the problem. It has to be created by trying out various alternatives and this process is often guided by the preconceptual content of the mathematical concepts, the mathematical magnitudes that appear in the problem. These magnitudes may not be obvious, but often enter indirectly through the goal we wish to achieve.

The model is usually visual, but not always geometric. It can also consist of symbols. Sometimes this model may make the solution obvious, but in other cases it is merely a part, a byproduct of the pantomime that represents the solution.

In any case, successful work in a field requires the development of a whole "landscape", a successful visualization of the field incorporating the basic magnitudes and how they are interrelated. I.e., it requires making obvious the relations between these magnitudes, thus allowing one to guess new relations before he attempts to prove them. No matter how well we know the relevant theorems, we do not have a true "grasp" of this field until we have a (geometric or symbolic) model that makes all these relations obvious, i.e. easily explainable.

This search for plausibility is not idle, but very important for further work in this field. For instance, very often in Mathematics, we generalize to spaces of infinite dimensions concepts like "orthogonality", "convexity", "tangent plane", "contact point" etc., which have a visual meaning only in spaces of two or three dimensions. This happens because the correspondence of the generalized concepts to those of the two-dimensional or three-dimensional space helps us to grasp visually the generalized concept. Similarly, many of the theorems that refer to square matrices can be derived very simply, if we consider the matrix as a vector of vectors. Without this visualization, the derivation, e.g. of the formula for the inverse matrix, is very complicated.

The processes that participate in mathematical research thus seem to be the following:

(1) Logical synthesis of a model of sorts and study of the properties of the model.

(2) Addition of factual information from our memory and modification of the model by means of an inward dialogue that tries to bring out the most promising features of the model.

(3) Visual-kinetic processing of the formulae substantiating (constituting or accompanying) the model, on the basis of known relations and some "feeling" of symmetry, of simplicity of transformations, that would render this processing more promising.

(4) Visual-kinetic processing of graphical representations of the situation that is being investigated[21].

(5) Pre-conceptual search in order to bring out some specific features of the concepts involved or in order to expand the concepts involved, so that they acquire certain new features expressing, e.g. our intentions!

All these processes are used alternatively, in an arbitrary order determined by our inward dialogue. The motivation that determines in which direction this investigation goes is very important, but also maintains our interest for this general line of research.

A Short Review

In this chapter we have combined what has been said in the previous chapters, in order to create a unified view of the mental system.

Considering the Long Term Memory system as a network structure, we have first considered how such a semantic network can be built up. Our main guide was the view that the conceptual system must not only be functional in its final form, but also during its creation.

This led us to the rejection of the idea that, for its creation, some initial, elementary concepts are connected, no matter whether they are determined lexicographically or by sensori-motor procedures. In both cases, the conceptual system would not be functional at the first stages of its creation.

Thus, we ended up with the hypothesis that the chronologically first connections (nodes) of the memory network are simply associative connections based on interactive experiences with the environment. Already these first, very unspecialized sensori-motor structures (i.e., composite sensory and motor structures), were called "preconcepts" because they consist a common substratum of all concepts. These connections become ever more specialized by the development of further empirical associations, which differentiate the initial memory structures.

[21] A very good example of visual-kinetic thinking is Poincaré's discovery of chaotic phenomena in planetary motion. The immense complexity of the orbits around a "homoclinic" point, which was discovered by Poincaré about the end of the 19th century, is very nicely described by a simple visual-kinetic presentation in Ivar Ekeland's book "Mathematics and the Unexpected", Univ. of Chicago press, 1988.

The whole process is some kind of evolution and not a complete restructuring at each stage. Previous developmental stages of a concept are not replaced by later ones, but remain in the memory as permanent constituents of the concept.

However, the associative connections are considered here as empirical associations (based on experiences) and not as logical associations (connections of logical categories), as in most network models. It is not logical components (categorical components) that are being connected but various sensori-motor schemata, which tell us how we can interact with an object, i.e. how we can observe it, touch it, handle it or eventually, follow its movements.

This leads to the question: how do concepts, as we usually understand them, enter into this system? Where are the concepts in this system? This happens when we start developing words and linguistic expression. The words are our marking system for concepts, which makes them discernible, while language is the system for expressing more composite meanings by means of structured sets of words.

The main features of the proposed mechanism for LTM can thus be outlined in the following way:

1. For the encoding of an object are not used any outward, supposedly "objective" characteristics of it, but the way we can handle it (touch it, feel it, move it etc.), look at it, or react to its possible changes.

2. If two objects can be partially handled or perceived in the same way, then this partial mechanism (structure) is their common "preconcept". The encoding does not start each time all over again but uses already developed structures as much as possible.

3. Words, i.e. names for things, actions etc. are simply tags placed on some nodes of this network, when it has already been developed to some extent. They are necessary for communicating and for thinking in logical categories but they are not necessary for the storage (encoding) of meanings in their initial form. Rather, they serve the construction of mental models and so the conception of more complex meanings.

The lack of concept names, i.e. words, at the first stages of the development of the conceptual system, seems to be also the reason why many contents of the memory are not easy to trace. They are simply not in a form that allows linguistic reference to them. They do not have a linguistic "address".

4. The "meaning" of preconcepts has increasing haziness, increasing non-specificity as we go to deeper (earlier) stages of concept formation. Thus, very different concepts may have some common preconcept or rather, preconceptual structure. An indication for this is the way in which young children initially "overgeneralize" the meaning of words.

Here it was noted that the above kind of modeling of the LTM does not face the problem of many concepts determining each other anymore and it is not obvious which comes first. There is no necessary serial order of concept formation, but all of them emerge gradually from the vast network of empirical associations. Such a network is functional, even if not very precisely, already in its first stages. The incomplete formation of empirical perceptual schemata does not hinder their use, since they are based on sensori-motor procedures. In contrast to this, logical categorical networks are descriptions of abstract structure and must be complete in order to be useful. They are like the computer programs, which are uninterpretable by the computer, if even a comma is missing.

To the question, what is the reason for the existence of logical categories in such a system, which is based on empirical and not logical associations, we gave the answer that logical categories are necessary for the communication between people. Unless we undergo a process of "standardizing" of our concepts, reducing them to some more or less common usage, there is no way of communicating with other people. The standardizing of the meaning of words involves by necessity categorization into nonambiguous categories, i.e. logical categories.

It was further observed that the introduction of "fuzzy categories", which is based on the observation that objects belonging to the same category are not considered as equally typical representatives of it, is not enough in order to explain higher mental abilities. The importance of motives was also stressed. These are equally responsible for any processing in the brain as the outward stimuli. They choose which of the incoming stimuli will be more closely attended to and decide in which direction the further processing of incoming information will go. They start from the very elementary inborn drives (control hierarchies of our responses) and become ever more refined as the conceptual system develops.

Next, we considered the question: how and to what extent can the proposed mechanism be revealed by appropriate experimentation? Here it was proposed that a possible method is the observation of very young children. However, in order to discover preconceptual content, the observation must begin, not at the liguistic stage, but already when the child forms its first sensori-motor experiences with his body. Furthermore, the observation of children born blind could make the study of preconceptual structures easier, since all experience is collected in this case only by touching, feeling and handling the objects.

Finally, we came to the central question of how the intuition operates, which was defined in **Chapter 2** as the function that allows us to make logical jumps to interconnect concepts that do not seem logically connected.

Its mechanism is possibly the tracing back, i.e. in depth, of some common preconceptual structure of the factors involved in the problem we study. I.e., it is a localization of factors, important for the solution of the problem, through their common preconceptual content.

The basic difficulty encountered here is not finding deduction steps, but rather becoming aware of some common preconceptual feature, i.e. "explicating" it, making it explicit. Not only do we have to locate it, but also to model it in terms of higher developed concepts in order to give it linguistic expression. The only way available to us, to initially grasp this preconceptual affinity, seems to be the use of pantomime and allegory. This may also explain why so many creative mental processes are unconscious, while they are not traumatic.

Thus, thinking seems to be an interplay between two processes:

a. Tracing back of empirical associations, which usually happens unconsciously and is expressed by symbols, pantomimes and allegories.

b. Logical processing, logical explication, in which mental models may also play an important role.

In order to support what has been said, two personal experiences of grasping new ideas and discovering new mathematical relations are described and an attempt is made to explain them on the basis of the above mechanism.

The chapter ends with an analysis of the role played by the motivation in mathematical research. This analysis shows that many mathematical concepts are only a formal expression of the dominant motive of the research. This happens when no answer is possible by using previous concepts.

BIBLIOGRAPHY

Donaldson, Margaret: Children's Minds, Fontana/Collins, Glasgow, 1983

Hadamard, Jasques: The Psychology of Invention in the Mathematical Field, Dover Publ., New York

Hawkins, D.M: On the choise of segments in piecewise approximation, J. Inst. Math. Applic., 9 (1972), pp.250-256.

Hunt, Morton: The Universe Within, Corgi Books, 1984

Kioustelidis, John B.: Algorithmic Error Estimation for Approximate Solutions of Nonlinear Systems of Equations, Computing 19 (1978) pp.313-320.

Kioustelidis, John B.: Optimal Segmented Approximations, Computing 24 (1980) pp.1-8.

Lewis David: The Secret Language of your Child, Pan Books, London, 1980

Miller, Arthur: Imagery in Scientific Thought, MIT Press, Cambridge Mass., 1986.

Moscowitz, Breyne Arlene: The Acquisition of Language, Scientific American, Nov.1978, pp.82-96.

Poincare, Henri: 'Mathematical Creation' pp.2041-2050 in Newman, James (ed.): The World of Mathematics, Simon & Schuster, New York, 1966.

Powell, M.J.D.: On Best L_2 Spline Approximation, pp.317-339 in ISMN 9, Birkhauser, Basel, 1968.

Schumaker, Larry: L_2-Approximations by Splines with Free Knots, pp.157-182 in: G.Meinardus (ed.): Approximation in Theorie und Praxis, Bibliographisches Institut Mannheim, 1979.

Stewart, Ian: Concepts of Modern Mathematics, Dover Publ., New York, 1995.

Stewart, Ian: From Here to Infinity, Oxford U.P., Oxford 1996.

Vilenkin, N.Ya.: Stories about Sets, Academic Press, New York and London, 1968.

Chapter 7
COMMUNICATION AND LANGUAGE.

There is a widespread opinion that no thinking can exist without language, because language is supposedly the only possible carrier of thoughts and meaning. However, the model proposed here introduces the idea that a great part of the LTM network has no relation to language. It corresponds neither to lexical items, nor to propositions which have a language-related form. It is therefore important to find out whether meaning can exist without language or not.

7.1 Body Language

Can we communicate without words or word substitutes like the signs of the deaf-mute sign language? This question seems strange at first, until we remember that we also use gestures, e.g. in order to emphasize what we say. A more appropriate question is therefore: "How much can we say without words, or word substitutes?". Surprisingly, the answer to this question is: "Quite a lot".

For instance, behavior researchers (ethologists) have established, by observing carefully filmed encounters between people, that there is a rich "vocabulary" of various messages communicated not only by gestures, but also by fleeting facial expressions and slight changes of posture and tenseness of the body. These are called "micromovements", in contrast to the more obvious gestures and longer lasting changes of facial expression, which are called "macromovements".

An example of such a message is the unconscious "eyebrow flash", which people do when they meet a friend, even before they have the time to think of greeting him (see Eibl Eibesfeld, [1976, pp.179-183] or D.Morris, [1982, p.16]). This message is quite universal, used e.g. even by Amazon-Indians who had never met white men, as Eibl Eibesfeld has found out.

Micromovements are usually unconscious. They are unconsciously performed, unconsciously perceived and semantically evaluated, while many (yet not all) of them are universal.

As a matter of fact, even macromovements are often unconscious. When we speak, we seldom notice the gestures we do with our hands, although they are semantically significant. What is more important, micromovements and macromovements consist a continuous and coherent flow of non-verbal conversation (see D.Lewis, [1980, p.38]). This body language seems to be the main language of feelings. It is the basic way in which we express our emotions and establish emotional contact.

Young children develop this non-verbal communication system much earlier than the verbal one and use it in social encounters with great dexterity. By means of it, they can call attention to themselves (e.g. when babies smile to somebody) or express their inward reaction to what another child does. They can indicate rising anger, threaten or show defiance as variations rather of mood than of feelings, i.e. without being very explicit and obvious, so that the other child can react not by immediate actions, but by the same kind of silent talk.

Combining micromovements with actions, children can actually conduct quite lengthy "silent conversations". Examples of this kind are described, for instance, by D.Lewis ([1980, pp.35-37]).

One must be aware that, although micromovements have meaning by themselves and even more in context with others, they cannot be precisely translated into linguistic form (D.Lewis uses transcriptions of their meaning, but he certainly does not mean them to be accurate translations). We cannot say that the meaning of a specific micromovement in a specific context is expressible by a unique sentence, or series of sentences, in such a way that one may not change any word in the sentence without altering the meaning of the micromovement. Actually, as we shall see later, even the meaning of spoken messages can be hazy and not uniquely determinable.

In view of these observations, how does the meaning of micromovements get encoded in the memory system? It seems very unlikely that it is encoded by some kind of propositional connections. It is much simpler to assume that this is done by direct sensori-motor and motivational associative connections (direct associative connections of sensations, motives and movements).

Actually, some micromovements, e.g. body tension, may simply be the first stages of a preparation we make, in order to proceed to a certain course of action, just as many macromovements seem to be ritualized patterns of action. Beckoning, for instance, is very similar to the movements we make when we try to reach something and it might be a stylized form of these movements.

Here we must be careful to note that some of these associations may not be empirical but preformed, since part or all of a certain behavior pattern may be inborn.

For instance, we know that smiling is inborn and not learned by experience, because even children born blind do smile (see [Eibl Eibesfeld, 1976, pp.20-27]) or ([D.Morris, 1982, p.18]). Laughing, just like crying, is also inborn since children born blind-deaf do laugh [Provine, 1996].

7.2 Language, Mental Models and Thought

But then what is the role that language plays in thinking? Development of linguistic abilities is certainly not a necessary prerequisite for the development of intelligence. Highly intelligent people have developed their linguistic abilities very late compared to other children. This was, for instance, the case with Albert Einstein (see Hoffman & Ducas, [1982, p.14]).

On the other hand, thinking seems inconceivable without language. When we start a new course of thought, usually we engage into some kind of inward talk, even if we do not care to communicate our thoughts to other people.

A plausible assumption, which may explain these facts from the point of view taken here, is that language is necessary in order to put together a first "mental model" (in Johnson-Laird's terms), on which further logical or associative thinking will be based.

Subsequent associative processing may play an important role in the thought process but first of all, a certain initial model of the situation has to be put together. Some initial associations have to be addressed and related to each other in a certain composite structure by the use of words. Otherwise, the search for common preconceptual content cannot be carried out.

Words are thus basically names, i.e. addresses for concepts, while sentences play the role of structural descriptions, relating concepts with each other in order to create some kind of model, based on the associative content of these concepts.

7.3 Language and Meaning – Is the Meaning of Linguistic Expressions Unique?

Since communication is possible without language, as we have seen, there are also "meanings" without linguistic expression.

"Meaning", in this sense, is of course either a memory content we intend to communicate to somebody else or a change in our view, in our model of the present situation, due to received communication. It may have linguistic form, but it may also consist merely of activation of other parts of the associative memory network.

But how close is the relation between meaning and language? How is meaning translated into, or extracted from linguistic form? We will come to this last question later. But before that, let us consider whether there is a basic difference between the meanings of verbal and non-verbal communications.

At first glance, it might seem that, while non-verbal communication has somewhat hazy and imprecise meaning, the meaning of verbal communication, i.e. of spoken language, is always precise and unambiguous.

Certainly, the use of language allows a high degree of precision of meaning, since extensive logical analysis and categorical classification has given very precise meaning to words, delimiting their meaning in relation to each other. Also, the study of syntax has established basic rules according to which, words should be put together in order to produce sentences with logically precise and unambiguous meaning.

But is this a feature of spoken language or basically due to the way we are taught to use language?

Let us consider more closely the relation between language and meaning. Theories on semantics are usually based on rules, which allow only a unique interpretation of a sentence. However, it seems very peculiar that a parallel processor, like the brain, should produce only serially ordered “meanings”. After all, there may be more than one memory content competing to be spoken out, due to parallelly running mental processes.

In fact, the serial, the sequential ordering of meanings is only an illusion due to the fact that in our basic education we are trained to express ourselves, especially when we are writing, in a precise and unambiguous way.

Spoken language of everyday life simply isn't of this kind. In everyday conversations we very often use unfinished sentences or even syntactically and logically incorrect ones. In spite of that, they are meaningful to those who listen, because they know the context (of events and thoughts) in which these sentences are spoken.

Sentences of everyday conversation are seldom conveyors of exact meaning. What they are meant to do, is not describe in precise terms something, but rather trigger some course of action, or reaction. They are motivated and meant to carry over this motivation and not some precise syllogism.

For instance, what does a mother mean when she says to the child: “If you don’t come immediately here ...”, leaving off the rest of the sentence? Possibly, even she does not know exactly what the second part of this sentence would be, because what makes her speak is the urgent wish to make the child come to her. In the back of her mind, various possibilities of substantiating the threat may compete at the same time, without ever taking precise linguistic expression even in her thoughts.

On the other hand, this sentence may induce various pictures of possible punishments in the child's mind or none at all, if he knows that his mother is lenient.

In order to see how loose the connection between linguistic expressions and meaning is, let us consider another example, based on personal experience. Returning by car from the countryside, at some point of the road I said to my wife, who was driving: “Next traffic light is Filolaou” (“Filolaou” is the name of a street). What was the possible meaning of that sentence?

First of all, the sentence was in various ways incorrect. The traffic light has no name, but also the street crossing our route at that traffic light is not called “Filolaou”. “Filolaou” is the name of the extension of the street we met at the crossing.

Furthermore, the sentence was certainly not meant to convey factual information about the name of a street, since my wife knew the street. Its meaning becomes somewhat more clear if one knows that the street we met at those traffic lights is a possible detour we might make, in order to avoid the center of Athens, which has usually very high traffic density.

But even if one knows this fact, can he correct and complete the sentence so that it precisely conveys what I meant? By no means. There are too many ways in which this sentence can be reasonably completed. But what is more, not all these sentences are equivalent in meaning.

I might mean that we should make that detour but I might also just suggest that we should discuss whether it would be advantageous to do so. I might also suggest to my wife to judge the possible density of the traffic at that time herself and decide on her own which course she should take.

What I "felt" that I meant, was all three possibilities and none of them specifically. All three possibilities "flashed" through my mind while I was speaking. But what I wanted to convey primarily, was a feeling of urgency to make a decision. This, and possibly the wish to remind my wife that we were nearing a crucial crossing ,made me speak in the first place.

This is a typical example of a sentence meant to "induce" various mental processes in another's mind, rather than convey any precisely definable message.

Everyday language is very often ambiguous like that, with many possible meanings. It is more like the ambivalent body language, than the precise, demotivated language of scientific papers (in scientific papers, at least those pertaining to physical sciences, "meaning" is carried over ideally only by the formalism, while the linguistic expressions connecting the formulae should not have any implicit meaning relevant to the subject considered. This is done for the sake of precision).

7.4 The Basic Stages of Linguistic Development and their Significance

In order to see how language operates, let us take a closer look at what is known about language acquisition.

As said before, language is learned only interactively. A non-impaired child of deaf-mute parents was unable to learn to speak by being exposed to TV programs, while he learned to communicate fluently in sign language (Moscowitz, [1978, pp.88-89]). However, contrary to the difficulty of finding universal syntactic rules, i.e. features of grammar and syntax common to all languages (see [Deacon, pp. 103-104] or [Aitchinson, pp. 176-177]), it has been readily established that language acquisition follows a very regular pattern: the early words used are primarily concrete nouns and verbs, later followed by adjectives and more abstract words. Newly acquired words are first used for naming and later for asking questions.

The first stage of linguistic performance is characterized by a maximum sentence length of one word, followed by a stage during which the maximum sentence length is two words.

It might seem that, at least in the one-word stage, children are not able to describe somewhat complex situations. However, this is not so. As R.Scollon has observed on a 19-month-old girl named Brenda, a child uses in dialogue a "vertical construction", a series of one-word sentences, in order to express what an adult might say with a "horizontal construction", a multiword sentence (see Moscowitz, [1978, p.91]).

This phenomenon is a further indication that children are capable of coherent thought, even when they are not yet able to express their thoughts in a coherent sentence.

The two-word stage is characterized by binary syntactic relations, corresponding to binary semantic relations. Typical examples are:

(a) subject noun - object noun, e.g. "Mommy sock", i.e. a possessor - possessed relation.
(b) subject noun - verb, e.g. "Johnny go", i.e. an actor - action relation.
(c) verb - object noun, e.g. "Read it", i.e. an action - object relation.
(d) verb - location, e.g. "Bring here", i.e. an action - location relation.
(e) noun - location, e.g. "Box there", i.e. an object - location relation.

The description of more complex relations is, at this stage, again done by a vertical construction, i.e. a sequence of two-word sentences. Surprisingly, a three-word period does not exist. The child passes directly to multiword sentences.

A possible explanation of this phenomenon is that ternary, or more complex relations can always be described by successive binary relations (e.g. the relation [John plays ball] can be seen as a combination of [John plays] and [play ball]). The construction of more complex sentences could therefore be based on combining binary relations.

The beginning of the multiword period is called period of "telegraphic speech", because it is characterized by short sentences made up primarily of content words (usually nouns and verbs), while it lacks function words (articles, prepositions, conjunctions, tense endings, plural endings etc.)[22].

All these features are subsequently gradually acquired, following always the same pattern: a rule is first applied non-specifically in all possible cases and then gradually restricted to its common usage. For instance, plurals are first built by adding always the suffix "-s" at the end of the nouns (e.g. "toy-s", but also "man-s").

In a similar manner, as we have said before, every newly acquired word is soon applied non specifically to describe various objects with feature similarities and then gradually restricted to a more specific usage. Some examples of this kind reported by [Moscowitz, 1978, p.92]) are given in the previous chapter.

We note again that the term "overgeneralization", used in both instances in the bibliography instead of "non-specificity", may be somewhat misleading because at this early period the child has not yet developed abilities for consistent use of inductive logic.

---

[22] Functional words are a typical example with sensori-motor encoding, rather than encoding by logical categories. They are symbols, which we learn to use in a certain way as designators of meaning, without having themselves a full semantic analysis.

7.5 The Analysis and Synthesis of Linguistic Expressions and the Role of Binary Relations

What are the possible implications of these observations for a fully developed system of language? A plausible assumption is that linguistic processes in adults follow the main paths, the main techniques, which have been formed during their language acquisition period.

In fact, it seems unreasonable and wasteful to assume that a child adopts, in each developmental stage of language acquisition, totally new strategies. So let us consider more closely what the implications of this evolutionary point of view are.

Both speaking and listening are motivated actions. We don't speak or listen unless we have some interest, some motivation. This means that a young child, who tries to say something, will arrange the one-word or two-word sentences of the vertical structure it possibly uses, according to their importance to him at the moment.

First comes the central object of interest or the most important action, and then additional specifications according to their relative importance at the moment. This importance criterion should, therefore, also determine the arrangement of phrases in a horizontal construction.

Both telegraphic speech and the absence of a three-word period seem to imply that multiword sentences are constructed by the use of binary relations, arranged according to their momentary degree of importance. This may also explain the impression raised by many experimental results in language understanding, that the basic storage units in LTM are "propositions", i.e. meanings corresponding to some kind of short phrases.

Does some other kind of phrase structure rules, i.e. rules generating complex sentences as entities, possibly exist at this early stage? If there were such a rule, then at least some of the function words should be present from the beginning of language acquisition. While they are mostly useless for expressing binary relations, they become much more important in the context of a complex sentence. Function words seem to be mainly means for subsequent "polishing" and "streamlining" of a telegraphic sentence.

The same principles should also determine, to some extent, what happens in an adult's mind when he produces a sentence. It may be, for instance, that he sometimes uses a ternary relation as an entity, if this relation happens to be very common, but by and large the sequence of processes taking place could be as described above.

These principles do not represent only a mechanism for sentence generation, but also give a rough idea of how parsing and understanding of a sentence we hear may proceed.

Parsing must be a systematic search for binary or, in some cases, ternary relations. This means that, possibly, everything we hear or read is kept in reserve in the echoic memory until the first part of a binary relation has been identified. This may be the first noun in a sentence but also a function word like "if", which indicates a binary relation "if-then".

Subsequently, everything else may again be kept in reserve until the second part of this binary relation has been identified. The same evaluation principle would then apply to the rest of the sentence.

A possible indication for the existence of such a process is the difficulty we have in remembering sentences with many interconnected secondary clauses. It is very difficult to keep in mind a primary clause, whose parts are separated by a series of who-, which-, while-, although-, nevertheless- etc. clauses. In order to see this difficulty, consider for instance a sentence like: "The man, whom I have adviced to go to the doctor, who had treated my friend, who helped me when I was on a journey, although he had problems of his own, because he had had an accident, is my neighbor".

In this case, it would actually be easy to remember what has been said, if we adopted the parsing strategy of identifying the first appearing binary relation each time, instead of the main one. This would yield: "I have advised somebody to go to a doctor", "The doctor had treated a friend of mine", "This friend had helped me when I was on a journey", "My friend had problems of his own", "My friend had had an accident", "The man, whom I have advised, is my neighbor".

Therefore, it seems that our difficulty in understanding and remembering the sentence lies in our parsing strategy. Obviously, hearing such a sentence, we keep everything else in the echoic memory until we have identified the main binary relation, the one that refers to the first noun, in this case, "man". As long as the chain of "who" continues, we know that the second part of this relation has not yet been reached and defer interpretation. Thus, if this chain is too long, we will probably forget a great part of the sentence, because we fail to interpret it in time.

The fact that too great distance between the two parts of a binary relation causes difficulty in understanding, is well known in Mathematics. This is why stating a theorem with many assumptions in the form, "If ... and ... and ... and ..., then …", is avoided and forms like "Suppose that...", "Also let ...", "Then ..." are preferred.

As said above, the reason, why binary relations play such an exceptional role, is probably,that they have a very important property. They allow the description of any more complex relation in the form of a series of binary relations. E.g. "John eats grapes" = "John eats" (Answer to the question: "What is John doing?") + "eats grapes" (Answer to the question: "What is he eating").

The parsing strategy proposed here is obviously content- and context-dependent, i.e. based on semantics. However, there are also opposite opinions, saying that syntactic processing is autonomous, i.e. independent of semantic and pragmatic factors.

Such views are probably influenced by the great success of generative syntactical systems (or transformational generative syntactical systems) in producing correctly structured sentences. However, they don’t take sufficiently into account that speaking is a motivated process. We do not speak in order to produce word-strings, but in order to say something that is important to us.

The whole process of speaking is based on meaning and not on structure, as we have seen in the previous chapter. Besides, psychological experiments have consistently provided evidence against the existence of an autonomous mental syntax (see [Johnson-Laird, 1983, pp.334-336]).

Furthermore, the theory, that language production and understanding is based on transformations from “deep-structure” to “surface-structure” (in Chomsky’s terms) and vice versa, has also failed to get any experimental support. People, who do not expect their memory to be tested, retain the general meaning of a sentence, but not its deep structure. They can not decide correctly, which of two sentences with similar meaning, but different deep structure they have heard [Johnson Laird, 1983, p.278].

7.6 The Role of Logical Categories and Language.

As said in the previous chapter, in order to be able to describe a certain situation to other people, we need not only attach “names”, i.e. words, to various nodes of the empirical association network, as it is seen here, but we must also adjust the meaning of these “names” to a common understanding, avoiding ambiguity as far as possible. This normalizing of the meaning of words involves, by necessity, categorization in nonambiguous categories, i.e. logical categories.

Logic, from this point of view, is neither an inborn, “instinctive” mechanism, nor a mere product of philosophical thinking, but a tool developed in order to enable communication between people.

Is there any psychological indication that this is so? “'Normalizing” of the usage of words is a process easily observable in young children. For instance, they often create their own words, their own names for things, which they then abandon in favor of commonly used words. Also, the process of replacing non-specifically used words, like “bow-wow”, by more specific ones, can be interpreted as a process of normalizing. In fact, even adults constantly adjust their usage of language by adopting new expressions, like “plastic money”, “mobile” (used for a wireless telephone receiver), “globalization” etc.

7.7 The Role of Language in the Mental System

This interpretation of the role of logical categories leads to another question: what is the role of language in our mental system?

First of all, we must abandon the idea that language is absolutely necessary for communication. As said before, another, very important but often not sufficiently appreciated way of communication is by "body language", i.e. by body postures and expressions of the face.

We also have to abandon the notion that language is absolutely necessary for thinking, although it is certainly very important for performing complex thought processes. As we have seen, young children are, even in the one-word stage of linguistic performance (when they use only single words), able to express more complex meanings by using in dialogue a "vertical construction", a sequence of one-word sentences, instead of the "horizontal construction", i.e. the many-word sentence we would usually use. This shows that they have more in their minds than single words, although they are yet unable to put words together in a composite linguistic expression.

Then what is the use of language? First, it provides a tagging system, an addressing system that makes what we have in mind more discernible. We can address special features of what we perceive and concentrate our attention on them. Furthermore, it allows us to put together complex models of some situation, so that we can mentally operate on them. Finally, composite sentences simplify communication considerably, because they transmit a whole model of some situation. Alternatively, this model could be transmitted only slowly, in a more primitive form, through a conversation conducted by an exchange of one-word sentences and gestures.

A Short Review

In this chapter we have discussed the relation between meaning, communication, language and logic.

At first, we have noted, on the basis of the observations of Ethology, that communication can also exist without linguistic expressions, by means of the so called "body language". We call body language the system of the usually unconscious, slight movements, changes in facial expression and body posture, which are mainly the "language" of feelings. This is used extensively and with great dexterity by young children, but it is concealed and replaced by linguistic expressions as they grow up. In combination with actions, this "language" can transmit an extensive repertoire of messages.

Then what is the role of spoken language in thought? A basic contribution of language seems to be that it allows the construction of composite mental models and so facilitates the mental processing of information.

Considering how unique the linguistic communication is, we saw that it is usually ambiguous, just as the body language. What is more, very often it does not carry over some information, but intends to trigger in the listener a series thoughts and reactions of his own. I.e., it is intended to transmit motivation rather than information.

After this, we have considered the main stages of language development in a child. We saw that the meaning of recently acquired or formed words is initially hazy, non-specific and only gradually becomes specialized to the common usage of the word. We also saw that language is acquired through interaction with the surrounding and not merely by listening to it.

With respect to the linguistic development, we have noted that it goes through three stages. The first is the stage of one-word sentences. Nevertheless, even in this stage, it is possible to express composite meanings, simply by a sequence of one-word sentences. It thus replaces the multi-word sentence. The second stage is the stage of two-word sentences, which are usually of the form: possesor-posessed, actor-action, action-object, action-place, object-location. The third stage is not a stage of three-word sentences, but immediately of multi-word sentences. These have initially a "telegraphic" form, i.e. they lack articles, prepositions, conjunctions, tense endings, plural endings, etc. All these are added gradually.

These data were used afterwards in order to form a plausible hypothesis about the wa, in which the analysis and synthesis of sentences is performed. According to it, the sentences are sequences of meanings, arranged in a receding order of importance to the speaker. I.e., the main meaning is placed first and the secondary ones afterwards. In this way, a series of one-word or two-word phrases composes a multi-word sentence. Similarly, the analysis and interpretation of a sentence we hear, seems to be an attempt to establish binary relations following the ordering of the words. An indication for this is the fact that we find it very difficult to understand a sentence, in which many secondary meanings are interposed between the two parts of the main binary relation.

Finally, we have considered the relation between logical categories and language. Our position here is that logical categories, i.e. categories with distinct features, are created in order to avoid ambiguous meanings as far as possible. I.e., they result from the necessity of normalizing the words, in order to make communication possible. From this point of view, Logic is neither an inborn "instinctive" mechanism, nor simply a product of philosophical thought, but an instrument developed in order to make communication between people possible.

BIBLIOGRAPHY

Aitchison, Jean: The Seeds of Speech, Cambridge U.P., 1998.

Deacon, Terrence: The Symbolic Species, Penguin, 1997.

Eibl Eibesfeld, Irenaeus: Liebe und Hass, Piper, Muenchen, 1976.

Morris, Desmond: The Pocket Guide to Manwatching, Triad/Granada, 1982.

Lewis, David: The Secret Language of your Child, Pan Books, London, 1980.

Hoffmann, Banesh & Ducas, Helen: Albert Einstein, The Human Side, Princeton U.P., 1982.

Johnson-Laird, P.N.: Mental Models, Cambridge U.P., 1983.

Moscowitz, Breyne Arlene: The Acquisition of Language, Scientific American, Nov.1978.

Provine, R.P. : Laughter, American Scientist, vol.84, 1996, pp. 38-45.

Chapter 8

DRIVES, INSTINCTS AND MOTIVES

Most writings on artificial intelligence are concerned with the question of how intelligent behavior may occur but they never consider the question of why it occurs. What causes anybody to ask questions about his or her surroundings and seek answers to them?

This is another manifestation of the curious neglection of the evolutionary aspects of intelligence, by many cognitive scientists (possibly a remainder of the damaging influence of behaviorism on this whole field of study).

We should not only ask how a concept develops into its present form, but also why it begins developing at all and what coordinates its development. To this question, the answer given by many biologists is that the crucial driving mechanisms for all behavior are innate releasing mechanisms, called instincts or drives.

8.1 The Difficulty of Clear Determination of Instincts or Drives

Instincts or drives are the "batteries" of the mental system (but also its "steering rod"). Just as the engine of a car needs the spark provided by a battery to make it work, the mental system needs the motivation given by an instinct or drive, in order to be activated and directed towards a certain goal.

It should be said right here that there is still no clear concept of what is an instinct or drive, i.e. how it operates, and no clear distinction between them. There are many reasons for this:

First of all, we can establish the existence of inborn mechanisms (i.e., automatic action or reaction schemata, not acquired by experience) only indirectly, by observing the stereotypical behavior of very young children or animals, or experimenting with them. This, however, does not tell us exactly what is the driving force behind that behavior. There may be different interpretations of exactly what kind of mechanism lies behind it, how general or how specific it is.

Second, some such mechanisms have no fixed form, but rather an initial direction of evolution, on the basis of acquired experience. The most prominent example of such a mental mechanism may be the language, which seems to rely on specific functions of some particular regions of the brain, but also has to be learned from one's parents and social surrounding.

Developmental aspects are present in all inborn neural mechanisms. They may be inborn, but they are also becoming increasingly differentiated as the conceptual system develops. For instance, social education may prevent someone from exhibiting aggression directly, or may even give aggression a more subtle direction. Similarly, hunger may be directed towards certain kinds of food, which experience has proved to be more tasty.

Besides that, inborn driving mechanisms may have individual variations, i.e. different strength in different children. These initial variations, the first rudiment of a personality, become further differentiated by experience, due to personal exploration and social influence. Thus, they create the complex personality of later years, whose motives are not always easy to establish. Clearly, these refined motives have their origin in the first individual mixture of drive variations everybody has when they are born, but also to the all embracing family and social influence.

As is also the case with the conceptual system, we can have definitions of instincts at the functional, macroscopic level or at the electrochemical and neurological, microscopic level. However, research in this field has not gone very far, in spite of its utmost importance for the development of intelligence.

One thing is clear; the behavioristic view of a mind free from any inborn driving forces and information filters is false.

8.2 On Behaviorism

Behaviorists have held the view that there are no innate (i.e., independent of the surrounding) factors that influence behavior, but that all behavior is due to acquired, i.e. learned, reactions to outward stimuli. What others call instincts or drives are, according to them, not inborn but simply acquired conditioned reflexes, that is, specific modes of reaction to certain outward stimuli, acquired by repeated pleasant or unpleasant, gratifying or frustrating experiences.

However, in saying that, one is already obliged to admit that there is some inborn mechanism directing the organism towards gratifying, i.e. pleasant, experiences. Thus, a behaviorist can possibly only reduce all behavior to one instinct, the seeking of gratification of basic needs of the body.

But this view is easily disproved by the observation of typical instinctive reactions almost immediately after birth, or even in absence of the possibility of acquiring responses by conditioning (social or natural). Such is the case of smiling, laughing and crying, which are responses appearing even in children born blind.

In addition to that, there are patterns of behavior that do not tend towards the fulfillment of personal needs, but are only explainable as means of improving survival of the species. For instance, the so called “social stress” can lead a sheep to loss of reproductive abilities, weight loss and even death [Wickler, 1972, p. 72].

This happens if we place two sheep of the same sex (e.g., male) in too narrow a space. At first, there is a fight between them for the establishment of superiority. Then the loser, without being further attacked, starts losing weight, its testicles return to the inside of the abdomen and the production of semen ceases.

Obviously, such a reaction does not serve the narrowly undestood interests of an individual but tends towards the survival and reproduction of the dominating animal. Such kind of behavior cannot result by a stimulus-response reinforcement, but rather phylogenetically (by evolutionary survival of the animals that have this property).

Another objection to using the term "instincts" is that it is not always easy to distinguish between them, as we have already said above. But this does not mean that we can simply ascribe, a priori, all behavior to a single drive, the seeking of pleasure or gratification.

There are two directions in distinguishing innate reaction mechanisms. One is to look at the basic biological needs of an organism and try to distinguish between them. Another is to observe typical reaction schemata of animals and try to distinguish those acquired only by experience, i.e. by learning, from those that seem to preexist before any opportunity to learn.

The first direction was chosen by theorists and experimentalists like Watson, Skinner and Clark Hull. The second was chosen by the school of researchers like Eric von Holst, Nikolas Tinbergen and Konrad Lorenz, which call themselves "Ethologists". Hull and his successors (and partial opponents like Tolman) speak of "drives", while the ethologists use the term "instincts".

We will look into these two directions of research starting with instinct theory.

8.3 The Investigation of Instincts by Ethology

Observing animal behavior, we see that in most cases it is a reaction to outward stimuli. However, there are also cases in which the animal acts spontaneously, without any outward cause.

For instance, a starling brought up by Konrad Lorenz sometimes flew to the ceiling, snatched with its beak at the air, as if it was catching something, returned to its nest and made movements as if it was killing and swallowing some prey. However, there was no prey there to be caught.

Such unjustified appearance of a certain kind of behavior is ascribed by Ethologists to the existence of inward motives which, if not satisfied for a certain period of time, seek relief even without an outward stimulus. I.e., the repetition of some stereotyped behavior, without any outward stimulus, reveals that the central nervous system in some instances produces its spontaneous activation, thus leading to the relief, i.e. release, of some concentrated "inward energy", which is connected to that specific behavior.

This "energy" may be due to the concentration of hormones or other chemical substances (e.g. neurotransmitters) or even to outward factors, like the interchange of light and dark (day and night) or the change of the seasons. For instance, many animals' migration or reproductive beahavior depends on the seasons.

According to the definition given by N.Tinbergen (see [dtv-Atlas zur Biologie II, S.389] [Bunk – Tausch, 1975, p. 213]): "instinct" is an inborn, hierarchically organized nervous mechanism, which reacts to certain triggering and coordinating impulses, inward as well as outward, and answers to them with well coordinated life- and art-survival actions.

Just like the intensity of the outward stimulus, so is the inward readiness variable and depends on the instance, as well as the biological constitution of the individual.

The various kinds of stereotyped behavior, e.g. nest building, attracting a female, caring for the youngsters, defending one's territory, are on the one hand independent modes of behavior, but they also belong to the wider class "Reproductive Behavior".

This wider class constitutes a center of instinctual behavior, which coordinates and integrates its various, more specific modes into one whole entity. These more specific modes of behaviour again constitute subordinate centers, to which even more specific modes of behavior belong.

A principal instinct, like the reproductive instinct, has an immediately subordinate center, which we might call mating instinct. This again can be subdivided into a variety of lesser sequences of behavior structures, like the zigzag "dance" of the male stickleback, a sweetwater fish that leads the female to the nest in order to lay the eggs, the fertilization of the eggs by the male, etc.

There is, therefore, a hierarchy of instincts, coordinated with each other and either cooperating or competing with each other, in which case the dominant one suppresses the functioning of the others. For instance, defending one's territory or caring for the nestlings cannot appear at the same time as attracting a female.

This means that different instincts of the same level (e.g. reproductive and hunting instincts) inhibit each other. I.e., different instincts of the same level, like the reproductive instinct and hunting instinct (food search instinct), each suspends the activation of the other.

Each instinct seems to have its own reaction energy, stored in centers of the nervous system, but usually also requires the appearance of appropriate outward stimuli in order to manifest itself, i.e. to cause the appearance of a certain kind of stereotyped behavior. Then we say that the behavior connected with the instinct is "released" by those specific stimuli of the surrounding.

For instance, young silver seagulls open their beaks, asking for food, as soon as they see a red spot on the beak of their parent and they also do the same when shown a spot of this colour on a vivid yellow pencil or other dummies similar to a beak ([dtv-Atlas zur Biologie II, p. 389], [Lamprecht, p. 38-39], [Gellatly-Zarate, p.50]).

But not all sequences of behavior that lead to the satisfaction of an instinct are entirely mechanical. Tinbergen distinguishes between "appetitive activity", associated with the search for a goal, and "consummatory activity", consisting of the responses occurring when the goal is reached.

Consummatory activity is rather simple, having the form of a fixed chain of responses. But appetitive behavior allows a great variety of responses, from mere reflexes based on primitively learned reactions, to even insightful behavior.

Tinbergen uses as an example the behavior of the peregrine falcon. Random searching of a wide area leads to the spotting of a potential prey. Then the falcon initiates a strategy for catching the prey, still within the limits of appetitive activity. Only when the prey is caught is the chain of consummatory behavior enacted.

There is a vast and very interesting observational and experimental material collected by ethologists, showing that some inborn neural mechanisms are not merely directed towards the fulfillment of basic organic needs, as drives are supposed to do, but influence even social behavior, i.e. fulfill social needs.

Such instincts are the ones leading to the building of social hierarchies on the basis of ritualized fights between animals living in groups (the, so called, "pecking order" of a poultry yard).

We know that at least part of this social behavior is instinctive, for instance because there are typical inhibiting postures of the defeated animal, which hinder the dominating one to kill the loser (these are usually imitations of the postures by which young animals ask for the protection of older ones).

Observation shows that the winner of the fight may still be in a very aggressive mood, but the surrender-ritual of the defeated hinders him absolutely from directing his aggression towards him. This accumulated aggression is then directed e.g. towards trees or other objects of the surrounding.

Such inhibitory signals exist primarily in species who have dangerous natural weapons, like sharp claws and teeth, while they do not always operate successfully in animals without such natural weapons (e.g. humans or pigeons) [dtv-Atlas zur Biologie II, p. 403].

A typical instinctive behavior is the highly species-specific song of birds. These signals are uniform and highly standardized within each species, in order to attract mates and repel intruders.

The act of singing is certainly a fixed action pattern, although it may be complex. However, in order to develop the typical repertoire of melodies, the bird, e.g. a canary, must have the opportunity to imitate other birds of its species during its youth. Birds, e.g. chaffinches, reared in isolation, sing a song having the right length and the right number of notes, but that is not recognizable as the typical song of their species (see [Barnett, p.229]). What is inherited in such a case is not a full behavioral repertoire, but rather the ability to develop it through learning by imitation.

Acquirement of human language seems to be a similar instinctive behavior, since the mother tongue is acquired by imitation and social interaction.

A simple example showing how universal and detailed the instincts are, is our attraction to babies, even animal ones, and our tendency to protect them. Babies are recognised by certain physiological features: oversized eyes compared to the size of face, puffy and round cheeks, a very large head compared to the size of the body etc. We find any animal with such features attractive. That is why babies are often used in advertisements, while almost no advertisement shows old people.

In fact, the instinct to protect babies, even those of other species, is active in all mammals. See [Eibl-Eibesfeld, p. 58] about the strength and importance of this instinct.

Let us now briefly consider drive theory.

8.4 Drive Theory

Clark Hull's theory for the analysis of behavior was the most influential theory in the psychology of learning and motivation, in the English-speaking countries during the nineteen thirties and forties. It remains influential even today, although it has been considerably modified, as we shall see.

Following the spirit of behaviorism, Hull believed that learning is a formation of habits, i.e. bonds between stimuli (S) and responses (R) and that these S-R connections explain the direction of all acquired behavior. Such connections were supposed to be created and fixed by the operation of reinforcement, i.e. by constantly rewarding a specific response to a certain stimulus.

However, since an animal will not start responding to a stimulus in any way, unless it is pushed to action by some inward factor, Hull introduced the term "drives" to describe such inward forces.

In his view, all drives were due to primary needs for self preservation, i.e. the needs for water, food, air, correct temperature, rest, sleep and avoiding damage to the body. He also believed that habits are only strengthened if they succeed in diminishing the intensity of a drive, i.e. that the nature of reinforcement is drive reduction.

This stimulus-response psychology of motivation was contradicted already in 1930 by Tolman and Honzig [1930], on the basis of their experimental results. Tolman and Honzig performed an experiment with three groups of rats, each of which were introduced into a maze. The first group received a reward each time they reached the end of the maze, the second group never received a reward and the third group only after the eleventh trial.

As expected, the first group steadily improved its performance in learning the right way to the end, making almost no more errors after the twelfth trial, while the second group never improved its performance. The surprising achievement was that of the third group, which learned the way to the end of the maze almost as good as the first group, only in two or three trials after the reward was introduced.

This contradicted Hull's view that habit builds up slowly as a result of reinforcement and must be due to a considerable increase of the motivation of the animals, after food reward was introduced as an incentive.

According to Hull, there can be no learning without reinforcement, in the form of drive reduction (e.g. of hunger). Habit strength is measured on the basis of the number of reinforced trials. But the rats of group three reached perfection in their task very soon after introduction of an incentive, although they had received no reinforcement up to then.

Incentives bring about an almost instant learning of the correct way through the maze, if previous experience with the maze exists. Obviously, the animal already knows much about the maze at the eleventh trial, when it is motivated to choose a certain route.

Reinforcement is thus not necessary for the occurrence of learning. The animal is not simply mechanically driven by inward and outward forces into a sequence of responses ending in the goal box. What it really does is to acquire information about how stimuli tend to go together. It learns stimulus - stimulus (S-S) connections, rather than stimulus - response (S-R) connections. When it is given some incentive to select a certain goal, it utilizes almost instantly the experience, the knowledge of the environment, that it has already acquired.

Later drive theorists, trying to explain animal behavior that is not directed toward the satisfaction of primary biological needs, have introduced the view of the existence of secondary, i.e. acquired drives, but practically they have considered only one: fear.

Experiments with animals receiving, e.g. unpleasant electric shocks, together with some other stimulus, show that fear can be learned as a drive and is extremely resistant to extinction. The animal may never cease to respond to the warning stimulus, although it is no more accompanied by a shock. Thus, in addition to drives satisfying primary biological needs, there are also drives directing towards avoidance of other stimuli.

Fear has been used in order to explain, for instance, why we spend so much time of our life in order to earn money. Money may be a secondary reinforcer, but this does not explain why so many people try to hoard ever more money. Most of our primary biological needs can be satisfied with relatively little money. Millionaires would have no real reason for trying to increase their fortune.

Trying to explain this behavior, drive theorists say that a person hoarding ever more money is just afraid of not making money, since it is very highly valued in human society. From the point of view of Ethology, trying to become ever richer is one way of gaining more power in human society, a behavior that may be instigated by the instinctive tendency to reach a higher rank in the human social hierarchy (in the human pecking order).

Today's drive theorists admit the existence of many more drives than those directed toward the satisfaction of primary biological needs, including social drives etc. in their theory.

D. S. Wright, Ann Taylor and their coworkers [1974, p.207] give the following definitions for the terms "drive" and "need":

Drive is the purposive activity, which is initiated by both the internal state and external stimulation of an organism.

If in a form of directed behavior the internal state of an organism can be identified as the primary instigating condition, the internal condition is called a "need".

They also point out that many drives appear to be initiated more by external stimulation than by any clear definable state of need.

Such a position does not, however, explain why some animals feel the drive to maintain a personal territory, defending it against intruders even in the case of overabundance of food. Such a reaction is only explainable by the phylogenetically established necessity of maintaining a sufficient personal territory for hunting or grazing.

Some of the basic drives are activated in order to serve the purpose of acquiring the necessary nourishment and fluids for the body. Others, however, like the sexual drive, need a sufficient outward, as well as inward stimulus in order to be activated. Also, some drives, e.g. hunger or thirst, exhibit a reduction of the stimulus after satisfaction but others, like the seeking of pleasure, may not do so. Experiments show that some animals never tire of pressing a bar that activates the pleasure center in their brain, through an electrode.

Variations in the internal state of the organism are, to a large extent, regulated by mechanisms of homeostasis (from the Greek words ὁμοιος = similar and στάσις = staying). We call homeostatic those mechanisms that try to restore the physical conditions of the organs, like temperature and the concentration of vital substances, e.g. Sodium and Potassium of the blood, to certain optimal levels.

When body temperature increases, the organism tries to reduce it, e.g. by perspiration. Similarly, the level of sugar or fluids in the bloodstream, necessary for efficient functioning of the body, is periodically restored by eating or drinking. All disruptions of this equilibrium are subjectively felt as needs and it is quite customary to refer to the physiological disequilibria as homeostatic or "biogenic" needs.

Some theorists suggest that such needs are the ultimate springs of all action, while behavior that cannot be seen as arising directly from homeostatic needs, has been assumed to derive indirectly from them, through learning.

However, this is an oversimplification, because some drives do not seek the restoration of an optimal level of physiological parameters. For instance, both animals and people often seek food containing an excess of sugar, although the body has no need for it. Similarly, mice given the alternative to press levers, either providing food or activating the pleasure center in their brain, constantly chose the second alternative and ignored their hunger.

Drives are usually classified into two categories: basic drives and acquired or learned drives, although it is not always easy to distinguish between innate and learned behavior.

However, certain motives seem universal and appear without any discernible opportunity to be learned, so that we can be fairly confident that there are inborn. These motives include, first and foremost, those called "self-preservative" and "species-preservative", but also some other drives, like the exploratory drive or curiosity and the proximity-seeking drive (the tendency to form attachments to living things), which do not spring from biogenic needs in the narrow sense. In spite of that, on the basis of experimental evidence, they also appear to be innate rather than acquired.

8.5 The Basic Kinds of Drives

Here we will discuss the basic drives, starting with those due to biogenic needs, like hunger, thirst and the sexual need. Then we will turn to the cognitive drives, i.e. general stimulus seeking (the tendency to maintain an optimal level of stimulation), proximity seeking and curiosity. Last we will discuss fear and aggression.

8.5.1 Thirst and Hunger

Thirst is controlled by centers in the hypothalamus, which respond to feedback concerning the state of cellular dehydration. There are two kinds of thirst: volemic, resulting from overall lowered blood volume, and osmotic, resulting from the hypertonicity of body fluids, i.e. from the increase of concentration of salts in it.

Hunger is also controlled by hypothalamic centres, whose activity is directed towards a homeostatic regulation of food intake, by means of feedback, e.g. of information about sugar availability in the blood. Destruction of these centres abolishes hunger, while their electrical stimulation induces eating in satiated animals.

8.5.2 The Sexual Drive

Although this drive depends on the presence of sex hormones in the bloodstream, in higher mammals and human beings it is not entirely controlled by them. Removal of testes and ovaries, which secret the sex hormones, does not necessarily abolish sex behavior, while removal of parts of the cerebral cortex eliminates sexual activity completely.

Sexual arousal depends partially on tactile stimulation and reflex responses, mediated in vertebrates by the lower spinal cord, and coordinated by a particular region of the hypothalamus. The drive intensity depends also on the presence of appropriate outward stimuli, visual, olfactory, auditory, etc.

Especially in species lower in the evolutionary scale, these outward stimuli have usually the form of inflexible patterns of behavior. The appearance of the male releases a particular pattern of behavior in the female, this evokes an appropriate courtship behavior of the male, which causes the next response of the female, etc.

In higher mammals, the sexual behavior is less rigid, being modified by experience, learning, and social standards.

Parental behavior is related to the sexual drive in many species, because it is controlled by hormonal secretions. The maternal drive, for instance, depends in mammals, to a considerable extent, on the presence of the hormone prolactin in the bloodstream. Injecting prolactin into adult virgin female rats makes them accept and care for infant rats.

8.5.3 The Seeking of Stimulation

Sensory deprivation studies indicate that animals and people require some optimal level of sensory stimulation in order to function normally. Otherwise, hallucinations may appear in order to compensate the lack of sensory inputs.

This is one of the findings that do not agree well with the classic drive theory that tries to reduce all behavior to homeostatic needs. Homeostatic mechanisms raise stimuli, which cause the organisms to try to reduce this stimulation by satisfying certain needs. Thus, an aching stomach tells us that we have to eat something in order to eliminate this ache.

However, while some mechanisms in the body indeed try to keep the level of internal and external stimulation low, it seems that there are other "stimulus-seeking" mechanisms (drives), which try to avoid a deficit of stimulation.

For instance, while "proximity seeking" is important but not always predominant in young animals, they seem to always seek stimulation, no matter the kind (visual, auditory, tactile etc.).

This need for outward stimulation may be directed toward sensory stimulation but also toward cognitive stimulation, which is related to what we often call curiosity.

8.5.4 Proximity Seeking

This behavior has been observed both in newborn birds and newborn mammals. Very soon after birth, they tend to approach any object or source of sound that stands out against the background. If this object moves away, then the newborn tends to follow it.

The animals showing this tendency are so very young, that acquirement through reward learning is very unlikely. Indeed, this response itself is a starting point of learning, because the young animals responding in this way gradually become attached to the object they follow, so that they later prefer the proximity of this familiar appearance and sounds to other ones.

In the first months of life, proximity seeking is prevailing, while proximity avoidance of unfamiliar objects and people shows itself only later.

Studies of young primates show proximity seeking to be a normal pattern of behavior towards anything that resembles a furred animal. Thus, infant monkeys cling to "cloth mothers" in absence of any reward, as if the contact itself gave them comfort.

As noted above, although proximity seeking is important for young animals, it is not always predominant , while they always seem to seek stimulation.

8.5.5 Curiosity and Playing Drive

Under the influence of behaviorism, most motivational theories between 1925 and 1955 were dominated by the rather simplistic view that all behavior could be regarded as evoked by homeostatic needs, sex, or painful experiences; if not directly, then indirectly by conditioned reflexes. A small number of innate drives, together with conditioned learning of various reflexes, was considered enough to account for all motives.

More recent studies, however, show that organisms can be active even in the absence of any of the above motivating conditions.

When animals and people seem to wander about, exploring the surrounding, they may be considered to be feeling an urge for locomotion or sensory stimulation, but they may also be seen as being simply curious, i.e. looking for new experiences without any particular purpose.

Curiosity seems to be a basic drive, especially in higher developed animals. Exploratory or investigative behavior is systematically studied since the early 1950s. Its basic difference from apettitive behavior is that it is not directed mainly towards a consummatory act, like eating or mating. It is also similar to it in seeking a variety of goals. Typically, there is a tendency to approach something unfamiliar, then withdraw, approach some other unfamiliar thing, then possibly return to the first etc.

An indication that curiosity is a genuine drive are the experiments of Montgomery [1954], which showed that the exploratory tendency in rats is so strong that it can induce conditioned learning, the same way biological needs do.

Thus, the opportunity to explore unfamiliar regions acts as a strong incentive to learn to distinguish stimuli that lead to this goal. The attractiveness of this incentive does not seem to diminish if the rats had recently the opportunity to explore something else [Halliday, 1966].

Similarly, experiments with rhesus monkeys showed that they learned to choose between yellow and blue colors for the reward of viewing their surroundings for a short while [Butler, 1954]. They were more interested in viewing and hearing other monkeys than other moving objects or food, but viewing anything new was preferred to seeing the same old monotonous environment.

Very much related to curiosity is the playing drive ("Spieltrieb" in German). Its significance is obvious; play is a means of exercising, i.e. of acquiring skill with certain action or reaction schemata. Similarly, curiosity is a means of undirected search for new possibilities, which may provide very useful new experiences.

The existence of these mechanisms at the macroscopic level seems obvious, but the nature of the corresponding neurological, electrochemical or hormonal triggering mechanisms is not yet clear.

8.5.6 Fear

Fear is an emotion based both on innate motives and on experience. The existence of innate fear reactions was shown experimentally by Walk and Gibson in 1961. They found out that very young animals of various kinds (rats, chicken, lambs, turtles, kittens, puppies, monkeys) as well as human babies, which did not have any experience of falling, showed clear signs of fear to move on, when placed on a surface with a "visual cliff" (a visual cliff is a transparent floor extending over a real cliff).

Experimental evidence shows that young animals respond to stimuli of any type, either by approaching whenever the stimuli are weak, or by withdrawal when they are strong. Thus, it seems that there is an inborn mechanism of fear towards strong stimuli.

Fear, of course, is also brought about by learning that certain, not necessarily strong, stimuli are related to unpleasant experiences or danger to the health or very existence of the animal.

Hebb observed that the strongest feelings of fear are evoked by objects sufficiently similar to familiar objects to arouse interest, but also sufficiently unlike them to disrupt the memory patterns built up by previous experiences with them.

People also show an affinity towards experiences familiar to them from childhood and an animosity towards anything strange and unusual, not fitting their habits.

8.5.7 Aggression

Aggression can be evoked, first directly, by unconditioned reflexes, when a prey is perceived or when the boundaries of the personal teritory, e.g. the hunting grounds, are violated by another male.

Second, it is often evoked in frustrating situations and directed towards innocent weaker or subordinate bystanders or towards objects, in order to release some of the emotional strain built up by the frustrating experience.

Third, it is evoked by acquired conditioned reflexes or as a reaction mode which is learned because experience has shown that it leads to success, in certain circumstances or situations.

Konrad Lorenz considers aggression in people as very dangerous because people , like all animals without natural weapons like claws or sharp teeth, lack the restraining instinctive mechanisms which hinder, e.g. carnivorous animals to kill an opponent of the same species during a conflict.

However, experiments with animals show that the rearing conditions can greatly influence the exhibiting or not of aggressiveness (dogs, cats and mice can live together peacefully, if they have grown up together and never learned to be aggressive to each other).

8.6 The Conflict of Drives

Inward conflicts are an integral part of life, because it steadily requires decisions. Having to choose between two or more alternative actions, means that we have to resolve an inward conflict, i.e. the incompatibility of two or more motives acting at the same time.

There are two basic kinds of inward conflicts. The first, called "approach-approach conflict", is the situation of having to choose among two or more desirable actions, e.g. what menu to choose at the restaurant. This conflict is easily resolved once a choice has been made, possibly on the basis of some superficial criteria, because the positive effects of one action overshadow the expectations from other choices.

The second kind of conflict is called "approach-avoidance conflict", because the very goal one wants to approach is also something he wants to avoid. We may be offered a very tasty dish which, however, is very fattening. Shall we succumb to the temptation or rather think about the addition of calories, which we have to reduce by subsequent fasting?

We are possibly faced with a far more serious decision of this kind when choosing a profession. We may feel that we have a talent for a profession like painting or writing poetry, which is seldom well paid. Shall we dedicate our lives to it or rather choose a boring but more secure and well paid profession?

How do we reach a decision in the case of such conflicts? Superficial criteria are not enough when our whole future may depend on our decision. Making a decision in such circumstances requires an internal maturing, a basic decision concerning priorities in our lives. Such inward conflicts are the most difficult to resolve. That is why they often lead to neurotic indecision.

8.7 Assessment of the two theories

"Drives" and "instincts" today seem to be almost equivalent terms, since drive theorists have included social instincts as well as the exploratory drive, i.e. curiosity, in the drive repertoire. They have also abandoned simplistic views about drive mechanisms. As a matter of fact, connected with the drives discussed abov there is a vast repertoire of stereotypical behavior patterns, which govern various aspects of these drives or instincts. Such behavior patterns have been extensively studied by ethologists.

In a sense, whether the above drives are the only ones, or whether others can be distinguished is a matter of classification. Certain forms of behavior can be ascribed to some basic drive/instinct or considered as a separate action schema.

From our point of view, the realization of the instrumental role played by the motives in the creation of new ideas and concepts, is more important. This is a subject that we are going to discuss more extensively in the next section.

Another important question, raised by the admission of the existence of drives or instincts is: how automatic are all these schemata of action and reaction? I.e., to what degree does a person control his actions and reactions?

What we consider as freedom of will may simply be the result of increasing differentiation of the instinctive behaviour due to different experiences, which lead to a different conceptual development of each person.

Since these are rather philosophical questions we are going to discuss them in **Appendix 8.1: An investigation of the freedom of will**.

8.8 The Role of Motives in Spiritual Work

As said before, no scientific or artistic work is possible without motivation. However, motives are also important for scientific work for yet another reason: they predetermine what kind of answer the scientist will look for. Very often, the inward preferences are responsible to a great extent for the nature of the theory a scientist will create.

Georg Cantor, for instance, was a transcendentalist. He believed in the existence of the Transcendental. It was, therefore, natural for him to create a theory of infinite sets and infinitistic counting concepts. He states, for instance [R.Rucker, 1984, p.43]: "The fear of infinity is a form of myopia that destroys the possibility of seeing the actual infinite, even though it, <u>in its highest form, has created and sustains us</u>[23], and in its secondary transfinite forms occurs all around us and even inhabits our minds".

Cantor made infinity discernible. By distinguishing between many different orders of infinity, he made them self-existing. He gave them an independent

[23] Underlined by the present author.

existence, in contrast to the shadowy, tentative existence that infinity had only as a limit up to then.

This also made the existence of a self-existing "highest form of infinity" more accessible. A rationalist, believing in a finitistic description of the world, would never look for such concepts, no matter how talented he might be.

Of course, one can always claim that these concepts lie in the nature of the integers and real numbers, so that they would inescapably be discovered at some time or another. But this view too simplistic. Many mathematical concepts lie not in the nature of mathematical formalism in which they appear, but require an act of bold imagination by a mathematician in order to be created. They are, so to speak, desperate measures, trying to overcome a basic difficulty by all means and in spite of common sense.

Such is for instance, as we have said before, the case of the imaginary numbers, like 2i = $2\sqrt{-1}$. They were created in order to make all quadratic algebraic equations (e.g. $x^2+4=0$) solvable, although they were strange to common sense. There was no "proper" number whose square would equal -2, since the squares both of positive and negative numbers are positive. This is also why they were called "imaginary".

More than two centuries after their creation, they were made plausible (accessible to common sense) by identifying them with points of the plane, i.e. two-dimensional vectors. But this had nothing to do with their initial conception, which was merely due to Girolamo Cardano's bold imagination.

More specifically, with respect to Cantor's cardinal and ordinal "numbers", we should not forget that, even today, many mathematicians do not like the concept of static infinity as an independent number, but prefer to interpret infinity as the abstraction of approaching ever larger numbers.

Thus, motives are not merely a blind driving force for research, but often influence immediately the nature of the results obtained.

Considering Physics, we see that the "beauty" of a theory, its symmetry, coherence, universality, simplicity etc., seems to be a basic motivating force for its admission. The physicist Hermann Bondi, for instance, reports that Einstein would accept or reject a suggestion on the basis of such criteria. "He was quite convinced that beauty was a guiding principle in the search for important results in theoretical physics." Such guiding principles are necessary because, as Einstein put it, "experience may suggest the appropriate mathematical concepts, but they most certainly cannot be deduced from it." [Roger Newton, 1994, p.14-15].

These esthetic values are connected to the exploratory drive. Investigation of a situation always seeks the simplest, the most parsimonious explanation or solution that is compatible with all data, because such a solution is easier to keep in mind. Thus, Mendeleev's Periodic Table of the Elements is much easier to work with than a vast catalogue of elements and their properties. Its "beauty" lies in its symmetry, which makes it easier to handle.

Similarly, if we put mice in a maze, they always seek the shortest way out of it.

8.9 The Relation of Instincts to Unspoken Meanings and Messages

Cognitive scientists very often consider the linguistic concepts as the only carriers of meaning. This, however, is a too restrictive notion of "meaning".

Outward stimuli, incoming through the senses, e.g. a sudden change in temperature, are also carriers of "meaning" for the organism, which do not have to be expressed in concepts. The withdrawal reflex when we touch something hot is a sensible and meaningful behavior, which presupposes no conceptual analysis of any kind.

All outward stimuli that may cause an instinctive response of any kind are obviously "meaningful", in some sense. For instance, many animals mark their territory by urinating on trees or rocks. Thus they leave olfactory messages for others of the same species, which roughly say: "Stay away, this is my territory". All triggering signals are meaningful, since they cause reactions. However, their meaning can be primarily instinctual and emotional instead of conceptual.

We must also note that the way we conceptualize incoming information from the senses depends on how we evaluate it. We constantly filter incoming information, selecting certain signals to which we pay attention and overlooking and ignoring what does not seem important to us generally, or at the moment.

We may note that there is a cleaning shop in our neighborhood only when we need one, although we have passed it hundreds of times up to that moment. This evaluation of outward stimuli is obviously based on sensitization of our attention for specific stimuli, like the sign at the front of a shop, but it takes place before we consciously know what we have perceived. Here we have meaning before conscious understanding.

General filtering out of certain outward stimuli is obviously ultimately based on evaluating reflexes, inborn or acquired. Such evaluations presuppose a preliminary selection of meaningful information. Every animal focuses its attention not on everything on the surrounding, but on what is a potential food source, a potential mate or a potential danger. However, this focusing presupposes meaning before the animal knows exactly what caused its attention. Perhaps this is done by an active search of all incomming stimuli, in order to find stimuli connected with what is of immediate interest, what is timely.

A Short Review

In this chapter we have discussed what kind of inborn mechanisms are the motives of human actions.

These mechanisms are called Drives or Instincts and are discovered by means of observations of stereotypical behavior of people and animals, especially of young ones. However, their clear determination, their clear distinction, is difficult for two reasons:

(a) Stereotypical patterns of behavior do not appear individually, but usually consist parts of a composite pattern of actions and reactions, which are due to a whole hierarchy of motives.

(b) Such inborn patterns of behavior are not absolutely immutable, but develop continuously together with the biological development of the organism and are influenced by the environment.

I.e., they are not based on a fully developed neuronal mechanism, but rather constitute the outward manifestations of various stages of its development.

As the most prominent/impressive example of this phenomenon, the language instinct was mentioned, which is of course based on the functioning of certain cerebral centers, leading not to a stereotypical set of linguistic expressions, but to the assimilation, each time, of the language of the social environment.

We have also discussed briefly the behaviorist view, that there are no innate (independent of the environment) factors which influence the behavior, beyond the elementary instinct of self-preservation. According to this view, all behavior is due to acquired reflexes to outward stimuli. We have rejected this view, mentioning various examples of behavior which are not acquired reactions of self-preservation, but can only be explained as means for preservation rather of the species than of the individual.

Considering the efforts to find inborn mechanisms of reaction, we saw that they have two different directions: one is to look at the basic biological needs of an organism and to try and distinguish between them. The other is to observe typical reaction schemata of the animals and try to distinguished those acquired only by experience, i.e. by training, from those which seem to preexist, before any opportunity to be learned.

The researchers of the first direction usually speak of “drives”, while those of the second direction speak of “instincts” and call themselves “Ethologists”.

Now, we have first considered the results of Ethology or Behavior Research. We saw that the Ethologists establish, on the basis of observation, that the inborn action and reaction mechanisms called “instincts” include a great variety of forms of behavior. They also establish that these forms of behavior are organized in hierarchies of instincts, which are coordinated with each other and either cooperate or compete with each other, in which case the dominant one suppresses the functioning of the other competing ones.

Considering the views of the drive theorists, we saw that, influenced by behaviorism, they initially believed that all drives are due to primary needs of self-preservation. I.e., they distinguished the needs for nourishment, rest and avoiding damage to the body. They also believed that habits are only strengthened when they succeed in the reduction of a drive’s intensity. However, it was shown experimentally, already in 1930, that this immediate stimulus-response connection is wrong. Similarly, it was gradually established that the drives are of many more kinds than initially believed.

Subsequent drive researchers have distinguished and studied experimentally, beyond those mentioned above, also drives for sexual satisfaction, seeking of sensory stimulation, proximity seeking for curiosity or play, fear and aggression. Thus, they came closer to the views of the Ethologists.

Considering, after that, the relation of instincts and spiritual work, intellectual creation, we saw that, very often, the dominant motives of an individual pre-determine the kind of answer he will seek to a scientific problem, the kind of theory he will create. Similarly, it is often that esthetic values determine the kind of theory that will be created in Physics. Such values are, however, indirectly connected with instinctive propensities.

Finally, we have considered the relation of instincts with unspoken meanings, unspoken messages. We saw that the instinct provides meaning for every incoming message, since it evaluates it. The instinct is thus creator or carrier of meaning, which does not have a linguistic expression of any kind.

BIBLIOGRAPHY

Barnett, S.A.: 'Instinct' and 'Intelligence', Penguin/Pelican, 1970

Bunk, B. – Tausch, J.: Grundlagen der Verhaltenslehre, Westermann, Braunscweig, 1976.

Butler, R.A.: Incentive conditions which influence visual exploration, J.exp.Psychol., vol.48, 1954, pp.19-23

Eibl-Eibesfeld, Irenäus: Der Vorprogrammierte Mensch, dtv, München, 1976.

Evans, Phil: Motivation, Methuen & Co, London 1975.

Gellatly, Angus - Zarate, Oscar: Mind & Brain for Beginners, Icon Books, Cambridge, 1998.

Halliday, M.S.: Effect of previous exploratory activityin the exploration of a simple maze, Nature, vol.209, 1966, pp.432-433

Lamprecht, J.: Verhalten, Herder, Freiburg, 1972.

Lea, S.E.G.: Instinct, Environment and Behaviour, Methuen, London and New York, 1984.

Montgomery, K.C.: The role of exploratory drive in learning, J.Comp.Physiol., vol.47, 1954, pp.60-64.

Newton, Roger G.: What Makes Nature Tick?, Harvard U.P., Cambridge, Mass. 1994

Rucker, Rudy: Infinity and the Mind, Paladin, Granada, London, 1984

Squires, Euan: Conscious Mind in the Physical World, Adam Hilger, Bristol & New York, 1990

Vogel, G. - Angermann, H.: dtv-Atlas zur Biologie, Band 2, Deutscher Taschenbuchverlag, Muenchen, 1968

Wickler, Wolfgang: Sind wir Sünder? Naturgesetze der Ehe, Drömer - Knaur, München/Zürich 1972.

Wright, D.S.- Taylor, Ann: Introducing Psychology, Penguin, 1974

Tolman, E.C. and Honzig, C.H.: Introduction and removal of reward and maze performance in rats, Univ.Calif.Publ.Psychol., vol.4, 1930, pp.257-275.

Chapter 9
CONSCIOUS AND UNCONSCIOUS

Various features of the memory model proposed here, for instance the haziness of the meaning of preconcepts and the linguistic nonadressability of lower level preconceptual structures, seem to be related to unconscious mental processes. In order to investigate the nature of this kind of unconscious processes, we must first make clear what has to be considered as conscious.

9.1 Unconscious, Semi-conscious and Conscious Information

Obviously, incoming stimuli cannot be considered as conscious information until they have reached a parsimonious enough representation (i.e. interpretation), mainly in terms of known concepts, so that they can be more specifically attended to. Before they reach this stage they could be called, as A.Marcel [1983] proposes, "preconscious" at most.

However, conceptual interpretation and attention are not enough for consciousness. When I walk on a street, thinking about some problem (be it a personal problem or a scientific one), I am certainly more conscious of the problem than of the surrounding, although I am able, at the same time, to "recognize" obstacles and avoid them or recognize familiar clues, which allow me to choose my way without ever interrupting my course of thinking. These recognition processes are ,at best, "semiconscious" and not remembered afterwards, although every thought related to the problem under consideration is well remembered.

9.2 Consciousness as Self-relatedness, i. e. Correlation with the Ego

What makes incoming information unconscious (or at least, semiconscious) is seemingly that it is not closely enough related to one's immediate interests; that is, his feelings, wishes, goals and recollections about his recent and past personal history, in short, his momentary self-image.

Therefore, such information seems to be evaluated (used) only for present purposes and then discarded, i.e. very soon eliminated from the Short Term Memory, without being stored for permanent use in the Long Term Memory.

Apparently, the brain does not simply process information for the sake of information processing. Its main purpose is to serve the goals of the organism. It therefore pays much more attention to information which is directly related to the organism's motivation, keeping it in the STM as long as the motivation predominates, but also storing it permanently in the LTM in a more or less detailed form, if it is important enough.

Such information, related directly to one's goals and past history, seems to form a kind of coherent "personal diary" in LTM, which is probably what we call "ego" or "myself".

When somebody says "I", he does not basically mean an organism, in which various biological processes take place at the particular moment, but rather a being with certain motivations and goals, related in various ways to other persons, to the present environment, to other environments kept in memory and to various recent and more distant past events. All these memory contents form the "personal diary", which is some times called "ego" (i.e. "I" in, Greek) or "self".

Adopting this terminology, we may say that consciousness is ego-relatedness, or rather that degree of consciousness is degree of ego-relatedness. Similar definitions of consciousness are given by many writers: C.G.Jung defines consciousness as the function that connects the psychical contents (obviously meaning the active memory contents) with the Ego, while Marvin Minsky says that consciousness involves one part of the mind monitoring the behavior of other parts [Horgan, 1993, p.15].

9.3 Awareness and true Consciousness

Antonio Damasio [1997-98, p.90] defines consciousness as the biological phenomenon that permits us to survey the contents of our minds: our feelings, thoughts and knowledge. He notes that consciousness is neither the feelings themselves, nor the accumulated thoughts and knowledge in our minds. It is rather an elaborate rite of passage into these feelings, these thoughts and this knowledge.

Damasio also distinguishes between essential (i.e., elementary or basic) consciousness, which he calls "awareness" and extended consciousness, which is what we usually call "consciousness".

Both occur automatically and are similar, letting a living organism sense that the contents of its thoughts are its own and that it can act on those thoughts. But awareness refers mainly to the present and can be found not only in people, but also in other animals, while actual consciousness encompasses a much broader spectrum of thoughts that portray not just the organism's present, but also its past and anticipated future. It performs, for this larger mental repertoire, the same tasks as awareness does for its much narrower range: it places mental contents in an individual perspective and it gives the owner the sense that he can act on them.

Damasio also notes that, "ironically, the machinations behind consciousness are non-conscious". For example, when we retrieve a word to name a particular object, we have no awareness of how we perform this task.

This seems to agree with the idea of a preconceptual search for clues connecting the present experience with past ones, stored in the LTM.

9.4 The Psychoanalytic Conception about Conscious and Subconscious or Unconscious

The psychoanalytic school of thought, which goes back to Sigmund Freud, Alfred Adler, Carl Gustav Jung and others, calls "Conscious" the part of (long term) memory which can be spontaneously recalled in imaginal, echoic, but mainly linguistic form (and obviously, related to our "self").

All psychoanalysts place, in contrast to this, another part of the memory, which they mainly study, the "Subconscious" or according to Jung, "Unconscious". That is what they call the part of the memory which cannot be easily recalled or not at all (presumably because it refers to deeper motives), although it is part of one's personal history and may nevertheless strongly influence his behavior; it manifests itself e.g. in dreams or impulsive reactions.

9.5 Distinction of Psychoanalytic Theories from Phenomenology

Since psychoanalytic theories are considered today, especially by cognitive psychologists, as fanciful inventions and not as scientific theories, before discussing their relation to what is said here ,we should consider their credibility.

This credibility concerns us here only with respect to the phenomenological basis of these theories, i.e. what is directly observable or a reasonable assumption based on direct observations. All extensions beyond the phenomenological basis, as well as the effectiveness of these theories as instruments of psychotherapy, do not concern us here.

First of all, we must give Freud credit for establishing the relation of the symbolic contents of dreams to our personal history, our experiences and our drives. No matter whether his interpretations of dreams are valid or not, it is important to note that dreams are not inspirations of mysterious external spiritual powers, but of our own mind, referring indirectly, i.e. symbolically, to personal experiences.

Second, we cannot doubt the existence of an unconscious part of the mind, no matter whether its core is always infantile sexuality or not (today, even psychoanalysts doubt the general existence and dominant influence of an Oedipal Complex in the formation of human behavior [John Horgan, 1996, p.74]).

The contents of this Unconscious are certainly related to each other, i.e. build complexes or rather, "emotionally accentuated complexes" ("Gefühlsbetonnte Komplexe"), as C.G.Jung has initially called them, no matter what their specific content may be.

Certain experiences or drives are certainly buried in the Unconscious and may suddenly become active, leading us to inexplicable behavior, which may be triggered by seemingly insignificant clues.

Thus, when we speak here of the Unconscious and of Complexes, we do not mean the interpretations of psychological phenomena presented by

Psychoanalysis or Analytic Psychology, but only the common basis of all these theories.

9.6 Experimental Indications of the Existence and Nature of the Unconscious

Referring to the existence and nature of the Unconscious, Daniel Goleman [1996, p.60] notes: "As Freud made clear, a great part of emotional life is unconscious; feelings that stir within us do not always cross the threshold into awareness. Empirical verification of this psychological axiom comes, for instance, from experiments on unconscious emotions, such as the remarkable finding that people form definite likings for things they do not even realize they have seen before. Any emotion can be – and often is – unconscious."

A proof of this is that, as Goleman reports, "the physiological beginnings of an emotion typically occur before a person is consciously aware of the feeling itself. For example, when people who fear snakes are shown pictures of snakes, sensors on their skin will detect sweat breaking out, a sign of anxiety, though they say they do not feel any fear. The sweat shows up in such people even when the picture of a snake is presented so rapidly that they have no conscious idea of what exactly they just saw, let alone that they are beginning to get anxious. As such preconscious emotional stirrings continue to build, they eventually become strong enough to break into awareness. Thus there are two levels of emotion, conscious and unconscious. The moment of an emotion coming into awareness, marks its registering as such in the frontal cortex."

9.7 The Contents of the Unconscious according to Freud, Adler and Jung

According to Freud and Adler, the unconscious part of the memory contains personal experiences, which have been repressed, because they are very painful (traumatic). Such experiences are due to conflicts between personal wishes, based on sexual, social or other drives and restrictive social values. They also tend to cluster around the repressed wishes that have caused them, forming what is called "complexes".

Here we may note that not all conflicts are due to a discord between wishes and social norms. Conflicts can also be due to different inward drives. We learn to hide wishes, not only when they are socially not acceptable, but also, for instance, in order to improve our bargaining position in the social interplay. This may go so far that we may hide wishes from ourselves, if we think that they can not be fulfilled.

According to Jung, the Unconscious is divided into two parts; the "Personal Unconscious" and the "Collective Unconscious". The Personal Unconscious contains not only the above repressed experiences, but also other factors determining our personality, most predominantly the unconscious parts of the Ego-Complex.

As said before, Ego is the memory imprint of our personal drives and wishes (the vast reservoir of our personal history) in their temporal development. This is a vast network of interconnections, which can never become activated as a whole, all at the same time! Even wishes which are not repressed are not always active, i.e. present in our conscious mind, while there are also deeper layers, which never become conscious.

9.8 The Various Kinds of Complexes according to Jung

C.G.Jung calls this Ego-network a "complex", which is the most important of all such complexes, the center, so to speak, of all conscious and unconscious network connections.

According to his theory, a complex may have conscious, semi-conscious, as well as unconscious parts. Similarly, the "Ego" has a conscious part and a much wider unconscious part.

What other complexes can there be except the Ego-Complex? Complexes due to underdeveloped personality traces, which may lead to infantile behavior, unlike our usual mature behavior, which stems from the developed part of our personality. The other complexes are, so to speak, infantile side personalities, which had never the opportunity to mature and get integrated in our main personality (e.g. to adjust themselves to an acceptable social behavior). That is why, when they incidentally become dominant and express themselves outwardly, they sometimes give the impression of a separate personality living inside us.

Such a sudden predominance over our usual self may occur because of some outward provocation, as well as a long lasting neglecting or repression of the drives that are behind it.

9.9 Drives and Unconscious

Since not all drives operate at the same time, but one dominates the others, all our actions may get reevaluated from another drive's point of view, when this drive becomes dominant, which can happen, e.g. if it has been neglected for a long time.

Therefore, drives sometimes manifest themselves as independent personalities, which operate from the Unconscious with irresistible power and may drastically and at first glance, inexplicably, change our behavior[24].

---

[24] A dramatic story, exemplifying how catastrophic a long suppressed instinct can be, is told by Klaus Mehnert, a political scientist, in his book "Der Sovietmensch" (The Soviet people), Rohwolt, Stuttgart, 1967, p.39:

The self-preservation drive combined with the sexual drive may account, for instance, for the sudden urge of some people, reaching middle age, to seek erotic adventures. Diminishing sexual powers are felt like a threatening loss of a certain part of their self-image and unconsciously they try to reconfirm it. There is nothing strange in this interpretation, no matter whether we wish to accept the existence, e.g. of the Oedipal Complex, or not. We may consider following up various courses in life, but finally we choose one. Everything else remains in the background.

In my view, a theory of the Unconscious is acceptable up to this point. All further extensions may be valid for certain persons, or not at all. On the other hand, I think that it is nonsensical to try to deny the great power of our personal drive-blend (which is at the core of our personality, our Ego) and its extensions, our various personality traces.

Drives are not merely there; they constantly get developed and differentiated. This means that they form modes of action and reaction, which represent emerging personalities. Underdeveloped evolving personalities of this kind are carefully hidden, just as an artist hides the work he considers as immature and presents only the work he considers to be mature.

9.10 Collective Unconscious, Preconcepts and Creativity

The "Collective Unconscious" is a deeper layer that Jung considers as the origin of all creative powers. He describes the Unconscious as an "ocean", from which the Conscious "gradually emerges like an island".

Jung deduces the existence of the Collective Unconscious partially from the existence of impulses to action without conscious motivation, which appear uniformly and regularly in all people (all races) [Fordham, 1966, p.23]. This kind of behavior is called "instinctive" and considered to be determined by the past

---

"In the village near Moscow, where we used to spend our vacation before the First World War, the story spread one day that a hermit, a holy man, came from the North for a visit to his family, which he had abandoned to serve God.

Along with the others, I too ran to the hut, which they pointed out to me. There, on the wooden bench in the small garden, sat a man with a gray beard, who seemed very old to me, but must not have been over fifty five. He seemed totally ragged and wore a long white shirt with holes and frays. From his ascetic face looked piercing eyes, set in deep holes. All this corresponded exactly to my idea of a saint. The hermit spoke seldom. I could not understand what he said, but it was received with reverence by the peasants. …

However, not even a week from his return had passed and one morning I was raised from bed by a neighbor boy. Something terrible and extremely interesting had happened. We ran to the hut of the holy man. It was no more there. One saw only smoking ruins. […] The wife and the daughter of the hermit were searching the ashes, crying. The hermit himself was absent. From the crowd standing around, I learned that the holy man had drunk too much in the evening, then he had broken everything, he had hit his wife and he had done to his daughter something terrible, incomprehensible to me. Then, he had set fire to the house and had left. On that same day it became known that, after he woke up from his intoxication, he had surrendered alone to the police and had already been transported to the prison in Moscow".

history of mankind, in the sense that the brain has been shaped and influenced by the remote experiences of mankind.

In other words, the various adaptations of the human species to changing outward situations have left their trace in its mental system, by shaping its modes of action and reaction and its way to comprehend and experience life in general.

From the point of view of Cognitive Psychology, the existence of a Personal Unconscious seems perfectly acceptable, but outside its scope. On the other hand, assuming the existence of a Jungian-type Collective Unconscious seems totally unnecessary and is simply ignored.

However, empirical associative mechanisms, as they are discussed here, somehow provide a counterpart particularly to this kind of Unconscious, although this interpretation may not be exactly according to C.G.Jung's views. For instance, it is difficult to say what Jung's "Archetypes", which are the main contents of the Collective Unconscious, might be.

More specifically, a memory system like the one described here, must have some contents which are unconscious, not because they are repressed, but because they are not easily expressible, or not at all expressible in linguistic form. They are not merely unconscious, but also not consciousable.

As said before, lower level (earlier) structures of the associative network, for instance those built before the age of two years, certainly have no names. Such memory structures are also increasingly hazy (nonspecific) as we go from developed concepts to lower levels, i.e. previous organizational stages.

In some sense, this non-specificity, or haziness, is some kind of generality, because the same preconcept is applicable in various instances, which may differ from each other. A common feature is thus detected not by induction, but by use of a nonspecific recognition device.

From another point of view, however, increasing non-specificity means increasing difficulty in describing something. As the distance from the fully developed concept, which underlies some word, increases, the correlation between the preconceptual structures and this specific word decreases. So it becomes ever more difficult to describe the distant preconceptual structures by that word or words related to it, in a logically well-defined and obvious way.

Trying to approach lower conceptual, i.e. preconceptual, levels is thus in some ways like trying to tell somebody how to find his way in a city without street names. It would become increasingly difficult and very soon impossible, both for the speaker to give precise instructions and for the listener to understand them. Lower level memory structures are unconscious in a more fundamental way than merely repressed experiences. They are not consciousable, consisting an Unconscious which is not related to repressed traumatic experiences. What is more, they play an important role in creative mental processes, similar (but possibly not identical) to the role attributed by Jung to the Collective Unconscious.

However, a reference to a Collective Unconscious raises the question: is there anything "collective", i.e. common to all people, in lower memory structures? Certainly, similar experiences will lead to similar associations in different people. But what is more important, deeper layers of this system, the initial ones, consist of inborn structures.

Even so, specific means of nonlinguistic communication, such as smiling, are basically inborn and not acquired by experience, as the smiling of children born blind shows. Therefore, a vast number of initial connections must be pre-formed, or rather formed basically independent of outward experiences.

Various external factors may influence the development even of an embryo but, at least before birth, heredity must be the predominant factor.

Considering the main character of these initial connections, we must note that the most important hereditary factors for mental processes must be the basic drives, the initial control hierarchies, which gradually get organized into more subtle motives and determine our behavior and even what experiences we collect.

As said before, there is usually a great variety of outward stimuli, which impinge on our sensory organs. Which of these stimuli will be more closely attended to and how we will react to them, depends on the drives predominant at the given moment.

9.11 The Drives as Molds for Thought and Ethics

Thus, drives play a crucial role in the forming of experiences and therefore, of the empirical associations. They are molds which form in a very general, but also very decisive way, our thinking and our conceptual system.

Our concept of right and wrong behavior depends, for instance, very much on the relative strength of social, versus basic, exploratory drives in us. If social drives predominate, any behavior opposed to the norms set by the society is considered wrong. In the opposite case, we feel that it is important to explore a new situation, no matter what restrictions are imposed by social norms.

From our point of view, the important point in this analogy is the central role that C.G.Jung ascribes to the Collective Unconscious as a source of all creative powers of the mind, which is exactly paralleled by the role of the preconceptual system.

9.12 The Communication of Unconscious and Conscious – The Role of Symbols and Allegories

What we have said about not consciousable mental processes leaves an open question: how can the results of such mental processes be communicated to our Conscious?

Obviously, the most direct way, if not the only possible, is by symbols and allegories. Any common feature between a present situation and past experiences, discovered at those lower memory levels which are intrinsically hazy in meaning, cannot be easily put into words or not at all, while it can be very easily "enacted", i.e. put into some form of allegory or pantomime, which stresses that feature.

An example of this kind is Kekule's dream of snakes biting their tails, which indicated to him that the molecule of benzene may have the form of a ring (see [M.Hunt, 1984, p.257]).

On the role of symbols in psychical processes, Daniel Goleman [1996, p.239] notes: "One way to get at the picture frozen in the amygdala [by traumatic experience] is through art, which itself is a medium of the unconscious. The emotional brain is highly attuned to symbolic meanings and to the mode Freud called "primary process"; the messages of metaphor, story, myth, the arts.

This avenue is often used in treating traumatized children. Sometimes, art can open the way for children to talk about a moment of horror that they would not dare speak of otherwise."

Symbols and allegories seem to be not only the appropriate means of expressing "hazy" relations, but also the only available at those lower levels of the memory system, which are not yet directly connected with words and language.

On the other hand, this idea is again much closer to the role attributed by Jung to symbols and allegories appearing in dreams, than to their characterization in all instances as mere disguises for objects related to painful experiences, as is done by Freud.

A Short Review

In this chapter we have discussed the relation of the considered mental mechanism and the theories about the Unconscious.

Initially, we saw that many mental processes are semi-conscious, because the incoming messages are so many and change so quickly, that there is no time for them all to become conscious. Only those related to the objects of dominant interest for us, at that time, become conscious.

After that, we have discussed what can be called Consciousness and we have seen that the main feature of becoming conscious is the interconnection of the events with the complex of motives and memories about ourselves, our personal "diary", which we call "Ego". Furthermore, we have considered Consciousness as a graduated function and we have defined as degree of Consciousness, the degree of interrelation of the incoming information with the Ego, with our present interests and motives.

Then, we came to the views of psychoanalysts about the Subconscious or Unconscious. We saw that they call so the part of the memory which cannot be easily recalled to mind, in spite of the fact that it is a part of each one's personal history, or cannot be recalled at all (presumably, because it is related to deeper motives). Nevertheless, it can strongly influence a person's behavior and it is revealed, e.g. in dreams or inexplicable impulsive actions.

The opposition of many present day psychologists to the psychoanalytic theories about the contents of the Unconscious has led us, here, to make a distinction between what is phenomenology in the Psychology of the Unconscious and what is possibly an imaginative interpretation of phenomena.

Here we have presented experimental indications about the existence of unconscious feelings. When we show pictures of snakes to people who are afraid of snakes, detectors placed on their skin reveal the release of sweat, a sign of fear, even if they say that they do not feel any fear. What is more, the sweat appears in such people, even when the picture of a snake is presented so quickly that they have no conscious idea of what they have seen.

After a brief presentation of the views of Freud, Adler and Jung about the contents of the Unconscious, we have focused our attention on the concept of (emotionally accentuated) complex, according to Jung.

Jung considers as “complex”, a network of experiences stored in memory, which are concentrated around a central drive and are evaluated by it. We saw that he considers the Ego as the dominating complex, which has both conscious and unconscious parts and determines our conscious personality.

According to this view, secondary complexes represent secondary, undeveloped or partially developed personalities. If any external factors or long negligence grant priority to the drives, which consist the kernel of such a complex, then this can dominate temporarily the Conscious, influencing our behavior and manifesting itself as a personality alien to our usual self.

Considering now Jung’s view that, beyond the Personal, there is also a Collective Unconscious, we found a parallell of it in the network of preconceptual connections, which we consider as a substratum of all the conceptual system. According to Jung, the Collective Unconscious is the source of all creative psychical processes, while the network of preconcepts is, similarly, the source of all intuitive inspirations.

We have also discussed briefly the role of drives as molds of thought, but also as sources of ethics. As we have seen, our behavior depends mainly on what motives, e.g. social or exploratory motives, have priority for us. Therefore, what is correct and what is not correct, for us depends on the priority we grant to the various motives.

Finally, we have discussed the means of communication between Unconscious and Conscious and stressed again the fact that the symbols and allegories, which appear so often in our dreams, are the only effective means of rendering preconceptual content, preconceptual relations, conscious.

BIBLIOGRAPHY

Antonio R. Damasio: A Clear Consciousness, Time, Winter 1997-1998, pp.88-90.

Fordham, Frieda: An Introduction to Jung's Psychology, Penguin, 1966.

Goleman, Daniel: Emotional Intelligence, Bantam 1996.

Horgan, John: Profile: Marvin Minsky, Scientific American, Nov.1993, pp. 14 -15.

Horgan, John: Why Freud isn't Dead, Scientific American, December 1996, pp.74-79.

Hunt, Morton: The Universe Within, Corgi Books, 1984.

Marcel, Anthony J.: Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition, Cognitive Psychology, 1983, v.15, pp.197-237.

Marcel, Anthony J.: Conscious and Unconscious Perception: An Approach to the Relations between Phenomenal Experience and Perceptual Processes, Cognitive Psychology, 1983, v.15, pp.238-300.

Chapter 10

THE INCOMPATIBILITY OF ASSOCIATIVE "MEANINGS" WITH LOGICAL CATEGORIES AND THE SIGNIFICANCE OF MOTIVATION

Considering the dualistic modelization presented here, we should not fail to note one of its basic features. Empirical associations and logical categories are basically "incommeasurable[25] magnitudes", incompatible magnitudes. Empirical associations do not admit a finite expression with the help of logical categories.

Empirical content can be approximated in terms of logical categories, but it can never be exhausted in this way. A model based on empirical encoding and logical explication is therefore a carrier of potentially infinite logical information.

10.1 Examples of Incompatibility of Logical categories and Empirical Associations

10.1.1 The Impossibility to Express Linguistically Most Observable Shapes

Consider, for instance, the shape of an object as it is experienced by sight. How would we go about to name it? Of course, there are many straight lines or circular shapes around us. But not all shapes have names and relatively few shapes have outstanding defining features.

Certainly, we may call a face round, spindle-like, almond-like etc., but this is not enough. Most shapes can be described only analytically, by functions. E.g., in the case of plane figures we must use mathematical expressions, either of the form $x=f(t)$, $y=g(t)$ or of the form $F(x,y) = 0$, where $f(t)$, $g(t)$, $F(x,y)$ are appropriate functions. Most of these functions have no name, even in mathematics.

Actually, there are infinitely many such functions, even uncountably infinitely many. So there is no appropriate linguistic description for most of the shapes around us.

---

[25] "Incommeasurable" (in-com-measurable) are called magnitudes, which do not have a common measure, i.e. cannot be precisely described as finite expressions of a common measure. Thus, e.g. the side and the diagonal of a rectangle have respectively length approximately $10m_1$ and $14m_1$ or $100m_2$ and $141m_2$ or $1000m_3$ and $1414m_3$ etc., where $m_1 = 0.1$ $m_2 = 0.01$, $m_3= 0.001$. However, there is no common measure, m, so that both are integer multiples of it.

We use the likening with this mathematical concept on purpose, because it shows very expressively that two incompatible magnitudes can, nevertheless, be "almost compatible". I.e., we can express them approximately by a common basis of description, but different each time, depending on the degree of accuracy by means of which we wish to express their relation.

10.1.2 The Impossibility of Expressing Linguistically, with Precision, Feelings as well as Social and Biological Phenomena

Similarly, it is u

can we describe how intense it is? There is no precise linguistic scale of graduations of joy or of any other sentiment.

What about social situations? Is there a finite number of concepts, a finite language which would allow the description of social relations in an arbitrary society? An exhaustive set of concepts of this kind cannot exist, because beyond all typical social phenomena, there are always the variations due to the different personalities that participate in a social system. A charismatic personality can sometimes transform a society beyond any typical instances and stereotypical social relations. Remember, for instance, how much the campaigns of Alexander the Great changed the face of the ancient world or how much certain societies were transformed by creators of new religions, like Budha or Mohamed.

What about biological phenomena? Do we have a complete language for them? In view of the progress in molecular biology, it seems that there is an immense number of new biological organisms that may be created. The organic phenomena possible in all these organisms are even more numerous.

10.1.3 The Impossibility of Expressing Linguistically, with Precision, all possible Physical Structures, all possible Machines

What about Physics? Even if we assume that a final concise mathematical description of all basic physical phenomena will eventually be reached, this may exhaust Theoretical Physics but it certainly does not exhaust Applied Physics, the study of all kinds of mechanisms, i.e. Engineering.

There is a potentially infinite number of machines that can be created and studied, since every computer algorithm is equivalent to a special purpose machine. As Allan Turing has shown, an appropriately programmed computer can simulate an arbitrary digital machine, a discrete state machine [Turing, p.2106].

Moreover, continuous state machines are even more numerous than discrete state machines, i.e. algorithms, since they admit an infinity of input forms and internal state forms.

Just as irrational numbers, like $\sqrt{2}$, can be approximated by rational ones (fractions), but never exactly represented by them, in a similar way empirical associations can undergo various logical classifications, but never be exhaustingly described in this way. This makes creativity intrinsic to them. They are carriers of potentially infinite logical possibilities, which can never be exhausted.

10.2 A Comparison of the Mechanism of Evolving Empirical Associations with "Learning Automata"

We may also note that there is a basic difference between the, so called, "learning automata", which have already been extensively investigated and the mechanism considered here. Learning is here not merely a collection of additional factual information, but also a process of concept formation, i.e. a process that also changes continuously the way in which new information is encoded. It is a continuous reshaping or reformation of the cognitive mechanism.

From this point of view, the grasping of new ideas, e.g. creating new scientific theories is a neverending dynamic process, because it is to a great extent based on a process of concept evolution.

Especially with respect to Mathematics, we may note that progress brings rather an increase than decrease of the number of open questions, in spite of the settling of old ones. The reason for this is that this progress reveals an ever richer growing structure of the already known formal systems, besides often creating new structures and new objects of study.

10.3 The Necessity of a Dynamic, i.e. an Evolving and not Static Model of Memory – The Primordial Significance of Motives

No matter how successful the way of modeling presented here is, it is very important to recognize the necessity of considering the memory system not statically, but dynamically, i.e. time-dependently. The question of how a certain memory structure has evolved into its present form is basic and cannot be ignored. Continuing evolution makes the whole memory system, as we have seen, a carrier of potentially infinite possibilities.

Equally important is the recognition of the primordial significance of motives, even if not much is yet known about the mechanisms of drives and instincts. The study of such parameters of the formation of the conceptual system may seem unnecessary because, to many people, they do not seem to affect the basic structure of Long Term Memory, in any way.

However, speaking of the mental system without reference to motives, is like speaking of a machine without reference to the source of energy that makes it move. What is more, this system is not merely powered by drives, but also formed by them.

Although drives, i.e. control hierarchies for actions and reactions, do not seem to be present, to extend in higher levels of the conceptual system, they are, nevertheless, ubiquitous, since they are at the basis of this system. The whole system is not formed only by experiences, but also by the drives. The dominating drives, at a given moment, select the outward stimuli which will form our experiences.

10.4 The Complexity of the Proposed System

Such a system, as the one presented here, may seem complicated and not easy to imitate. However, its complexity does not arise from the fact that it is based on too many principles, but only on the fact that, in some sense, it is being continuously restructured. The principle used here is, in fact, only one; evolutionary procedural semantics. Those who seek some other principle, resulting in an easier imitable system, overlook what is at stake. We do not wish to produce merely any information processing system, but a creative one.

This attempt to give a unified view of the mental system leads also to a basic philosophical question: what is the meaning of Truth, as a general concept, in such an evolving system? Is there something we may call absolute or ultimate Truth and what is its nature?

What many people would like to believe, is that behind all varying phenomena there is a fundamental truth, an absolute truth. Then, we may try to approach this truth and thus attain an absolute compass for life. This gives us mental security, tranquility of mind.

We will discuss this philosophical question, which is also the background of Platonism, in **Appendix 10.1**.

A Short Review

In this chapter we have emphasized again the fundamental incompatibility of logical categories and experiential associations. We have also likened this relation to that of “incommeasurable” magnitudes. In this way, we have been able to explain that this incompatibility does not hinder the approximate description of empirical associations by means of logical categories. This may, eventually, happen with ever higher accuracy, but this process has no end. This feature makes the empirical contents of memory, carriers of potentially infinite logical information.

In order to support these observations, we have initially considered the shapes of objects of our environment and we have seen that they have so many different forms, that most of them cannot be expressed lingustically; they cannot be reduced to simple logical categories. Instead of these, the mathematicians use mathematical formulas, whose number is literally infinite.

Similarly, we have considered the vast world of sentiments and we saw that it is impossible to give a precise linguistic description of all these feelings. We have also made similar observations for social and biological phenomena.

Finally, we have considered all possible physical systems and saw that they must be infinite, since the number of all discrete state machines is already infinite. Such a mechanism can be simulated by a computer program and the number of such programs is infinite.

Comparing the proposed mechanism with the, so called, “Learning Automata”, we saw that these cannot exhibit the versatility of a dynamically evolving conceptual system. We were thus led to the conclusion that, even if the proposed mechanism is not very successful, any creative mechanism for storing and processing information must not be static, but dynamically evolving. I.e., it must not only store information by means of a fixed conceptual system, but also steadily reform, remodel the storing system, the conceptual system itself. We have also stressed again the necessity of recognizing the significance of motives for the functioning of thought.

Finally, we have noted that the complexity of the system proposed here is due, not to the use of many different mechanisms, but only to its flexibility. Any arbitrary simplifications may end up to a not truly creative system.

BIBLIOGRAPHY

Turing, A.M.: Can a Machine Think?, Mind, 1950. Reproduced in J.Newman (ed.): The World of Mathematics, New York, Simon and Schuster, 1956, pp.2099-2123.

Szabo, Istvan: Einführung in die Technische Mechanik, Springer-Verlag, Berlin 1958.

Chapter 11
A GENERAL OVERVIEW
(Or: How to See the Obvious)

11.1 The Basic Elements of the Mechanism of Cognition

Today there is a widespread view that the mechanism of thinking is an impenetrable mystery. Many neurologists, cognitive psychologists and investigators of the Theory of Automata express pessimism about the possibility of directly composing an overall organizational schema of cognition[26]. It seems difficult for them to see how the vast amount of data, which has already been collected about the mechanism of cognition and the functioning of the brain, can be synthesized to form a coherent whole. Thus, they believe that further long-lasting research is needed, until this stage is reached. Other, more daring researchers (mainly physicists), were thus lead to the following conjecture: that the mechanism of information processing in the brain is based on mysterious molecular mechanisms, quantum phenomena and as yet unknown fields[27].

Nothing of the sort is necessary, however. The solution of the mystery is simple, if we carefully utilize the data provided mainly by Developmental Psychology, i.e. if we consider how the cognition develops with time, as the child grows up. The difficulties, mentioned above, arise only if we insist to consider the conceptual system as something static, something already determined.

This misconception has been possibly supported by our great wish to create thinking machines. Therefore, up to now, it has been attempted to describe the brain as a discrete state machine, which analyses all cognitive processes by means of logical categories. I.e., it was attempted to somehow determine the "atoms of meaning" and compose more complex meanings as some kind of "molecules", by means of these "atoms", these "elementary meanings" that supposedly compose all thinking. Thus, many people have been misled to seek features that do not exist in the mind and have overlooked its obvious properties.

---

[26] See John Horgan: The Undiscovered Mind, Touchstone, N.Y. 1999. See also: Jerry Fodor: The mind doesn't work that way, MIT press, 2000 or Gary Stix, 2001: A Scorecard, Scientific American, January 2001, σελ.26.

[27] See, e.g., the article "What is Thought?" by V. Mesheryakov, in QUANTUM, July/August 2001, in which he proposes the existence of an interaction between the brain and elastic waves, with wavelength about one nanometer. He also makes the conjecture that an analytic study of the forces that act between molecules, atom complexes et al. will possibly provide an answer to the question, "What is thought?" Similarly, the mathematical physicist Roger Penrose proposes in his book, "Shadows of the Mind", that the higher synthetic abilities of the mind are due to quantum phenomena of instant mutual influence of various parts of the brain, which happen at the level of microtubules. These are minute tunnels of protein that serve as some kind of skeleton for the cells.

More specifically, cognitive psychologists have tried, or try still, to describe the mind as a system based on a network of logical type concepts, i.e. logical categories. They think that, in this way, logical thinking can be reproduced, which they consider a central part of thinking.

However, those who think in this way overlook that a system of any kind, which is based on logical categories, can never be gradually created and stay functional during the period of its formation. Thus, they do not take in account the obvious, that young children form their conceptual system gradually, during a period that lasts at least until their adolescence and that they are able to think, albeit possibly awkwardly, during all this time (see the research of Jean Piaget and more recent developmental psychologists, like Tom Bower, Margaret Donaldson, Riley & Trabasso and others).

The categories of logical concepts of any kind have a grave disadvantage: they are interconnected and interrelated. The concept "bird" is a determining element of the concept "wing", but also the concept "wing" is a determining element of the concept "bird". Each of these concepts is a part of the logical definition of the other one and cannot precede it. We cannot define what a wing is without mentioning that it is an anatomical part of a bird, but we also cannot define what kind of animals the birds are, without mentioning that they have wings. Thus, whole categories of concepts should be formed in the child's mind simultaneously, without any obvious indication of the kind of logical elements, by means of which they could be formed.

Certainly, some cognitive psychologists assume that beneath the level of logical concepts there is a level of empirical encoding, which determines the concepts that are immediately connected with experience. Thus, physical objects, like a lemon, are defined by empirical mechanisms, empirical "memory entries", like its smell and taste. But even so, the whole system remains obscure because it is not obvious how and when the logical super-structures, the logical categories, are formed in it. This group of cognitive psychologists also cannot explain satisfactorily how the complete logical categories will suddenly appear.

Incomplete logical categories cannot ever function syllogistically because they are like computer programs, which cannot operate if even a comma or any other sign is missing from some line of the code. Logical categories like space, time, quantity and number must be fully formed in order to be used syllogistically. Above the level of empirical memory entries, again we have a level of logical concepts which mutually determine each other, without any indication of which one is formed first, like the concepts "bird" and "flying".

As an example, consider the logical definition of the word "uncle", given by Johnson-Laird: "Uncle is the brother of a parent, i.e. of the father or the mother". We could possibly assume that the word "parent" is empirically encoded in our memory, i.e. that it belongs to the level of empirical memory entries. They are the father or the mother, with whom we have close contact from the day of our birth.

But what does "brother" mean? He is another child of the parents of the parent, i.e. of the grandparents. But what does "child" mean here? It is the immediate offspring of someone. But what is an "offspring"? It is very doubtful that a young child knows what sexual life, conception and child delivery are. Apparently, he therefore does not have the possibility of understanding what "offspring" means, what "child" means and what "uncle" means.

In reality of course, a young child does not care about the word "offspring" and considers as someone's "child" any other young human creature, which has an immediate dependence (for care and protection) from this grown-up person. This person is the parent, the father or mother of that other child. He recognizes that every child has parents. But for him, parents are not the biological bearers but rather the "natural protectors" of each child, the persons under whose care the child lives.

This is already an example of an "empirical" or "interactive" category. Such categories are not defined logically but rather on the basis of immediate experience. What is more, this empirical "definition" is totally wrong from a logical point of view. The caretaker of a child may not be his or her immediate biological parent. Nevertheless, this empirical "definition" does not hinder in any way a, not always but to some extent correct usage of the concepts "father" and "mother".

The empirical categories form the fundament of all thinking, because they are not restricted by logical definitions and are functional, even if only primitively, from the first moment of life. The logical definition of the concept "bird" as "an animal, which has wings" is not functional until what a "wing" is, is determined. However, a wing is an anatomical part of a bird, which leads us to a vicious circle of definitions.

In reality, what a young child initially understands as a "bird" is a creature that flies. I.e., this concept is initially determined or "defined" kinesthetically (by a sensori-motor schema) or interactively, in the way that we empirically observe a flying bird. Later of course, we find out that this flying is performed by means of wings and can define a bird as a winged animal. However, we also note that certain winged animals, like the chicken, do not fly, which raises the reasonable question whether they are "really" birds.

Here we see that, although empirical definitions may be imprecise and incorrect from a logical point of view, logical definitions can also lead us to contradictions. These are not logical contradictions, but contradictions to the initial determining experience.

But if the initial "definition" of concepts is empirical, based on sensori-motor interactive mechanisms, then how do the logical categories appear and why?

This happens when the child seeks wider communication and attempts to achieve social interaction. The logical categories are the means of communication in society. The initial empirical categories are based on personal experiences and may exhibit great variations from one person to another. It is, therefore, necessary to "standardize" them, in order to make communication between people possible. This gradually leads to the creation of logical categories, logical concepts or rather, logical standardization of the concepts[28].

The most important nodes of this mixed, partially empirical and partially logical, network are marked from the age of two years on, by words. These are used as names, as tags which make these nodes discernible. The words allow us to move in this vast network without being lost. They are like the street names or the demarcation of merchandise in a storehouse. By means of them we can express complex meanings by placing them in a sequence, i.e. by constructing a sentence. Thus, they allow us to build up a "mental model" of any situation (in Johnson-Laird's terms), which often allows us to recognize various, not immediately given, interrelations between the factors involved in this situation. It is experimentally observed that people, in order to make simple syllogisms, compose mental models of the situation at hand, which they mentally process or manipulate[29]. Thus, they are led to logical conclusions immediately, without the application of rules of Logic[30].

As we have seen, the network of empirical categories is a network of associative interconnections of various motor and sensory abilities of the body. These associations never seem to vanish from our memory, even after the beginning of logical categorization. This is obvious, e.g. from the fact that adults, as well as children, use interactive schemata rather than logical definitions, in order to explain the meaning of some word. Explaining what an egg is, usually they will not say that an egg is an initially unicellular organism, which is the result of a reproductive process and gradually evolves into a multi-cellular organism similar to its parents. What they are likely to say is that an egg is something edible, laid by a hen, forgetting that eggs are also laid by fish. I.e., in the mind operate two parallel and partially contradictory processes: the creation of empirical associations and logical categorization.

Based on the above hypothesis, that empirical associative connections remain permanently in the memory, we can now immediately understand why a great part of this network is unconscious. We need only take into account that the network of associative connections begins acquiring notation, by means of words, from the second year of life on. The earliest associative connections have no names. What is more, they have increasing haziness, reduced specificity, as we go backwards in time. They are hazy mental schemata, which grasp some common features of wider sets of empirical data. Thus, while they are important for the formation of concepts, especially as we use them in internal (mental) dialog, they are not easily discernible and become ever less discernible as we go to earlier stages.

---

[28] According to Piaget, the complete logical handling, especially of the basic concepts of space, time, quantity and number, is accomplished around the age of 12 years, with the concept of time developing last.

[29] See experimental results of P. C. Wason and P. N. Johnson-Laird, which are described in the second one's book: Mental Models, Cambridge U.P., 1983 or similar results of C. A. Riley and T. Trabasso, which are described in the book: Geoffrey Brown – Charles Desforges: Piaget's Theory, Routledge & Kegan Paul, London, 1979, p. 59.

[30] The logical standardization of the concepts seems to be based exactly on such mental models. The gradual assimilation of ever more complex "mental models", related to some concept, allows the progressively better handling of various logical aspects of the concept.

They also play another important part. They allow us to find common elements, common features in concepts, which seem logically independent. These preliminary stages of concepts, which are called "preconcepts", constitute non-specific recognition mechanisms for common empirical features of various concepts, which are not easily discernible or totally indiscernible, because they are too hazy.

Note that a part of the very early empirical content of various logically unconnected concepts may be common among them. All associative connections are stages of concept formation, which reduce the initial very hazy and unspecific conceptual structures to more specific ones.

For instance, the concepts "water glass" and "spoon" have the common preconceptual content [container]. The brackets denote here a common perceptual schema, a perceptual mechanism and not any logical definition of the concept "container". No matter how different a water glass and a spoon are, we know that they can be used to contain a quantity of liquid. Similarly, a young child understands that he can use as a container even a walnut shell. But how does he know that? Does he make logical analysis like: "The shell of the walnut is something hollow and hollow objects can be used as containers of liquids?"

It is rather improbable that he makes such syllogisms e.g. at the age of 1.5 years in which, according to Piaget, the stage of "symbolical thinking" begins. This is the stage during which some objects begin to be used as symbols, as substitutes of others. What is more, later developmental psychologists believe that this stage begins much earlier even.

What a young child does is rather that he uses the same unspecific perceptual mechanism for the recognition, the mental grasping of all these objects. Thus, it is immediately obvious that a walnut shell can be used as a container, without any logical processing. The perceptual mechanism [container] is simply a preconcept of all these objects, an unspecific perceptual mechanism, which grasps anything that can play the role of a container. I.e., it recognizes automatically a common content in concepts, which superficially are very different, like a thimble and the calyx of a flower.

Such a perceptual mechanism does not ever reach a final stage but is continuously developing. The concepts give the impression that they have a final form, because their use in discourse imposes their strict standardization. However, this standardization is only one of their components. Their empirical-associative content continues developing as long as we live. For instance, for a young child, a chair is a big object on which he may climb with some risk. For a young adult it is simply a piece of furniture on which he may rest for a while. For an old man it means restfulness or aching bones, depending on how comfortable it is. Similarly, abstract concepts like "love" and "freedom" have a different associative-empirical content in every period of life.

The existence of such a vast network of preconceptual associations also allows us to explain, in a way, the phenomenon called "inspiration" or "intuition" of a scientist. After a usually long lasting, persistent and futile search for the solution of a problem, many scientists have a sudden inspiration, which sheds light on the problem from a very different and unexpected point of view, leading to its solution. How does this happen?

Logically unconnected concepts may have common preconceptual content, common preconceptual aspects and possibilities. The realization, the recognition of this preconceptual content or of preconceptual possibilities is possibly what we often call the "intuition" of a scientist or intuitive thought.

Intuition is so mysterious a function of the mind exactly because it is a mental search in previous, more hazy and unspecific stages of the concepts. The only way to bring some important feature to consciousness, from that depth, is the use of symbols and pantomime, which stress this feature. There is no ready-made linguistic encoding at that early stage. This is possibly the explanation why symbolic and metaphoric representations play such an important role in creative thought processes which do not refer to traumatic experiences.

As an example of this phenomenon, we may consider the discovery of the ring structure of the benzene molecule by Augustus Kekule, a German chemist of the 19th century. For a long time, he was trying to find out how the atoms of Hydrogen, Oxygen and Carbon are connected in the molecule of this compound. No matter what he tried, some chemical bonds remained free. Then, while he had fallen in reverie before the fireplace, he started dreaming chains of atoms, which were dancing and twisting like snakes. Suddenly, one of these snakes bit its tail and formed a ring that was whirling around. He woke up immediately, recognizing the significance of this new possibility and worked on it during the whole night.

Here we see the Subconscious or Unconscious projecting symbolically, in the form of dancing chains or snakes, the possibility of a non-linear arrangement of atoms in the molecule of some compound. The geometric structure of the molecule is stressed as an important factor for the existence of properties that do not exist in other compounds with the same quantitative consistence, the same number of atoms of Hydrogen, Oxygen and Carbon. Something that may have helped Kekule's intuition in this instance was the fact that he had initially started studying Architecture, before turning to Chemistry. He had, therefore, a more highly developed sense for spatial relations and forms than other chemists.

The existence of the two systems of empirical and logical categorization also explains the creativity of thought. The two systems are, by their own nature, incompatible. Empirical categories can be partially described by logical categories, but their content can never be exhausted in this way. I.e., they have a relationship similar to that of the diagonal and the side of a square, whose lengths have a ratio of $\sqrt{2}$. This number is irrational and cannot be written as a quotient of two integers. However, it can be considered as a decimal number with infinite non-repetitive decimal digits: 1,414213562…. In the same way, empirical categories can have a logical description, but only approximately. Their content cannot be logically exhausted. They are carriers of infinite logical information.

Finally, an element of the mind's mechanism that must not be omitted is the role of the inborn drives, i.e. of the instincts. The instincts are both the moving power of the mechanism of thought, which make it operate, and its steering rod, which determines in which direction it will move. They are filters of incoming information, which choose what will be attended to and further processed and what will be ignored.

This very short sketch of the mind mechanism shows that this system is not very complex in its structure. However, it has great plasticity, so that it is doubtful whether thinking machines can ever be built. In order to construct them, we must equip them with a great variety of "instincts" or "drives" and allow them to form empirical associations, interacting with an environment rich in physical and social experiences. The result would be something like a primitive man with sentimental disorders and bad upbringing and education. The most important thing is, however, that such a structure will intrinsically have a free will and won't be fully controllable. It will not be an obedient machine.

11.2 Direct answers to certain fundamental questions

Since we have described here various mechanisms that participate in the phenomenon of "cognition" it would be useful to give direct answers to the question: what is the nature, the character, of certain cognitive phenomena. Therefore, we will discuss how some fundamental questions may /can be answered on the basis of what we have said.

1. What is "meaning"?

Here we should distinguish the "meaning" of a sentence from the "meaning" of an action or an event.

a) The meaning of an event is the "perception" (analysis, recognition and storing) of the event through sensori-motor "schemata"[31], which have already been formed during the development of our mental system (the trace it leaves in our mental system), as well as its further evaluation by the mechanisms of the motives. For instance, we often perceive/become aware of an approaching danger without a linguistic description and analysis of the corresponding experience.

b) The "meaning" of a sentence is given by the composition of a "mental model", based on the words being used in the sentence and its grammatical structure. The grammar determines the relations between the words and thus provides the "macroscopic" model. On the other hand, the words activate the standardized meanings that have been gradually formed in our mind, on the basis of interpersonal communication. However, underneath these standardized meanings there is the evaluation by means of deeper preconceptual layers, i.e. a complex system of perceptual "schemata".

---

[31] Here we call, "schemata", stereotyped mechanisms of action-reaction, i.e. interaction, which we have developed on the basis of our experiences up to now.

The understanding of a sentence is a process of assimilating meaning. Therefore, it is somewhat like the process of assimilating food by an organism. First comes the "chewing", the division into smaller pieces of meaning, which is achieved by grammatical and syntactical analysis. Each piece is now being interpreted, first on the basis of the standardized or typified meaning attributed to each word during the interpersonal communication. Such a typified meaning includes some typical ways of using the word, some stereotypical "models" connected with the way it is used. In this way, an initial model of each sentence is composed. However, many of its parts are still not "digested", i.e. not fully understood. Initially, the typical model that accompanies the sentence is activated. However, its deeper understanding is not restricted to the typical model and may also require the search for its associative content, its preconceptual content of the hazier and nameless perceptual "schemata", which constitute the substratum of each concept. I.e., just as it happens during the acquirement of food, initially many elements of meaning are missed/escape: they are not thoroughly "chewed" and "digested". We do not always perform a deep preconceptual investigation/analysis of each concept but restrict ourselves to certain of its features, which are sufficient for an understanding of the sentence, a composition of its semantic model. A deeper preconceptual analysis of a concept is performed only if a deeper understanding of what is being said becomes necessary. This deeper investigation often produces the intuitive inspirations.

2. What is the relationship between the conceptual system and the, so called, "declarative" and "non-declarative" knowledge?

First of all, we must remind the reader that neurologists and psychologists today distinguish between two kinds of acquired knowledge: the "declarative" or "explicit" and the "non-declarative" or "implicit". The first one can be expressed linguistically and is often acquired only after a single experience of exchange of information. The second one is the knowledge that accompanies acquired dexterity, for instance, the ability to use a bicycle.

This kind of knowledge requires many trials in order to be acquired and cannot be expressed with words (for instance, we cannot teach somebody how to keep his balance on a bicycle only by means of a linguistic description). In contrast to the "declarative", the "non-declarative" knowledge is also extremely hardly erased from the brain, in case of an injury of the brain or a stroke.

According to the view presented here, all knowledge is initially "non-declarative". The preconceptual storing of experiences in the mind is basically implicit encoding, as well as recording of knowledge.

The non-declarative knowledge is interactive knowledge. But that is exactly the kind of knowledge we acquire gradually, with respect to how we can observe or handle something. These interactive "schemata" (the stereotypical mechanisms of action or reaction to external stimuli) gradually acquire "tags", i.e. names, due to the social interaction and these names are gradually standardized in our mind. They are gradually connected with certain stereotypical "schemata" for perceiving and forming corresponding stimuli, i.e. with certain models of the way a concept can be used in communication. For instance, we gradually learn that the concepts of time imply a certain serial order of the events and certain duration of each one, no matter how we experience them internally.

Thus, the concepts have an interactive part and a standardized part, which is connected to the way that the name accompanying them is used.

3. What is the form of the interactive knowledge stored in a concept?

For instance, the concept of "space" is acquired interactively. In order to see this we need only ask ourselves: how do we recognize the stability of the shape of certain objects?

Many of the objects with fixed shape around us change their appearance because they are moving. Before we acquire experience with them, we cannot know the cause of the change of their appearance. For instance, why does a balloon seem to become smaller? Is it somehow deflated, has it been moved farther away or does it simply have the property of growing smaller, like a snowball in the sun? When we first become acquainted with the world, we do not know which objects have a fixed shape, nor what is the cause of a possible change of their shape.

According to Henri Poincaré, we recognize that some objects have a fixed shape, even if their appearance has temporarily changed, if we are able to restore their initial appearance by adjusting our own position relative to them. In this way, we gradually build up the sensation of a three dimensional space and the concept of "space" in general. Space initially refers to the fixed and invariant elements of a landscape and to the positions that other objects can have with respect to these invariant elements.

However, we must add here that the concept of the invariance of the form of many objects is not stored in memory only as an invariance of appearance from the same relative position, but also as invariance of the way we can handle them, hold them, feel them etc.

In the same way, we store the concepts "water glass", "cup" and "nutshell", partially interactively. These three concepts have a common "preconceptual" perception and handling schema, which could be described by means of the term [receptacle]. This term does not signify here some linguistic definition of what is a receptacle, but rather a sensori-motor schema of handling-perceiving all kinds of hollow objects.

By experience we know that we can, e.g. put our finger in them and that, as we go deeper, we will meet again some kind of shell. The concepts "water glass", "cup" etc. are then further differentiated (distinguished from each other) by means of additional interactive schemata, which tell us, for instance, how we can hold a water glass or a cup (the last one has a handle).

Thus, each concept has such a "non-declarative" interactive content, which we call "preconceptual", because many apparently or logically different concepts (e.g. "water glass" and "nutshell") have a background of common perceptual schemata. As we go to earlier stages of formation of the conceptual system, these perceptual schemata are increasingly fuzzy and non-specific, i.e. they cover a wider spectrum of specialized concepts. Such schemata are "non-declarative", because they have no name and they remain unconscious precisely due to their lack of specificity, i.e. because they cover a great spectrum of different specific concepts.

An indication of the fact that concepts have a "non-declarative" content is the fact that we often have inspirations without being able to describe with words how we have reached them. An explanation of our inability to describe linguistically how we have reached an inspiration is exactly because we have reached it not by using some kind of logical-syntactical processing, but empirical-interactive processing, i.e. an investigation of its deeper preconceptual content. Such an inspiration cannot be described linguistically, because, already when it originates, it does not have the form of a syntactic structure of simple symbols, e.g. words.

4. How is the standardized "meaning" of a word acquired?

The meaning, especially of abstract words, is gradually acquired by means of interactive experience and social exchange (communication). For instance, we learn to distinguish movements from deformations of objects empirically, "sensori-motorically". Also, by means of the experience of natural changes, we learn certain properties of the concepts of space, time, and quantity etc.[32]. Experience teaches us that the appearance of the surrounding space may change even when we are not there to observe it. Thus we learn that "time" keeps on "flowing" or "running" (passing) in some place while we are absent. It may also "run" in the place where we are, while we sleep. We understand the passing of time by means of the experience of outward or inward changes. We also gradually learn how to correlate the chronological ordering of our own experiences to the corresponding ordering of other events, which do not happen directly to us. Similarly, by means of experience we gradually acquire the concepts that are related to duration.

Here we must note that many aspects of the concepts exist also in animals. They simply lack the linguistic notation and the standardization accompanying it. For instance, how does an animal find its way in the forest? Obviously it has formed some concept of "space" (extension, depth, relative position of the objects and invariance or variability of their forms). Orientation in a complex environment cannot happen coincidentally or automatically.

[32] Concepts like that of time are not unique, but consist of a whole complex.

Otherwise, the animal does not survive. The animals have thus a highly developed concept of space (although the concept of time may be less developed, since they do not seem to have a sense of the farther past and the farther future). They simply lack the system of notation which allows the composition of more complex "meanings", the words.

5. How many different aspects does a concept have?

In spite of its apparent unity, it often has many, not directly interconnected properties or aspects. These properties are partially acquired independently from each other and are gradually unified in a unique model of usage of the concept. In order to make this understandable we present here some examples:

<u>Concepts of time</u>: The concept of time is apparently one, yet it constitutes a composition of many partial features, which usually have a linguistic expression. Thus, we gradually learn to distinguish:

(1) The temporal order of events:

a) The instantaneous order of events (past, present or future): Initially we perceive empirically the present, the "now", which is related to our short term memory and partially the past, the "just before now", to the extent that we maintain memories. We distinguish the future as a separate category when we begin to plan future actions.

b) The daily order of events (yesterday, today, tomorrow): All our activities are dominated by the distinction: day-night (action-rest, activation-sleep). This discernment carries the distinction, "before-now-after", over to daily periods and thus we have the concepts "yesterday-today-tomorrow". Later on we form the concepts "day before yesterday" and "day after tomorrow".

c) The yearly order of events (season, year): The periodic repetition of seasons leads to the distinction of seasons and years.

(2) The temporal correlation of events (whether they happen simultaneously or not): The ordering of external events within the frame of events that immediately concern us, leads to the recognition of the fact that certain events happen simultaneously.

(3) The temporal coherence: The coherence of temporal moments, the existence of a continuum of time, the "flow" of time, constitutes a silent admission without any linguistic expression.

All these aspects of the concept "time" are gradually assimilated during a period of many years.

Concepts of space: Similarly, we can distinguish different concepts that determine the understanding of space:

The relative position with respect to us (here-there)

(1) The relative magnitude with respect to us (large-small)

(2) The distinction of a change of appearance due to movement, from that which is due to distortion.

The gradual recognition of the fact that certain changes of appearance are due to a change of position and not to the deformation of an object, also make clear the fact that there are different outlooks, i.e. appearances from some other position.

Concepts of magnitude, number, quantity: The distinction of concepts of magnitude (large-small, much-little, slow-fast) is mainly due to our physical abilities: great, big or large is whatever requires a great effort on our part to perceive or hold; a great length exceeds the opening of my arms, a great distance requires some time and effort to cover it, a large quantity does not fit in my hands or in the receptacles I usually use, while a long time is connected with a delay of the fulfillment of my wishes or expectations.

There are also the comparative properties (equal, more, less) as well as the concept of conservation of quantity, independently of the form of the heap (with respect to a quantity of solid objects), the form of the receptacle (with respect to fluids) or the length, when we ply flexible solids.

Concepts of physical form: solid (it does not change form when I touch it) or liquid (it escapes my grip).

Many aspects of these concepts are "stored" in my memory sensori-motorically. I.e., with growing experience we build up perceptual "schemata". Subsequently, additional properties are stored, which result from social interaction. For instance, the view/contention that there is coherence of time, a uniform flow of moments of time, is due to the fact that we order our own experiences, as well as those of other people, in a unique arrangement.

If I am living alone, I register primarily the current temporal ordering (before-after) but, as the events move further to the past, they tend to get confused. For instance, I do not remember with what temporal ordering I have made various travels. My personal time is, therefore, fragmentary and consists of only certain important memories, often without a clear temporal ordering (which event preceded some other). However, the necessity to arrange the personal time of other people together with mine in a unique system creates the concept of a continuum of time. I am obliged to arrange the memories in a unique temporal ordering in order to be able to reach an understanding with others.

Similarly, the admission of the fact that certain repetitive phenomena have a specific duration is mainly due to social contact, to the necessity to adjust our own views to those of other people. Living alone, I may have a very subjective sense of duration, depending on whether some experience (e.g. a travel) is interesting or boring for me. Upsetting events, like an earthquake, seem to last for hours and not merely seconds. Interesting events seem to have a short duration (we say: "I didn't notice how the time went by"). Boring events seem to last for ever (for instance, standing in a queue).

The concept of conservation of quantity is also purely subjective. We understand that the quantity of a liquid does not change during its transfusion into another vessel if we restore it in its initial vessel, in which case we note that its level is the same as before. If all liquids tended to evaporate relatively quickly, we would not have the sense that the quantity of a liquid is preserved when we transfer it from one vessel to another. After some transfusions of the liquid back and forth between the two vessels, we would note that the liquid constantly diminishes and finally disappears.

The concepts of a unique space are also partially due to the necessity of social adjustment. We gradually become aware of the fact that in addition to our personal space, the space which we perceive and supervise immediately, other spaces must also exist. Otherwise where do other people go when they leave our environment?

The various spaces become gradually unified in our mind, as we get the opportunity to explore them, extending our personal space. Subsequently, we simply accept as a fact that there are also other spaces, when we hear other people speak of them and that these spaces have a certain relative position with respect to our own space. For instance, how do we know that America exists if we have never visited it? How do we know that the Earth is spherical? How do we know that, when we are in Europe, Australia is in some sense beneath our feet (in a diametrically opposite position to us)?

Therefore, we gradually learn certain logical properties of the concepts, certain stereotypical schemata for using them, certain micro-models, by means of which we put together the more complex models carried over to us by a series of written or oral sentences.

6. What is an "instinct"?

It is a (usually evolving) control mechanism of our reactions to outward changes. This mechanism may be electrochemical (neural system) or chemical (hormonal system). We might simulate such mechanisms by means of electromechanical or electronic systems, but we should not underestimate their complexity. There are not merely five or ten such mechanisms, as some people believe, but dozens, maybe hundreds, organized in hierarchies which lead from one reaction to another, while these hierarchies compete with each other.

7. What is Logic?

Logic is the art of consistent and clear (not confusing or contradictory) grasping and transmission of information. Logic is imposed on us in two ways:

a) Partially due to the necessity of consistent conception of our environment, the necessity of internal consistency and coherence of our mental models.

b) Partially due to the necessity of standardizing the meaning or part of the meaning of words, so that they can serve as a common means of exchange of meaning, a common "currency".

The first kind of "logic" exists even in animals. It is not based on rules but on the gradual accommodation of the sensori-motor "schemata" to different, constantly occurring outward situations. The second kind of "logic" is based on rules, but its main part is invisible. There is a vast number of stereotypical "properties" of a word-concept, which we assimilate with some effort by means of a process of trial and error, especially during the first decade of our life (see question 5). On a second level, this "logic" includes syntactic rules and rules for composing new sentences, i.e. rules for logical deduction from previous ones.

A Short Review

In this chapter we have described the integrated view of the mental system, as it results from the detailed investigation of its various aspects in the previous chapters. Here we have developed with greater coherence the central position of this book, that, in order to understand how cognition operates, we need not discover as yet unknown molecular mechanisms or other complicated physical phenomena in the interior of the neural cells, but only understand more in depth the nature of the concepts themselves. More specifically, the analysis of how the conceptual system is gradually built up, in parallel to the neural system, has shown that in the functioning of thinking, two different mechanisms are involved: the associative acquirement of the empirical features of the phenomena in our environment and their gradual, logical evaluation. These are two, essentially incompatible, but cooperating mechanisms and the study of how they cooperate has provided us with some reasonable explanations for phenomena like the existence of intuition (the function that produces unexpected insights or inspirations), as well as the existence of a fundamentally unconscious part of the mind.

In the second part of this chapter, we have investigated how certain fundamental questions about the cognitive system can be directly answered within the frame of the present position. Thus, we have investigated what could be called "meaning", how many different aspects a concept may have, the nature of instincts, the nature of logic and the relation of the conceptual system to the, so called, "declarative" and "non-declarative" memories.

BIBLIOGRAPHY

Poincarè, Henri: The Value of Science (first edition 1905).

Poincarè, Henri: Science and Hypothesis, Dover, N.Y., 1952.

Appendix 1.1

Cantor's concept of the power of sets. Comparison between the number of continuous real functions, that of real numbers and that of natural numbers

Cantor's great discovery was that all infinite sets are not equally numerous. Although this may seem strange, since one cannot imagine more than an infinite number of objects, there is a simple way to determine whether the elements of two infinite sets are equally numerous or not. One needs only establish a one-to-one correspondence between the elements of the two sets, as was done already by Galileo Galilei. If this is possible, the sets are equally numerous and we say that they have the same power (we might possibly use the terms "numerocity" or "multiplicity", but the term "power" is traditionally used). If, in the attempt to establish a correspondence, some elements of one set remain without a counterpart, we say that this set has greater power (i.e., is more numerous) than the other one.

Using this idea of trying to establish one-to-one correspondences, first we can prove that the set of real numbers is more numerous than the set of natural numbers and second, that the set of all continuous functions is more numerous (has greater power in the sense of Cantor) than the set of all real numbers.

We see that the set **R** of real numbers or even the set **R**[0,1] of those real numbers that lie between 0 and 1, has greater power than the set of all natural numbers **N,** using the, so called, diagonal argument of Cantor:

Suppose that there is a one-to-one correspondence between the natural numbers and the set **R**[0,1]. Then we can arrange the elements of **R**[0,1] (in decimal form) according to this correspondence:

$$1 \leftrightarrow 0.a_{11}\ a_{12}\ a_{13}\ a_{14}\ \ldots$$
$$2 \leftrightarrow 0.a_{21}\ a_{22}\ a_{23}\ a_{24}\ \ldots$$
$$\ldots\ldots$$
$$k \leftrightarrow 0.\ a_{k1}\ a_{k2}\ a_{k3}\ a_{k4}\ldots$$

where a_{ki} is an integer between 0 and 9, representing the i-th digit of the k-th rational number in this ordering. However, in this correspondence there are many elements of **R**[0,1] missing. For instance, any element:

$$0.b_1\ b_2\ b_3\ b_4\ \ldots.$$

is not contained in the above table, provided that the digit b_1 is different from a_{11}, b_2 different from a_{22}, b_3 different from a_{33}, and generally b_k different from a_{kk}.

In order to see that the set of all real functions has greater power than the set of real numbers, we will show that even real functions that assume only the values 0 or 1 are more numerous than the set of real numbers.

First, we note that the set of these real functions is at least as numerous as that of real numbers, since to each real number, r, we can correspond the function $f_r(x)$ with the value 1 for x = r and the value 0 at every other point of the real line.

If the two sets are equally numerous, then any other such function must belong to this ordering. But let us consider the function:

$$g(x) = 1 - f_x(x)$$

Suppose that it corresponds to the real number b. This means that g(x) = $f_b(x)$, i.e.,

$$f_b(x) = 1 - f_x(x)$$

The value of this function for x = b should be 0 or 1, since $f_x(x)$ can only have the values 0 or 1. But letting x equal b in the above equality we obtain:

$$f_b(b) = 1 - f_b(b)$$

i.e., $f_b(b) = 1/2$, which is impossible. This proves that g(x) does not belong in this ordering and therefore, the set of functions on **R**, with values only 0 or 1, is more numerous than **R**.

For an extensive and easily readable introduction to the Theory of Sets read the book: Vilenkin, N. Ya.: Stories about Sets, Academic Press, New York, London, 1968 (reprinted many times).

Appendix 1.2

A simplified description of Gödel's proof that there is an infinite number of undecidable arithmetical propositions

Gödel's derivation of his seminal results is based on a similar line of argumentation as the derivation of the paradox of Jules Richard (see **Chapter 1**). However, in order to be able to prove his results beyond any doubt, he had to first make a careful formalization of the axiomatic system and the derivation rules of arithmetic (Number Theory), in order to avoid the ambiguities of everyday language. He also had to distinguish very carefully between arithmetic and metaarithmetic, the system that studies arithmetic propositions, in order to avoid paradoxes like that of Jules Richard.

Gödel's ingenious idea was to make arithmetic talk about itself. He achieved this by mirroring metaarithmetical statements into statements of arithmetic. Thus, certain arithmetical formulas could be interpreted metaarithmetically. How did he do this? Simply by introducing a numerical coding for all arithmetical formulas. This is not difficult to understand and can be achieved in an infinite number of ways.

For instance, as we know, all information inputted and processed in a computer, text as well as formulas, is represented by a binary code, i.e. by numbers written in a binary form. Thus, a metaarithmetic statement about certain arithmetic formulas can be reformulated as a statement about numbers, the code numbers of the arithmetic formulas under consideration.

But let us describe directly such a correspondence between formulas and natural numbers: first we let every symbol used in our formalization of arithmetic correspond uniquely to some positive integer, containing no zeros (because we want to reserve zero for a special purpose). We can introduce, e.g. the following correspondence between the symbols of the first line to the numbers staying just below them in the second line:

+	-	*	/	(	)	>	<	=	1	2	...
1	2	3	4	5	6	7	8	9	11	12	...

8	9	0	→	∃	∀	~	A	b	...	s	..
18	19	21	22	23	24	25	31	32	...	51	..

Here "→" is a symbol for the implication "if... then", "∃" means "there is", "∀" means "for all" and "~" is the symbol of negation. Thus, e.g. "A → B" means "If A then B", "(∃x)(x>y)" means "There is an x that exceeds y", "(∀x)(sx = x+1)" means "For all x the successor sx of x equals x+1" and "~A" means “A is not valid”. E.g. “(∀x)~(sx = 0)” (“There is no natural number x, whose successor is 0”).

Note that in the formal axiomatic system of arithmetic all natural numbers are at first written as successors of zero. Thus, 1 = s0, 2 = ss0, 3 =sss0, 4 = ssss0, 5 = sssss0 etc., where “s” is used as a symbol of the succession operation: sx = x+1 and should not be used as a name of a variable like the other letters. This is done simply because, in a strict sense, when we start producing the whole system of arithmetic theorems, the decimal system is not yet given but must first be introduced (defined).

However, in order to simplify things, we shall assume here that we have proceeded with the development of arithmetic so far that we already have the decimal system and can use it in the usual manner.

In order to construct the number that corresponds to a formula, we simply replace each symbol by its code number and use zero as a separator of these code numbers, in order to avoid ambiguities. A sequence of formulas will be coded by a string consisting of the code numbers of the individual formulas, separated by double zeros.

Consider, for instance, the formula:

$$(\forall a) \sim(sa = 0) \rightarrow \sim(s0 = 0),$$

which will be denoted by F1. This formula simply states that, since zero is the successor of no integer, the successor of 0 cannot be equal to 0. The first part of this formula is actually one of Peano’s axioms for the natural numbers. The second is a valid conclusion derived from this axiom, obtained by replacing a specific integer for a in this axiom.

According to the above coding, this formula corresponds to the number:

$$<F_1> = 502406025050510310906022025050510210902106$$

while its second part, the formula F_2: ~(s0 = 0), has the code number:

$$<F_2> = 25050510210902106$$

Here we use the notation <F> for the code number of the arithmetic formula F (its Gödel-number, as it is usually called).

The metamathematical proposition that the above formula F_1 is a derivation, a proof of the formula F_2: ~(s0 = 0), now corresponds to the purely arithmetic proposition: “The number $<F_1> - <F_2>$ is divisible by 10^{17} and $(<F_1> - <F_2>)/10^{18} - 22$ is divisible by10^3 .” More precisely, this arithmetic proposition corresponds to the metamathematical proposition: “F_2 is the last part of the formula F_1, preceded in it by an implication arrow”. But this means that F_1 is a formal derivation of F_2.

The first relation in the arithmetic proposition is true, since $<F_2>$ has 16 digits and there is an additional zero before $<F_2>$ in the above string $<F_1>$, while the second expresses the fact that in the above formula F_1, before formula F_2 there is the implication arrow, which has code number 22. Since this last proposition is purely arithmetical, it can be written as a formula of the formal system of arithmetic.

Any proposition depending on free variables is called a predicate. Thus, the proposition: "The formula with code number m is a proof for the formula with code number n", with arbitrary integer variables m, n, is a metamathematical predicate. It becomes a true or false proposition if we insert specific values for the variables m, n.

Although discovering a proof may require "intuition", verifying a proof is an automatic, purely mechanical process, if a system of axioms has been fully formalized so that no tacit assumptions can creep into the supposed proof. Gödel was therefore able to show that the above metamathematical predicate corresponds to certain arithmetic relations between the numbers m and n, which can be formalized, i.e. expressed as a formula within the formal system of arithmetic. Since this formula still depends on parameters, it is a formal (arithmetic and not meta-arithmetic) predicate, which we denote by "PROV(m,n)".

Thus, the formula PROV(<A>,<B>) is provable in the formal system of arithmetic when A consists a formal proof of the formula B, in the same system and the negation ~PROV(<A>,<B>) is provable if A is not a formal proof of B.

In order to derive his results, Gödel then used an argument similar to the proof, given in **Chapter 1**, that there is no algorithm (there can be no algorithm) for deciding whether an arbitrary sentence is an algorithm for the computation of an integer-function or not.

He considered the formulas (arithmetic predicates) F(x), with one free variable x, and replaced in each one the free variable by the formula's own code number <F(x)>. He thus obtained the sentences F(<F(x)>), which have no free variables and can be true or false.

Let sub(n,n) be the code number of the sentence F(n), where n is the code number of the predicate F(x). I.e., n = <F(x)> and sub(n,n) = <F(<F(x)>)>. For any formula with a free variable the number sub(n,n) can be easily calculated. It may result in a very long string of digits but there is nothing difficult in repeatedly replacing each symbol by its corresponding code number.

Such a predicate is, for instance, $F_1(a) = (a>2)$ which has the code number:

$$<F_1(a)> = <(a>2)> = 50310701206$$

The sentence $F_1(<F_1(a)>)$ is now:

$$F_1(<F_1(a)>) = (50310701206>2)$$

and in the present case it is true. The corresponding code number sub(…) is here:

$$sub(50310701206, 50310701206) = <F_1(<F_1(a)>)> = <(50310701206>2)>$$

$$= 50150210130110210170210110120210160701206$$

The question now arises: "Is every sentence F(n) = F(<F(x)>) decidable? Can we prove or negate it, i.e. prove ~F(n)?" The answer is "No!".

In order to prove this, Gödel considered the formula with one free variable (the predicate) D(k):

$$\sim(\exists m)PROV(m,sub(k,k))$$

This is a formal arithmetic predicate, whose metamathematical interpretation is: "There is no natural number m, which is the code number of a proof of the formula with code number sub(k,k)". In other words, "There is no formal proof of the formula with code number sub(k,k)".

Then he asked the question: is the sentence (arithmetic proposition) D(<D(k)>) obtained by replacing the free variable in the predicate D(k) by its own code number, provable within the formal system of arithmetic or not?

The only possible answer is that this sentence is neither provable, nor disprovable. It can only be undecidable, since its metamathematical interpretation is that the formula with its own code number, i.e. itself, is not provable.

If a formal proof of this sentence could exist, this fact would belie its metamathematical interpretation. A formal negation of it would also lead to a contradiction because, metamathematically, it would mean that its interpretation is wrong, i.e. that it is provable.

Thus, provided that the formal axiomatic system of arithmetic is free of contradictions, the formula is undecidable. Nevertheless, by metamathematical reasoning, we know that it is true, because it must be formally not provable as its interpretation claims.

What does the conclusion that D(<D(k)>) is a true, but not provable statement of formal arithmetic mean? The formula itself is a formal arithmetical proposition referring to some property of the positive integers. Only the corresponding metaarithmetical expression has the strange content that it is formally not provable.

How can a mathematical formula be true but not provable? A nice example of how such a result is to be understood is given by Mark Kac and Stanislaw RID01224224224OTI pp. 135-136]. They consider the formula:

$$1 + 2 + 3 + ... + n = n(n + 1)/2,$$

which gives the sum of the first n natural numbers.

This formula is easy to discover and verify for any natural number n, by calling the above sum S(n) and writing it once in ascending and once in descending order:

$$S(n) = 1 + 2 + 3 + ... + n$$

$$S(n) = n + (n-1) + (n-2) + ... + 1$$

If we sum these equations, adding the numbers which are one above the other, we have:

$$2S(n) = (n+1) + (n+1) + (n+1) +...+(n+1),$$

where on the right side there is n times the number n+1. Thus, 2S(n)=n(n+1) and the above formula follows directly. Is this argumentation a proof of the formula? Not yet!

This technique can be used in order to verify the formula for any specific n, say n = 10 or n = 100, but in order to prove the formula for ALL natural numbers, we have to make use of the axiom of induction, which is the only axiom of Number Theory that deals with the natural numbers as a whole. It simply says: if a proposition concerning an arbitrary natural number n is valid for n = 1 and from its validity for n = k (where k is an arbitrary integer) we can prove that it is also valid for n = k+1, then it is valid for all natural numbers.

What would happen in an axiom system which does not include this axiom? Obviously, we might be able to verify a formula, like the one above, for any natural number but we would not be able to prove it for all natural numbers. In a similar manner, we may be able to verify such an undecidable formula for some natural numbers in the classical axiom system of Number Theory, without being able to prove it for all of them.

Note that the whole line of argumentation of Gödel does not depend on the specific axioms used. Thus, the construction of an undecidable formula can be repeated even if we incorporate the above undecidable formula in the previous formal axiomatic system. Any sufficiently rich mathematical axiomatic system can be shown to be necessarily incomplete, provided that it is free of contradictions.

This last problem, whether an axiom system is free of contradictions or not, has also been shown by Gödel, in the case of formal arithmetic, to be unsolvable with the means provided by this axiom system alone.

Gödel’s results lead us thus to distinguish between truth and provability. What is true is not necessarily provable within a formal axiomatic system (actually it was this discovery, which he made heuristically, that led Gödel to his results as Rudy Rucker [1984, p.288] reports).

BIBLIOGRAPHY

Kac, Mark and Ulam, Stanislaw: Mathematics and Logik, Dover, New York, 1992.

Rucker, Rudy: Infinity and the Mind, Paladin, Granada Publishing, London, 1984

Appendix 3.1
Comparison to Neural Networks

In **Chapter 3** we have discussed some of the general features of the human neural system organization, which give it its unique capabilities. Here we will discuss some of the details of this system, in comparison to artificial neural networks. This will make more discernible some of the intricacies of the human nervous system. It will also reveal in what way artificial neural networks differ from natural ones and thus fall short of expectations.

Since their beginning in the nineteen fifties, there has been considerable progress in the field of Neural Networks and Associative Memories and some of them are already applied for the solution of technical problems, like speech processing (pronouncing text or transforming speech into written text), allocating resources, noise filtering, industrial scene analysis, aircraft identification, terrain matching for automatic navigation systems etc.

Here we will refer only to their general characteristics and discuss in somewhat greater detail only multilayer Neural Networks, which seem to have a greater structural similarity with the nervous system and simulate, to some extent, the pattern-recognition capabilities of the mind.

Well written introductions to the basic known Neural Networks and Associative Memories are given, e.g. by [Beale – Jackson, 1990], [Davalo-Naim,1991] or [Fausett, 1994].

1. General Properties of Neural Networks

Artificial neural networks are information storage and processing systems that have certain functioning features, similar to those of biological neural networks. Such features are, for instance:

(a) Parallel processing of information in many locations at the same time.

(b) Associative storage in the memory. New memories are stored in locations related to their content. Thus, they are easily retrievable without lengthy search. Such a system is called "content-addressable memory".

(c) Distributed storage in the memory, i.e. storage of new content distributed among many locations, so that local damage of the brain does not disrupt memories appreciably.

In spite of such similarities, none of the artificial neuron network models is a true simulation of a biological neural system. They are rather inspired by the functioning of biological systems, in trying to replicate some of their features, than actual simulations showing how neurons are cooperating, how they are interconnected and interact with each other.

We will discuss these differences after giving some technical details about certain typical artificial neural networks.

2. Technical Features of Neural Networks

Artificial neural networks consist of a large number of interconnected basic processing units, called (artificial) "neurons" or simply "units". These units are called "neurons" because they resemble natural neural cells in their functioning. However, as we shall see, the collective organization of such units does not resemble the collective organization of biological neural systems well enough.

Artificial neural networks operate in discrete time intervals, called "time steps" or "periods". This is also a feature of biological neurons, since each such neuron needs a certain recovery time, after sending a signal, before it can fire again. However, natural neurons are usually not activated all at the same time, as artificial ones often are.

Each artificial neuron is connected to other ones by means of directed communication links of varying "weight", i.e. strength. This is also in accordance to natural neural connections, which are basically unidirectional. All chemical synapses between biological neurons that have been studied, proved to be unidirectional. Information flows only in one direction, from the pre-synaptic neuron to the post-synaptic one. The "weights" simply determine what fraction of a neuron's signal will be allowed to influence the subsequent, receiving neuron and represent the information stored in the network. This is basically the role played in biological neuron connections by the synapses.

Each neuron adds up these (positive or negative) fractions of signals ("action potentials") it receives from the preceding ones, within a period, and sends a signal pulse to the subsequent ones each time the sum exceeds a certain "threshold" value.

3. Multilayer Neural Networks

A very important class of artificial neuron networks is the class of Multilayer Neural Networks. It consists of neurons placed in layers, with the neurons of each layer sending signals, not to each other, but only to some or all of the neurons of the subsequent layer.

Such networks consist usually of three layers, since a three-layer network is enough for any two-dimensional pattern recognition task. According to a theorem by Kolmogorov, it can distinguish, i.e. separate from each other, arbitrarily complex patterns in the plane, provided that it has a sufficient number of neurons in each layer (see [Fausett, 1994, pp.328-330]).

The main goal of such a network is to associate a given set of input signals or patterns to another given set of output signals or patterns. In order to achieve this, it gradually modifies its connection weights during a “learning” or “training” phase, when each of the input-output pair of patterns, which are to be associated, is repeatedly presented simultaneously to the input and output layer of nodes respectively.

During each of these presentations, all weights are somewhat modified in order to achieve a better matching of the actual output of the network with the intended output. The modification is achieved by a mathematical process called “back-propagation”. It is basically a feedback of the discrepancy between the two output patterns: the actual and the intended one.

However, instead of a “threshold value”, which will determine whether the neurons should fire or not, multilayer neural networks use a “threshold function”. This function allows, instead of an all-or-nothing decision, a gradual increase of the percentage of the outgoing signal, as the sum of the incoming signals increases. This also seems to be in accordance to the functioning of biological neurons.

This feature is important, because the (so called, “sigmoid”) threshold function allows a better communication of the output layer with the input layer. If an all-or-nothing threshold is used, then the input is masked from the output by the intermediate layer (or layers). The output layer does not have an indication of what input signals have caused the signals it presently receives and there is no possibility of subsequent adjustment of the weights in order to improve performance.

4. The Different Way of Forming Associative Neural Connections in Biological Systems – Hebb’s Associative Mechanism

Here we should note that this is not the way a biological neural network builds associations, although biological organisms certainly tend to learn from mistakes. The biological neural network does not usually have a set of intended optimal responses to incoming stimuli, but rather builds free associations between input patterns appearing simultaneously or in close succession, or exhibits inborn stereotypical reactions. It improves its performance on the basis of success or failure and of rapidity of achieving success in tasks important for survival, like finding food, and not by a meticulous comparison of the actual reaction, i.e. “output” of the system, to some ideal output, some ideal reaction.

Free associations are often governed, in biological neural systems, by a simple rule which was proposed by Donald Hebb already in 1949: the strength of the connection between two neurons is increased when they are simultaneously activated by incoming stimuli.

This kind of Hebbian associative learning is actually replicated in some sense, not in multilayer neural networks, but only in Hopfield networks, Willshaw’s associative nets and their improvement and Austin’s ADAM (Advanced Distributed Associative Memory) net. In these networks, a modification, an increase of the weights between two connected neurons happens only if the two neurons are simultaneously active.

We are going to discuss biological associative mechanisms in greater detail later, in order to see their significance for information processing. Here we will first consider the advantages and shortcomings of various artificial neural networks. This will allow a later comparison with capabilities of biological systems.

5. Advantages and Shortcomings of Various Kinds of Artificial Neural Networks

Many artificial neural networks succeed in exhibiting all or most of the initially mentioned features of the nervous system (parallelism, associative addressing, distributed storing) but they are still far from being even remotely as successful as the brain.

Of course, most of these models of memory exist only as algorithms, i.e. as computer simulations of what a real technical device of this kind would do. This limits considerably their capacity (the number of neurons), even if we tried to model inputs from only one of the senses, e.g. vision.

But their shortcomings are more fundamental than that. Although biological systems consist of many more neural units than artificial ones, the main problem is not to build devices with sufficient capacity, but to make them more efficient in various ways.

The very popular multilayer neural networks (consisting usually of three layers) are able to classify correctly "noisy" (i.e. distorted) patterns and even yet unknown (never "seen" before) patterns, with others that share the same distinguishing features. Thus they are, in some sense, able to "generalize", although they are better at "interpolating" known patterns than at "extrapolating" from them. They detect known patterns in the inputs they are given, but they cannot extend, as easily, a range of known patterns to include new ones.

Their main shortcoming is that their training set has to be fixed. The associated input and output patterns are fixed beforehand, and they require long training times. Many presentations of the patterns, which yet are to be learned, are needed until they have adjusted the connection weights well enough to respond correctly, when presented with one of these patterns.

Furthermore, they are not so good at learning new information, a new pattern, after they have been trained. The new pattern disrupts the already settled weight values. So the whole lengthy training process may have to be repeated again from start.

Comparing the brain structure with multilayer neural networks, we may note that the brain seems unnecessarily complicated. Sensory areas consist of many layers of neurons, each reaching a certain stage of processing of incoming signals. For instance, there are at least seven layers of neurons reaching from the eye to the cortex. The brain does not attempt an immediate pattern classification, i.e. recognition in two or three layers, but rather dissolves the perceived pattern into more elementary perceptual features, which are gradually organized into more complex ones. Each layer of the visual system "recognizes" (responds to) a certain kind of features and is not directly involved in recognizing the whole pattern.

Every layer of neurons in the brain is content with organizing only certain aspects of perception and leaves other aspects for the next layers. We have a succession of classification systems rather than a single three-layer system doing the whole job.

A network that has the ability to perform unsupervised learning is Kohonen's Self-organizing Network. Here, no output set to be associated with the input signals or patterns is given. The categories, into which the input signals are to be classified, are not predetermined but rather created by self-organization of the network. Thus, free associations of the input patterns into categories, having somehow common features, are possible. The network consists of only one layer of nodes (artificial neurons), with all inputs connected to all neurons and feedback connections between neighboring neurons.

Input patterns are presented to all nodes of this network at the same time and the network then proceeds to adjust the initially arbitrary weights of the interconnections of neighboring nodes. Thus, it organizes the nodes into local neighborhoods, each of which represents a class of somehow similar patterns.

However, the number of classes or categories which are to be built up has to be given and the labeling of the neighborhoods with class names has to be done externally.

In the application of automatic translation of speech into writing, built up by Kohonen, phonemes of spoken language are presented to the nodes of the network as vectors of sample values. The network gradually settles into groups of nodes representing similar phonemes and these are labeled by letters, thus building a "phoneme feature map" of the Finnish language (a similar system has been built for the Japanese language).

Another interesting group of network models are ART1 to ART3, which are based on S. Grossberg's and G. Carpenter's "Adaptive Resonance Theory". They consist of two layers of artificial neurons with strong feedforward and feedback links. The network is able to associate a set of inputs with a set of outputs in a very short training time (usually, there is only one presentation of the training data). It is also able to learn new classifications without disruption of the old ones.

However, it performs poorly with "noisy" (i.e., distorted) inputs and does not offer distributed storing, since each output node corresponds to one classification category. Destruction of an output node will eliminate the corresponding category.

Hopfield's network has a single layer of fully interconnected nodes. Input signals are applied simultaneously to all nodes and the network cycles adjusting the weights of the interconnections, until it settles to a steady state that represents its response. It cannot encode data, because it has only a single layer, but it can function as a pattern recognizer that filters out "noise", i.e. associates corrupted inputs with certain ideal forms, preliminary trained into the network. It can also be used as a topological problem solver, settling down into a "best" or "nearly best" solution.

However, the network tends to get quickly saturated, so that new information cannot be stored without disrupting, e.g. corrupting the already existing ones. A network consisting of N interconnected neurons can store no more than about N/7 "memories". After this limit, it tends to confuse them. Similar advantages and disadvantages can be also found in other network models or associative memories.

Most of them do not build free associations, but need pre-selected sets of input and output patterns, even if they are subsequently able to associate unknown input patterns to one of the categories represented by the output patterns. They are rather models of conditioned reflexes than of free exploration of the surrounding, although they do not use the modulatory association mechanism described later.

Even Kohonen's self-organizing network requires external labeling of the built up "categories" and their number is preliminarily selected. It does not evaluate incoming stimuli fully automatically.

Many of them also need extensive re-training in order to add even a single pair of associations in their repertory. Those that learn with a single pass of the data are not good in "recognizing" noisy (distorted) inputs, since they tend to classify each new pattern separately.

All these models are thus pattern recognition devices, useful for certain very specific technical applications, rather than real simulations of human or animal memory. They succeed in partially realizing parallelism, associative addressing or distributed storing, but they are still far from being equally successful as the brain.

A basic reason for this is that the brain has some important additional features. One of these is that it uses various associative mechanisms for storing memories, different from those used by artificial neural networks. In order to see their significance, let us briefly discuss the various biological associative mechanisms.

6. Associative Mechanisms in Biological Systems

Speaking about associations in a human or animal neural system, we must distinguish between two levels: macroscopic associations and microscopic or neural associative connections. Macroscopic associations, those usually considered by psychologists, are of course due to microscopic ones but they can have different features because they are the result of a collective functioning of the former.

Macroscopic associations require only persistent or repeated proximity in time, or space and time. Two perceptions are linked if they repeatedly occur at the same time or almost at the same time, no matter in what order. Seeing a picture of a famous singer, we may remember his/her voice, while hearing his/her voice we may remember his/her face.

However, this is not the case at the microscopic level. A connection between two neurons is created or strengthened only if the stimuli appear either simultaneously or successively, within a very short period of time, but in a certain order.

There are more than one neural associative mechanism. Thus, we have:

(a) Hebbian Associations. These are due to repeated simultaneous activation of two neurons and are involved in connecting two or more simultaneous stimuli with each other.

The neural mechanism behind Hebbian associations seems to be Long Term Potentiation (LTP), a long lasting sensitization of a neural pathway, due to a high frequency signal of short duration.

(b) Modulatory Associations. These are due to a repeated sensitization of the initially weak connection between two neurons, by signals coming from a third neuron, the “modulatory” neuron. In this case, the synapse is strengthened only if the modulatory neuron activates the receiving neuron a short time before the actual signal from the pre-synaptic neuron comes.

Such associations are involved in learning of stimulus-response connections, i.e. sensory-motor connections, as those appearing after classical (Pavlovian) conditioning.

(c) Habituation. This is actually the opposite phenomenon to the formation of stimulus-stimulus connections, a desensitization of a neural pathway to a repeated outward signal, which has no consequences for the organism, i.e. the cancellation of a previously formed association.

The neural mechanism behind habituation is the, so called, Long Term Depression (LTD), a desensitization of a neural pathway, due to a low frequency signal of short duration.

Here we will consider the first two of these mechanisms more closely, since we are more interested in how memories are stored rather than in how they are erased.

7. Hebbian Associations

The first microscopic neural associative mechanism was proposed in 1949 by Donald Hebb. Considering two neurons, A and B, weakly interconnected so that when A fires, B does not fire, Hebb proposed that their interconnection is strengthened if they happen to fire simultaneously. If when A fires (sends out signals), B happens to also fire, excited through another neural pathway, this strengthens the synaptic connection between A and B. In this way, B can later be activated by a signal coming from A alone.

According to Hebb, when two or more neurons are repeatedly “fired in conjunction”, this repeated association in firing brings about structural changes in the neurons, which facilitate their interconnection. In this way, assemblies of cells are established in the sensory cortex, which are likely to fire in conjunction (each one can activate the others).

As perceptual experience accumulates, complex and stable patterns of firing are set up, which he called “phase sequences”. Different sensory patterns evoke different phase sequences, but their form depends also on the already existing neural associations from past experiences.

The community of phase sequences enables us, e.g. to identify the shape of newly encountered objects or judge our distance from objects, on the basis of various visual cues which we have learned to appraise through their relation to tactile and kinesthetic sensation.

Hebbian associations may combine different kinds of sensory inputs, e.g. acoustic, visual and olfactory, but also sensory inputs of the same kind, e.g. only visual. Thus, we may recall a certain place by a smell we have associated with it; but we also tend to associate the faces of people we meet only during their professional occupation with their uniform, that of a postman, an officer, a policeman, a nurse, a laboratory assistant, a worker etc. If we see these people without their uniform, we may not recognize them, although they may seem remotely familiar to us.

7.1 Long Term Potentiation (LTP) – Its Role for Storing Declarative Knowledge and for the Transition from the Short Term to the Long Term Memory

A neural mechanism that forms Hebbian associations was discovered quite late, in 1973, by Timothy Bliss and Terje Lomo. It was the, so called, “Long Term Potentiation” (LTP), a sensitization of a pathway of many interconnected neurons for a longer period of time, which is due only to a short initial excitation by a high frequency electrical signal (about one hundred pulses in a second). It does not change the whole post-synaptic neuron, making it more sensitive to any input, but only the particular synapses on the post-synaptic neuron that were involved in processing some experience.

LTP seems to be one of the basic mechanisms by means of which memories are kept in the Short Term Memory long enough to be transferred to the Long Term Memory. In particular, it seems to be important for explicit learning, the kind of learning that can be expressed in linguistic form, because explicit learning can result even from a single external experience, without any repetition. Implicit learning, improving task performance is, on the contrary, based on repetition.

By making some signal path more sensitive for a longer period of time, LTP keeps a stimulus in memory for some time, so that it can be associated with other stimuli appearing in that period. It can last for hours, days and even weeks.

Long Term Potentiation was first discovered in pathways of the midbrain’s hippocampus, which is a region involved in storing new explicit memories in the LTM; this conclusion is due to the fact that damage hinders the long term storage of new memories, although it does not interfere with old ones. Subsequently, this mechanism was also discovered in the brain’s amygdalas and in its cortex.

7.2 The Mechanism of LTP

LTP has two stages. An initial enhancement of the connection between a pre-synaptic and a post-synaptic neuron results from activation of both neurons simultaneously, i.e. by following a Hebbian mechanism. Due to this brief enhancement, the synapse is then “potentiated”, i.e. maintains the facility of the connection between the two neurons for a long time.

The mechanism of LTP is, in broad lines, the following (see [Kandel - Hawkins, 1992, pp.59-60] or [LeDoux, p. 218]):

A neuron receiving enough input signals at the same time, fires an action potential (a wave of electrical charges) along its axon[33]. The axon ends up in branches that approach many other neurons, but are separated from them by

[33] In natural neurons the information is coded in the form of frequency of electrical impulses, “action potentials”. With increasing frequency of the arrival of electric pulses, more neurotransmitter is released from a synapse and the post-synaptic potential, which activates the subsequent neuron, grows higher.

small gaps, called, synaptic clefts. These gaps are bridged during signal transmission by chemicals called neurotransmitters. They are the messengers that carry a signal, across a synapse, from one neuron to another connected with it.

Neurotransmitters released from axon terminals bind to appropriate receptors on the other side of the synapses and either result in excitation or inhibition. Excitatory transmitters increase the likelihood that the post-synaptic cell will fire and inhibitory transmitters decrease it.

The major excitatory transmitter in the brain is the amino acid glutamate. Packets of glutamate released from the axon terminal cross the synaptic gap and usually bind to a special class of glutamate receptor molecules, the AMPA receptors. When this happens, the post-synaptic cell fires impulses along its axon.

There is also another kind of glutamate receptors on the membrane of the post-synaptic cell, the NMDA receptors. These are normally blocked by magnesium, so that the glutamate reaching them has no effect. They become available to bind it only if the post-synaptic cell happens to fire, activated through some other pathway. Thus, the NMDA receptors bind glutamate only if the pre-synaptic and the post-synaptic neuron happen to fire simultaneously. In this way, they serve as coincidence detectors, as a means for forming associations between coincident stimuli.

Then, the molecular changes that happen strengthen and stabilize the connection between the pre-synaptic and post-synaptic neuron. Thus, simultaneously occurring events come to be associated as parts of the memory of an experience.

While the induction of LTP depends on the simultaneous firing of the post-synaptic cell, the maintenance of LTP is due to a different mechanism: the enhancement of transmitter release from the pre-synaptic terminal. This increase in transmitter release is only possible if some message is sent from the post-synaptic to the pre-synaptic neuron. The retrograde messenger performing this task is nitric oxide. It diffuses rapidly out of the post-synaptic cell, across the synaptic cleft, into the pre-synaptic terminal. There it enhances the transmitter release, but only if its diffusion is paired with activity in the pre-synaptic neuron.

Thus, LTP uses a combination of two independent associative, synaptic learning mechanisms: a Hebbian mechanism based on properties of NMDA receptors and a non-Hebbian mechanism of pre-synaptic facilitation during activity.

8. The Role of the Amygdalas for the Fixation of Memories – The Effect of Adrenaline

LTP is only part of the mechanism that stabilizes memories and maintains them in the Long Term Memory. Another important factor that stabilizes memories is, according to psychological and biochemical experiments, the overall evaluation, the emotional appreciation of the events. This process is significantly controlled by the two amygdalas of the brain’s hemispheres.

When the amygdala is aroused by a sudden, pleasant or unpleasant event, it turns on many bodily systems and in particular the adrenal gland, which releases adrenaline into the bloodstream.

Adrenaline has been shown to play an important role in the solidification of memory processes. For instance, rats given a shot of adrenaline, right after learning something, show an enhanced memory of the acquired learning. Similarly, emotionally arousing stories are better remembered by people than similar ones without emotional accentuation, while no such enhancement of memory is seen if a drug that blocks the effects of adrenaline has been given to the subjects.

Adrenaline release, as a consequence of emotional arousal, interacts with cerebral systems like the hippocampus and strengthens the explicit memories being created there, while adrenaline blockade prevents the memory enhancing effects of emotional arousal [LeDoux, p.212 and p.206]. Actually, adrenaline stabilizes and strengthens both implicit and explicit memories, i.e. the formation of all associations.

Here we see that biological associative learning is based on a very different adjustment mechanism than that of artificial neural networks. Improvement of performance is not based on better pattern-matching, but on an overall evaluation of the incoming information. This overall evaluation is very important for the permanent storage of new information in biological neural circuits. By enhancing LTP, a macroscopic evaluation is able to influence a microscopic learning process, like the associative connection of two neurons.

Pattern-matching associations are used by the brain only in specific instances, for specific tasks. E.g., in order to send a billiard ball into the hole, to aim at a fixed target etc. In such instances, there may preexist in our mind an optimal trajectory, which we try to replicate.

9. How the Amygdala is Aroused – Its Role in Causing Instinctive Reactions

But how does the amygdala become aroused? The amygdala receives signals both from the sensory thalamus and the sensory cortex. Signals from the sensory regions of the thalamus reach at first this system and provide already a crude image of the external world, allowing a quick “instinctive” (impulsive) reaction, if there is an indication of danger.

The amygdala receives more detailed and accurate representations somewhat later from the sensory cortex, together with non-sensory information about the general situation, which comes from the hippocampus. This allows a more careful appreciation of the situation and an eventual moderation of the initial hasty reaction which, nevertheless, may be important for survival in cases of acute danger. The amygdala is thus involved in appraising emotional meaning and triggering quick reactions.

There are various indications that the amygdala is indeed able to evaluate incoming stimuli before they are processed by the cortex. We often react to an outward stimulus before we know what exactly has caused our reaction. For instance, a quickly approaching object makes us jump aside or duck to avoid it, before we know what it is! An interesting story supporting this view is told by D. Goleman [1996, pp.19-20].

Obviously, this is done on the basis of very specific stimuli, without notice of further details. For instance, we know that instinctive reactions are triggered by very specific “releasing” signals. Thus, according to experiments of N. Tinbergen, young seagulls are moved to ask for food by any somewhat inclined, lengthy, light-colored object with a red spot near its end, since these are the outstanding features of a gull’s beak. The way they ask for food is by picking on the red spot of their parent’s beaks. They pick on a crude dummy of a gull’s beak with a red spot at the proper place, almost as often as on a precise model of a full head of a grown-up gull, while they very seldom react to the model of a head without a red spot on the beak.

Similarly, a male stickleback (a kind of fish) will attack, during the mating period, any lengthy object whose bottom is colored red, because these are the outstanding features of a rival male, while it does not react to the same dummy turned upside down, i.e. with the red colored part on the top [Daumer-Hainz, 1987, pp.38-39].

10. Modulatory Associations

The second associative neural mechanism, the “modulatory” associations, was discovered by Ladislav Tauc and Eric Kandel in 1963. They play an important part in building up conditioned reflexes.

Modulatory associations consist of strengthening a synapse between two neurons, e.g. a sensory and a motor neuron, due to the activation of a third neuron, the “modulatory” neuron, within a short time interval <u>before</u> a signal passes from the pre-synaptic sensory neuron to the post-synaptic motor neuron.

The modulatory neuron, which meets the pre-synaptic sensory neuron in the neighborhood of the strengthened synapse, is activated by other sensory neurons, which cause an immediate motor reaction without conditioning.

Repeated activation of the modulatory neuron just before the, initially ineffective, pre-synaptic sensory neuron is activated, leads to a strengthening of the synapse with the motor neuron. The transmitter release from the sensory neuron is enhanced and, after a number of repetitions of this sensitization, a motor reaction is possible just by activating the pre-synaptic sensory neuron, without any activation of the modulatory neuron.

Since this mechanism strengthens a neural connection only after a number of repetitions, it is actually involved in acquiring conditioned reflexes and other forms of implicit learning (learning that cannot be linguistically described, e.g. acquirement of aptitudes).

We should note here that the activation of the modulatory neuron must precede that of the sensory neuron by a certain critical and often narrow time interval. If the succession is reversed or the time interval is somewhat longer, then no conditioning results. However, the optimal time interval may be different for different reactions. For instance, it is 0.2 to 0.5 seconds for eye-blink conditioning, 2 to 4 seconds for licking conditioning, 10 to 20 seconds for fear conditioning and 1 to 2 hours for conditioning a taste-aversion [R. F. Thompson, 1993, p.345].

11. The Storage of Declarative and Non-declarative Memories – The Continuing Change of Neural Connections

Thus, in biological neural networks we may also have weight-modifying signals, the modulatory signals, although they do not have a form corresponding to a "back-propagation rule" that corrects the deviation from an ideal output.

Experiments with various animals show that both explicit and implicit memory storage proceed in stages. The storage of the initial information for minutes to hours, i.e. the maintenance of Short Term Memory, involves changes in the strength of already existing synaptic connections, i.e. sensitization of a whole path to further signals. However, the long term maintenance of memories, which happens at the same sites, requires something totally new: the activation of genes and the production of proteins, which lead to the growth of new connections [Kandel-Hawkins, p.60].

Stimuli that produce Long Term Memory, e.g. classical conditioning in animals, lead to an increase in the number of pre-synaptic terminals. Similar changes occur in the human hippocampus after LTP. Does that imply that our brain's anatomy changes constantly? The answer suggested by M. Merzenich's work is: "Yes!"

Michael Merzenich and his coworkers have demonstrated in monkeys that the cortical map of the hand, its representation in the sensory area of the cerebral cortex, is subject to modification based on the use of the sensory pathways. They induced a monkey to touch a rotating disk with only the three middle fingers of its hand. After several thousand disk rotations, they found out that the area in the cortex devoted to the three middle fingers was expanded at the expense to that devoted to the other fingers. Practice, therefore, leads to changes in the cortical representation of active organs.

Even stress can change associative connections. E.g., exposing subordinate male tree shrews[34] to the presence of a dominant male for a longer period of time, reduces the branching and length of dendrites (causes dendritic atrophy) in the hippocampus [LeDoux, pp.242-243].

12. A Comparison of Artificial and Natural Associations

All the above associative mechanisms do not seem to have a counterpart in artificial neural networks, except for the Hebbian associations in Hopfield and other networks, which were mentioned earlier.

In particular, the weight modification in multilayer networks is non-Hebbian, since the initial activation and the subsequent feedback from the output are not simultaneous. It is also neither "modulatory", nor of the kind of long term potentiation. What is more, "weight" modification in artificial neural networks is a form of adaptation to given patterns. It does not result from an overall evaluation of the situation, which strengthens the associations involved in the present events.

Comparing biological systems with neural networks, we may also note that in the latter there is no distinction of Short Term and Long Term Memory, which is important for maintaining only significant information. There is also no distinction between explicit and implicit memory, i.e. of declarative knowledge or skills. The existence of these mechanisms provides biological neural systems with much higher flexibility.

However, there are yet other ways in which a biological neural system differs from the most widely used artificial neural networks. Let us look at them more closely.

13. Important Features of the Brain's Organization

The most important differences of a biological neuronal network from an artificial, multilayer neural network are two:

[34] They are mouse-like, insect-eating mammals.

(a) It performs graduated feature extraction and no pattern-matching.

(b) It improves its performance on the basis of drive-fulfillment and not of accuracy of reproduction of predetermined patterns.

For instance, a stick is, for the brain, far more than a lengthy rigid object. It is mainly and primarily an extension of the arm or, sometimes, the fingers, e.g. of the forefinger.

Although the brain builds up the largest part of its associative connections on the basis of experiences acquired after birth, it is already extensively pre-connected at birth, in order to achieve certain goals quickly or adjust itself quicker to outward experiences. It has various in-built reaction and perception mechanisms.

For instance, as Wright-Taylor [p.254] note, parts of the shape perception mechanism in the visual cortex are already present at birth, since there exist cortical cells which respond maximally only to patterns of light having a particular orientation[35].

All these "pre-wired" connections serve the graduated feature extraction mentioned above. The brain does not use three, but many layers and allows each layer to extract from the input signals certain features, which are important for the processing of the next layer. Each layer may be preformed to respond maximally to a certain kind of signal, but it is also formed by experience, which leads to new connections and strengthens existing ones.

Thus, not only do the sensory and motor areas of the cortex get increasingly organized, but these areas are also steadily further interconnected by neurons which belong to the intermediary, the association areas of the cortex, which make up 80% of its whole surface.

14. The Contribution of Sensori-motor Mechanisms for Pattern Recognition and for the Development of Perceptual Abilities

We must not forget here that biological pattern recognition is not only a matter of visual scene analysis. We can recognize somebody by looking at him from behind or from afar, not by the way he looks, but by the way he moves: the rhythm of the steps, the way he sways his body when he walks or his posture. This kind of recognition is not merely visual, but rather visual-kinetic. It

[35] Smiling, laughing and crying are also inborn reaction mechanisms, as the observation of blind borne children proves (see chapter 8).

incorporates perception of movement, experience of movement and not merely of shape.

An experiment which supports this observation is described in [Gellatly-Zarate, 1998, pp.52-53]. In this experiment, light emitting diodes (LED) are placed at various parts (mainly the joints) of an actor with a blackened face, who is wearing a black suit and moves in a dark room, being videotaped. When the video is shown to people, they perceive an irregular distribution of light spots as long as the actor stands still. But, when the actor moves, they soon recognize the pattern of a moving human body and even the kind of its movements, e.g. walking, running, dancing etc. They are even able to tell whether the actor is a man or a woman.

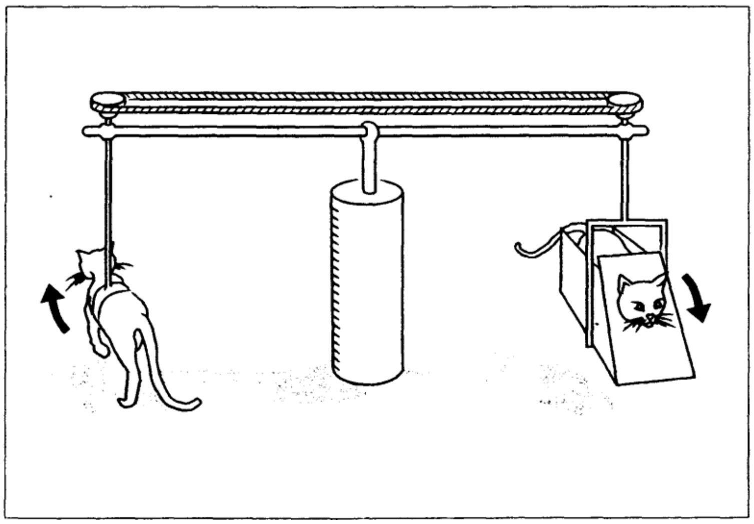


Fig. 3.1.1: The kitten carousel designed by Held and Hein in 1963.

Another interesting experiment which supports this view is based on the, so called, "kitten carousel", designed by Held and Hein in 1963. By means of it they demonstrated the importance of active, rather than passive, visual experience. Pairs of kittens, aged two to three months old, were placed in a carousel. One was placed in a carriage restricting its movements, while the other in a harness allowing it to move about and thus to turn around the carriage with the first kitten. The kittens were put in the carousel for three hours a day and were otherwise reared in darkness from their birth on. Thus, both kittens received the same amount of visual experience; but only for one (the "active" kitten) was visual experience correlated with movement. When the animals were submitted to various tests on depth perception, after about ten days of training, the "passive" kittens still showed impairment in responding correctly, while "active" kittens showed such impairment no more [Wright-Taylor, pp.263-265].

This finding strongly suggests that it is movement-produced, rather than passively received, stimulation which is important for the development of visual organization.

15. General Features of the Evolutionary Organization of Human Memory – The Functionality of Intermediate Evolutionary Stages

We should also note here that human memories are not only distributed among certain, somehow interconnected neurons, but they are also stored in many developmental stages, i.e. by gradually developing interconnections. Intermediary developmental stages are not merely parts of the encoding of certain final memories, but must consist of already functional, although initially hazy, memory structures that grow progressively more precise and specific (see **Chapter 6**).

If the memory structures were not already functional in each developmental stage, even if in a not very precise form, then the organism would be helpless and unable to sustain himself. We have called here these early memory structures “preconcepts” (**Chapter 6**), in order to indicate that they are not yet linguistically describable.

Every such developmental stage is the temporary form of a set of continuously growing interconnections, which never reach a fully fixed final form, although later they may not evolve as quickly as in their early stages. There is no end to “training time”, as Merzenich’s work has proved.

Such temporal developmental stages of the nervous system must not necessarily correspond to a spatial arrangement of different neural layers. Unlike artificial neurons, the natural ones can always grow longer projections to reach certain other neurons and build synaptic connections.

In a human or animal brain, patterns are empirically decomposed and classified by the already existing hierarchies of perceptual and kinetic mechanisms and stored as appropriate combinations of such mechanisms. Such hierarchies evolve steadily until they reach the form of logical-linguistic classes and even afterwards. The storing of a newly perceived pattern does not disrupt that of already stored ones, because it is based on connecting highly differentiated perceptual schemata and not on the indiscriminate connection of a set of neurons.

16. The Drives as a Mechanism for Evaluating Information in order to Store it Appropriately – Comparison with the Back-propagation Mechanism

The brain also has built-in drives, which are the moving force of this system. They set the goals, on the basis of which, performance is evaluated and improved (see **Chapter 8**).

The new information stored in each stage of memory development is selected, evaluated and filtered by drives, continuously pressing for "better" performance in certain directions determined by the organism's inherent goals. This is why each stage is already functional and not only the final stage, as in artificial networks.

Actually, a macroscopic drive consists of a hierarchy of specific patterns of behavior (action and reaction schemata), which are inborn in their basic features, but nevertheless evolve as the organism grows up; they also have to be developed in their specific form on the basis of experience. Some of these behavior schemata may even be shared by many drives. Thus, a kitten learns to run after a ball and catch it just for play, obeying to the, so-called, playing drive (see **Chapter 8**). But this skill, when acquired, will also be useful to a cat, driven by hunger, for catching a mouse.

We must also note that a drive is not merely a mechanism that evaluates reactions, as the back-propagation mechanism in artificial neural networks does, but it also provides action directions independent of outward signals. It is far more than a mechanism for judging the degree of pattern-matching and improving it.

The initial goals of an organism are body control, as well as control of the surrounding, through the body. A strong driving force in this endeavor is hunger, for instance, which is a collective evaluation mechanism, judging on the basis of success or failure of hunger reduction.

It is noteworthy that this evaluative mechanism does not operate merely electrically, as artificial mechanisms do, but also chemically. It translates chemical signals, e.g. a low level of blood sugar, into electrical signals of hunger in the hypothalamus. This activates various other systems, which seek food in order to still the "feeling" of hunger.

The body has more than one chemical control system (e.g., various hormonal controls), corresponding to various drives. They all play a very important role in evaluating inward and outward signals and thus form the experiences stored in the brain, just as strongly as the actual outward signals that reach it.

Success or failure in artificial neural networks is only a matter of good pattern-matching. There is no overall evaluation which may not only modify reactions, but even the drives, the evaluating mechanism. There are no macroscopic goals, like acquiring food, which evaluate and organize perception on the basis of its success in reaching these goals. The "optimal" reaction patterns in artificial networks are usually prescribed. But this makes them rigid. They are not adjusted to varying outward circumstances, like observation from a different angle or distance.

We may also note that, unlike a biological system, even such artificial networks that need only one training period for encoding a new association, cannot classify most new patterns until sufficiently many have been stored. They do not "pre-classify" in yet hazy empirical categories.

Artificial multilayer networks thus lack sufficient plasticity. Newly learned patterns disrupt the whole system because there is no distinction of stimuli into graduated empirical classes, which may maintain a more steady form.

A Short Review

In this appendix we have considered the main structural features and the properties of various artificial neural networks and we have compared them with biological neural mechanisms. We saw that the artificial neural networks partially simulate the basic features of natural ones, like the associative distributed storage in memory and the parallel processing. However, they are far less effective compared to biological neural networks, because the mechanisms for reinforcement of associations are totally different.

For instance, the very popular three-layer neural networks store, during the "learning" phase, a certain number of specific patterns as associative connections and reinforce the associative connections which store each pattern, on the basis of how much the output signals of the network differ from the pattern taught each time. Of course, this allows them to "recognize" such patterns, even distorted ones but, consequently, if even one more pattern is added to the initial set, the very time consuming process of "training" must be repeated for the whole set of patterns. Similar shortcomings are present in all other kinds of artificial neural networks.

On the contrary, biological neural networks do not aim at the storage of certain chosen forms, but each time evaluate the "responses", i.e. the reactions of the system, on the basis of success or failure in processes of immediate significance for survival. Pattern recognition is here based on mechanisms of feature analysis, related to the way we can interact with the corresponding object.

We have now considered, more closely, the biological associative mechanisms. More specifically, we have considered the Hebbian mechanism of associations, which is based on "Long Term Potentiation", as well as the mechanism of "modulatory associations". We saw that Long Term Potentiation can maintain even unique experiences in Short Term Memory, until appropriate associative connections are formed, which store them in Long Term Memory. This led to the conclusion that Long Term Potentiation is especcially important in maintaining "declarative" memories, i.e. memories that can be expressed linguistically, because these memories are usually stored after only a single experience. On the other hand, modulatory associations are primarily involved in the building of reflexes, i.e. of "non-declarative" memories, because these require many repetitions.

We have also seen that an important role for the fixation of the contents of Long Term Memory is played by the two amygdalas. They are the centers of instinctive or impulsive reactions, i.e. they make a first quick evaluation of each situation. When the amygdala is aroused, it causes the release of adrenaline by the adrenal glands. Adrenalin has the property that it enhances the long term storage of memories.

With respect to pattern recognition in man, we saw that it does is not based simply on some purely visual processes, but rather on visual-kinetic ones, i.e. sensori-motor processes. We have also seen, on the basis of the “kitten-carousel” experiment, that the perception of depth, as well as other perceptual abilities, requires sensori-motor experience and not simply visual experience, in order to be developed. These findings strengthen the view that all storage in memory happens interactively.

Considering biological neural mechanisms, we also saw that there are experimental data, showing that practice leads to changes in the cortical representation of active organs. New experiences change the neural connections that correspond even to functions as elementary as the sense of touch. This supports the view that all memory structures are steadily changing, steadily being modified on the basis of experience.

Concluding this appendix, we made a general comparison of the human memory system with artificial neural networks, emphasizing its plasticity, its flexibility and the fact that the way of evaluating new experiences plays a central role in making this system much more effective than the artificial ones.

BIBLIOGRAPHY

Beale, R. – Jackson, T.: Neural Computing. An Introduction, Institute of Physics Publishing, Bristol and Philadelphia, 1990 (last reprint 1998).

Daumer, K. – Hainz, R.: Verhaltensbiologie, Bayerischer Schulbuch-Verlag, München, 1987.

Davalo Eric – Naim, Patric: Neural Networks, Mackmillan, London 1991.

Fausett, Laurene: Fundamentals of Neural Networks, Prentice-Hall, 1994.

Gellatly, A. – Zarate, O.: Mind & Brain for Beginners, Icon Books, Cambridge, U. K., 1998.

Kandel, Eric – Hawkins, Robert D.: The Biological Basis of Learning and Individuality, Scientific American Sept. 1992, pp. 53-60.

LeDoux Joseph: The Emotional Brain, Phoenix Paperback, London, 1998.

Thompson, R. F.: The Brain, A Neuroscience Primer, Freeman & Co, New York, 1985, 1993.

Tsien Joe Z.: Building a Brainier Mouse, Scientific American, April 2000, pp.42-48.

Wright, D.S., Taylor, A.: Introducing Psychology, Penguin, 1974.

Appendix 4.1

Is Meaning Logically Decomposable?

(Philosophical Positions on the Nature of Concepts)

The position that meanings are logically decomposable, i.e. reducible to certain elementary ones, is very old but also faces much opposition. Thus, Ferdinand de Saussure insists that it is only possible to define a concept in terms of its relations to other concepts. He observes that if words designate pre-existing concepts, then they should have exactly equivalent meaning in all languages, which is not true [Johnson-Laird 1983, pp.183-184].

Johnson-Laird observes that some words, like "lemon" or "triangle", do have the same meaning in all languages, while others may not be easily translatable. Not easily translatable are, for instance, the words "smart", "glamour" and "pattern", the expression "o.k.", or the German words "Gestalt", "kess", "fies", "gemütlich", "Kinkerlitzchen", "Krimskrams". Obviously, it is occasionally possible to give an approximate translation of these words. However, it is not possible to give a unique translation, which is generally valid in all instances, i.e. which has an exact counterpart in the second language. For instance, the word "smart" means, according to Divry's English-Greek dictionary, sharp, keen, lively, clever, elegant, to ache, to sting. These words do not have a unique, single-word expression, either in Greek or in German.

Johnson-Laird also observes that [1983, p.185], in spite of their uniform treatment in language, there are different kinds of words. Language often treats different things as though they were akin to each other, by referring to them by the same concept designation, the same name. Thus, various containers are all called "vases" and various flying animals are called "birds". This led many philosophers and psychologists to believe that the different entities subsumed by a concept necessarily have some features in common, and that this set of features suffices to define the concept [Hull, 1920], [Vygotsky, 1962]; an opposite view was held, e.g. by workers of the Gestalt-school of psychology.

For instance, the gestaltist Smoke [1932, p.5] wrote: "As one learns more and more about dogs, his concept of "dog" becomes increasingly rich, not a closer approximation to some "bare element" which runs through all stimulus patterns through which it was learned". Smoke's own view was that what one learns is "the total pattern" which constitutes the concept.

A similar view to that of the Gestalt-school was held by Wittgenstein. In his *Philosophical Investigations* [1953, pp.31-32], he considers the example of the activities called "games" and remarks that there are many sorts of games: board games, card games, ball games and so on. Looking at them, he says, we cannot find something that is common among them all. We see a complicated network of similarities overlapping and criss-crossing, which are no common features of them all, but "family resemblances".

One might add that, actually, what counts as a game is not determined by the nature of the rules for playing the game, but rather by our intention. The word "game" is not defined by its extension (all possible games) but rather by our "intention", i.e. by "being in a playful mood" while being engaged in this specific activity, by entertaining oneself instead of working.

Indeed, almost all activities can be called "work" or "game", depending on our mood, as it was very nicely shown by Mark Twain in "Tom Sawyer". In a well known scene, Tom persuades the boys passing by to paint aunt Poly's fence and pay him as well, for giving them the opportunity to play this new game. Johnson-Laird and Wason [1977, p.177] remark, similarly, that the concepts used in daily life are seldom logical functions of independent characteristics, but tend to have a relational structure that cannot be captured by mere logical combinations[36].

For instance, the everyday concept "table" does not just correspond to a conjunction of a top and some legs. The legs of a table support the top, which is a feature that does not clearly come through in the usual logical definitions of this concept.

Thus, concepts seem to have no precise boundaries and the best way to clarify the meaning of a word is usually to give instances of its application. However, this view raises the basic question: f there are no essential characteristics corresponding to a term, then on what basis is it applied? Wittgenstein talked of "criteria", fixed neither by inductive generalization of experience, nor by logical necessity but by convention. According to Minsky [1975], such criteria should be considered as default values, as characteristics of an object that can be assumed unless there is evidence to the contrary.

A similar view of how concepts are defined is held by Johnson-Laird [1983, p.190]. According to this view which, as he remarks, can already be found in Fisher [1916], a concept is represented by a schema that specifies not a set of necessary and sufficient conditions, but the typical or default characteristics of the items it designates. Such a schema allows the construction of mental models as representative samples.

In support of this view, Johnson-Laird mentions the experimental work of Berlin and Kay [1969], who found out that there is little agreement, from one speaker to another, about the boundaries of colours. What is more, they found that even one subject's choices of color names can vary from one session to another. On the other hand, they also found that there is a considerable consensus about which are the best exemplars of a color, which suggests that there are indeed "focal points" in the space of possible colors.

This line of investigation was pursued further by Eleanor Rosh [1973,1976] who concluded, on the basis of questionnaires asking for the most typical

[36] As Johnson-Laird observes, contrary to common belief, logic is not innate but merely an organized system for processing statements. Thus, for instance, the distinction of cause and effect is based on generalizing and interpreting experiences and not on logical deduction. That is why the cause-effect correspondence is not always clear. Obviously, for the same reason, words are not composed of logical components.

members of a category, that many natural categories are mentally represented by "prototypes" of their most characteristic members and that not all instances of a concept are deemed equally representative. Thus, a robin is a prototype bird, whereas a chicken is not.

Johnson-Laird [1983, p.201] remarks that the intrinsic vagueness of language, which is a problem for logicians, legislators, lawyers and other people who need precise formulations, is exactly the property that allows a concise representation of objective reality in the mind. Otherwise, a tremendous effort would be necessary to map in the mind the exact boundaries between classes of all the objects existing in the world.

BIBLIOGRAPHY

Berlin, B. and Kay, P.: Basic Color Terms: Their Universality and Evolution, Univ. of California Press 1969.

Fisher, S.C.: The process of generalizing abstraction and its product, the general concept, Psychological Monographs, 21, No. 2 (whole No.90), 1916.

Hull, C.S.: Ouantitative aspects of the evolution of concepts, Psychological Monographs, 28, whole No.123, 1920.

Johnson-Laird, P.N. and Wason, P.C. (eds): Thinking: Readings in Cognitive Science, Cambridge U.P., 1977.

Johnson-Laird, P.N.: Mental Models, Cambridge U.P., 1983.

Minsky, M.L.: Frame-system theory. In R.C.Shank and B.L.Nash-Webber (eds) Theoretical Issues in Natural Language Processing. Pre-prints of a conference at MIT, 1975.

Rosh, Eleanor: On the internal structure of perceptual and semantc categories. In M.T.Moore (ed): Cognitive Development and the Acquisition of Language, Academic Press, New York, 1973.

Rosh, Eleanor: Classification of real-world objects: origins and representations in cognition. In S.Ehrlich and E.Turving (eds): La Memoire Semantique, Bulletin de psychologie, Paris 1976.

Smoke, K.L.: An objective study of concept formation. Psychological Monographs, 42, (whole No.191), 1932.

Vygotsky, L.S.: Thought and Language. Cambridge Mass.: MIT press 1962.

Wittgenstein, L.: Philosophical Investigations, Translated by G.E.M.Anscombe, Macmillan: New York, 1953.

Appendix 4.2

Solving a Problem by Means of an Appropriate Generalization[37]

Contrary to what is usually believed, the solution of a mathematical problem is not merely a process of combining its premises appropriately. Many mathematical problems require a great leap of the imagination in order to be solved. Consider, for instance, the equality:

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{1}{mn(m+n+2)}=\frac{7}{4}$$

Such an equality cannot be proved by constructing a mental model of sorts and scrutinizing it to detect a way leading from the double infinite summation to the simple result 7/4.

The only way I have found to prove it, is by embedding this summation in a more general one. Using differentiation and integration in an appropriate manner, I was then able to determine its value. This is a procedure that has no direct connection with summation. However, even this hint will not help most students to prove the equality, because they do not know how to generalize the problem.

How was, then, this identity discovered in the first place? Probably as a by-product of the study of certain integrals. But, if no hint of this fact is given, one needs both experience with all these concepts and imagination, in order to retrace the way of deriving the equality.

In this case, we succeed to find a thread that leads to the solution, only after we search in the most unlikely places of our memory to find relations that might be helpful. Our only guide in this search is mathematical experience, which tells us which course is more expedient.

Many summations can be evaluated, for instance by expressing the summed quantities as differences of simpler expressions, most of which then cancel each other. E.g.,:

$$\sum_{n=1}^{\infty}\frac{1}{n(n+1)}=\sum_{n=1}^{\infty}\left(\frac{1}{n}-\frac{1}{n+1}\right)=1$$

The first hint that something more subtle is needed comes from the failure of all attempts to find such a transformation. This makes clear to the experienced mathematician that such a transformation, or possibly another reduction of this sum, is only possible if he can remove one of the factors of the denominator, say m+n+2.

---

[37] This example requires the mathematical knowledge possibly of a first-year student in Natural Sciences. The mathematically versed reader will benefit more from the discussion presented here if he first tries to find a proof independently. He can then observe his own thought processes. The problem was proposed to me by Prof. Th. Rassias of the National Technical University of Athens.

How can this be achieved? If the terms of the sum had the form $t^{m+n+2}\big/mn(m+n+2)$, then differentiation with respect to t would be sufficient, since the derivative of this term is t^{m+n+1}/mn. Thus, we get the idea to try and modify the sum by introducing the magnitudes t^{m+n+2} as numerators of the summed quantities. This of course, changes the sum that we consider, but we can later transform the numerators again to 1, simply by letting t tend to 1.

If we differentiate each term of the sum:

$$S(t) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{t^{m+n+2}}{mn(m+n+2)}$$

we get

$$S'(t) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{t^{m+n+1}}{mn} = t\sum_{m=1}^{\infty}\frac{t^m}{m}\sum_{n=1}^{\infty}\frac{t^n}{n} = t\left(\sum_{n=1}^{\infty}\frac{t^n}{n}\right)^2$$

However, the sum in parentheses in the last expression is known. It is the Taylor expansion of log(1-t), the natural logarithm of 1-t. This means that the sum of S'(t) is equal to $t[\log(1-t)]^2$.

In order to recover the initial sum, we therefore need only integrate this expression with respect to t and let, as said before, t tend to 1. After some transformations of the resulting integral, we get the expected summation result, 7/4

Here we have described only the crucial observations that guide the mathematician. The detailed calculations and proofs of convergence of the series are given below.

Where is the model in this process? Both the expression S(t), as well as its fundamental property: $S'(t) = t[\log(1-t)]^2$, have nothing to do with the premises. They appear only as results of a guess, which is guided by our wish to get rid of the term m+n+2 at the denominator.

The mental model finally used here is the result of an evolutionary process, which is guided by considerations of utility. We make various attempts to find a solution method and are gradually led by their failure to a totally different approach. The generalization is, so to say, a desperate reaction, since everything else seems to fail.

In this case, the resulting “model” is of course a conversion of the initial expression, but it is by no means in any way directly implied by the initial data. We might try various other conversions or transformations of the initial series. For instance, there may be some transformation that allows a direct summation, in spite of my failure to find it. The solution method of a problem is by no means unique. For many famous theorems, literally hundreds of different proofs are known.

Proof of the double sum equality

In order to prove that:

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{1}{mn(m+n+2)}=\frac{7}{4}$$

we note, first, that:

$$m+n\geq 2\sqrt{mn}$$

Therefore,

$$\frac{1}{mn(m+n+2)}\leq\frac{1}{mn(m+n)}\leq\frac{1}{mn2\sqrt{mn}}=\frac{1}{2}\frac{1}{n^{3/2}}\frac{1}{m^{3/2}}$$

and

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{1}{mn(m+n+2)}\leq\frac{1}{2}\left(\sum_{k=1}^{\infty}\frac{1}{m^{3/2}}\right)\left(\sum_{k=1}^{\infty}\frac{1}{n^{3/2}}\right)=\frac{1}{2}\left(\sum_{k=1}^{\infty}\frac{1}{k^{3/2}}\right)^{2}$$

Since the squared sum is convergent, the first sum must also converge and have a finite value.

By a similar argument we can show that:

$$S(t)=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{t^{m+n+2}}{mn(m+n+2)}\leq\frac{t^2}{2}\left(\sum_{k=1}^{\infty}\frac{t^n}{n^{3/2}}\right)\left(\sum_{k=1}^{\infty}\frac{t^m}{m^{3/2}}\right)=\frac{t^2}{2}\left(\sum_{k=1}^{\infty}\frac{t^k}{k^{3/2}}\right)^{2}$$

Since the squared sum is absolutely and uniformly convergent for $|t|\leq 1$, the first sum must also be absolutely and uniformly convergent for $|t|\leq 1$. This means that if we differentiate it term by term, we obtain again an absolutely and uniformly convergent series with the same convergence radius:

$$S'(t)=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{t^{m+n+1}}{mn}=t\left(\sum_{k=1}^{\infty}\frac{t^n}{n}\right)\left(\sum_{k=1}^{\infty}\frac{t^m}{m}\right)=t\left(\sum_{k=1}^{\infty}\frac{t^k}{k}\right)^{2}=t[\ln(1-t)]^2$$

The last equality is due to the Taylor series expansion:

$$\ln(1-t)=-\sum_{k=1}^{\infty}\frac{t^k}{k}$$

which holds for $t\in[-1, 1)$.

Therefore, for x∈[-1, 1) it is:

$$S(x) = \int_0^x S'(t)dt = \int_0^x t[\ln(1-t)]^2 dt = \int_0^{-\ln(1-x)} (1-e^{-u})u^2 e^{-u} du$$

where we have made the substitution:

$u = -\ln(1-t)$ or $t = 1 - e^{-u}$

The last integral can be rewritten in the form:

$$S(x) = \int_0^{-\ln(1-x)} u^2 e^{-u} du - \frac{1}{8} \int_0^{-2\ln(1-x)} v^2 e^{-v} dv = -[e^{-u}(u^2 + 2u + 2)]_0^{-\ln(1-x)} + \frac{1}{8}[e^{-v}(v^2 + 2v + 2)]_0^{-2\ln(1-x)}$$

where v=2u.

Letting x tend to 1 from smaller than 1 values we obtain:

$$\lim_{x\uparrow 1} S(x) = -[e^{-u}(u^2 + 2u + 2)]_0^{\infty} + \frac{1}{8}[e^{-v}(v^2 + 2v + 2)]_0^{\infty} = 2 - \frac{2}{8} = \frac{7}{4}$$

On the other hand, it is:

$$\lim_{x\uparrow 1} S(x) = \lim_{x\uparrow 1} \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{x^{m+n+2}}{mn(m+n+2)} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{1}{mn(m+n+2)}$$

since the double series in x is absolutely and uniformly convergent for |x|≤ 1.

Appendix 4.3

Creating Irrational Numbers

We call irrational numbers the numbers which cannot be expressed as fractions, i.e. quotients of integers.

Do such numbers exist and how can we express them, if not as fractions? Such a number is, for instance, $\sqrt{2}$, the square root of 2, i.e. the number q with the property that $q^2=2$.

In order to define $\sqrt{2}$, we may consider for instance the monotonically decreasing ($x_{n+1} < x_n$) sequence:

$$x_{n+1} = (1/2)(x_n+2/x_n),\ n = 0,1,2,3,...$$

with x_0 an arbitrary positive fraction, as well as the monotonically increasing ($y_n < y_{n+1}$) sequence:

$$y_{n+1} = 2/x_{n+1}.$$

It is:

$$y_1^2<y_2^2<...<y_n^2<..<2<..<x_n^2<...<x_2^2<x_1^2.$$

and the limit both of the sequence $\{x_n^2\}$ and of the sequence $\{y_n^2\}$ when n tends to infinity is 2.

Thus, we have a decreasing sequence of fractions $\{x_n\}$ and an increasing sequence of fractions $\{y_n\}$, whose squares come ever closer to each other and have the common limit 2. But there is no fraction which would be the common limit of $\{x_n\}$ and $\{y_n\}$. They converge, coming ever closer, but do not have a rational limit. There is no fraction $d/b = \sqrt{2}$ with d, b positive integers.

If such a fraction would exist, it should fulfill the equation: $d^2/b^2 = 2$, or $d^2= 2b^2$. But such an equation is impossible! In order to see this, first we assume that the integers d and b have no common divisor, since we can always eliminate common divisors. But since d^2 is a multiple of 2, d must also be a multiple of 2, say, d = 2c. Then we have $4c^2 = 2b^2$, or $2c^2 = b^2$ and by the same reasoning we conclude that b = 2p. This leads to the result that both d and b are multiples of 2, which is impossible, since we have assumed that they have no common divisor.

In the same way, we may show that the sequences:

$$x_{n+1} = (1/2)(x_n + k/x_n),\ \text{and}\ y_{n+1} = k/x_{n+1},\ n = 0,1,2,3,...$$

converge respectively from above and below to the square root of k, which is irrational in all cases that k is not the square of an integer.

The fact that $\sqrt{2}$ is no rational number was discovered already by the Pythagoreans in the 6th century B.C. This led them to the conclusion that it is no number at all and that it is only meaningful as a geometric magnitude (the hypotenuse of a right triangle with vertical sides equal to 1).

The numerical approximation of solutions of algebraic equations, which are usually irrational numbers (although this was not initially known), is studied already by mathematicians of the Helenistic era. Irrational numbers in the form of series are studied already in the 17th or 18th century. But the strict theory of convergence of sequences was created in the 19th century by Cauchy, Weierstrass, Dedekind and others. This theory is necessary for the foundation of the theory of real numbers, i.e. of the union of the sets of rational and irrational numbers. Dedekind simply decided to require the existence of a new number, which would be by definition the common limit of an increasing and a decreasing sequence, whose terms approach each other ever more. As we see, the concept of the irrational number evolved gradually, possibly for 2000 years.

The transition from rational to real numbers was not so obvious as it might seem today. For instance, the reason why Diofantos (2nd or 3rd century A.D.) sought integer solutions to algebraic equations may be that the only known numbers in his time were rationals (fractions of integers) and fractional solutions were equivalent to integer solutions of a modified equation. For instance, if $p = a/c$, $q = b/c$ is a solution of the equation $p^2 - q^2 = 1$, then $x = a$, $y = b$ and $z = c$ is a solution of $x^2 - y^2 = z^2$.

What made irrational numbers finally tractable was, possibly, the introduction of the handy decimal notation, which corresponds to a decimal series expansion of a number (an expansion in powers of 10). For instance,

$$22/7 = 3.142857142857142857\ldots$$

In this notation, any fraction has either a finite decimal notation, like 1/2=0.5, or a decimal part which is repeated ad infinitum, as in the above example. An irrational number is, then, one with infinitely long, but not repetitive decimal part. For instance, a simple irrational number is:

$$0.122333[1]_4[2]_5[3]_6[1]_7[2]_8[3]_9[1]_{10}[2]_{11}[3]_{12}\ldots$$

where $[k]_n$ means a string of n times the digit k. This number cannot have an infinite repetition of a finite number of digits, because the strings of identical digits grow longer and longer. Therefore, it cannot be a rational number.

The decimal notation made irrational numbers representable and thus acceptable. In fact, it plays an even more important role in establishing properties of irrational numbers. Cantor's proof that the real numbers between 0 and 1 are much more than the set of all integers, is based on a decimal representation for the real numbers (see **Appendix 1.1**). Without a decimal notation (or more generally, a power series notation of numbers) Set Theory could not be easily created by Cantor.

Appendix 4.4
Imaginal Representation versus Propositional Representation And their Relation to Mental Models

The controversy on whether the processing of visual and other information in the brain happens in an imaginal form or is based on “propositions”, has been a central subject of Cognitive Science for many years.

As we shall see, the experimental study of this subject has given controversial indications, which shows that even experimentation is some times not sufficient in order to settle a dispute. A possibility to overcome these difficulties seems to emerge here from Johnson-Laird’s theory of mental models.

We are going to discuss this controversy in depth, because it shows very clearly how many difficulties arise during the attempt to create satisfactory models for mental processes.

1. The Theory of Imaginal Processing as a Basis for Mental Processes

As it often happens in Cognitive Psychology, the theory of imaginal processing developed by Paivio, Shepard and Kosslyn, has its origin in common experiences of everyday life and has been later supported by systematic experimentation. However, this has not proven it definitively because the interpretation of these experimental results was disputed by theorists like Baylor, Pylyshin, and Palmer.

Most people report that, when they solve everyday problems, they rely heavily on imagery. In order to find a solution to somewhat composite problems, they seem to build up imaginal representations of the premises and to examine or transform these imaginal representations in a manner similar to the one they would have used if the imaginal representations were true visual experiences. In this way, they try to make the solution obvious, i.e. directly observable on the image.

Not only spatial problems, e.g. a travel-route, are solved in this way but also temporal ones, e.g. when we build a time-table to make the serial order of certain events visible. Similarly, introspective accounts by most scientists, of how they think, show that they feel that imagery plays a central role in all their investigations (see [J.Hadamard, 1954]). For instance, mathematicians doing mental calculations basically transform the formulae under consideration by performing imaginal symbol manipulations.

As Arthur Miller [1986, p.223] says, thinking in modes that are purely syllogistic or verbal cannot be very productive because it can proceed only linearly or stepwise. Truly productive thinking “is to a high degree intuitive, moving freely over a multidimensional field of data that were either heretofore unconnected or connected inappropriately”.

He also notes that words achieve their meanings in context, which means that either imagery or, at least, intuitive aspects that go beyond the verbal material have to be invoked. In this respect, he cites Rudolf Arnheim's [1971] comment: "What makes then language so valuable for thinking cannot be thinking in words. It must be the help that words lend to thinking while it operates in a more appropriate medium, such as visual imagery".

A typical instance, in which imaginal processing is an important part of the mental processes, as personal experience tells us, is the case where we try to solve spatial problems. In solving spatial problems, people report that they use their imagination to form visual images of objects or scenes, which they can rotate or magnify.

Such observations have led a school of cognitive scientists, the "imagists", to the formulation of a theory of imaginal mental processing, which considers mental images as a distinct sort of mental representation with two basic properties:

1. An image is a coherent representation of a scene or object, with all parts of it simultaneously available to a process of scanning similar to perception.
2. Images can undergo seemingly continuous mental transformations (rotations, expansions, etc.).

Thus, images are representations of objects and not mere references to them, like propositions. However, another school of researchers claims that mental images are mere epiphenomena and that "propositions" are the basic medium of thought. By this they don't mean that all information in the mind is stored and processed in the form of propositions, but rather that the ultimate form of representation of information, whatever it may be, has propositional rather than imaginal features, i.e. it is more like a verbal description than iconic.

2. The Theory of Processing by means of "Propositions" – The Homunculus Problem

In face of the observations we have mentioned, that favor extensive use of mental imagery, why should anybody doubt the existence of imaginal processing in the brain? It is the "homunculus problem" that creates these doubts.

The propositionalists say that imaginal processing requires an inward observer, a "homunculus" (= little human being) in the brain, and this merely shifts the problem we have to solve. Instead of determining the thinker's mental processes, we now have to determine the mental processes of this inward observer (see [Pylyshyn, 1973]).

This brings them to the assumption that information stored in the brain must somehow have propositional form, which seems to the propositionalists as the only form that allows syntactical processing, i.e. putting together information elements to form an all-embracing representation of a certain situation.

3. The Equivocal Indications of Psychological Experiments

Psychological experimentation seemed to be the only useful way to resolve this long standing controversy. However, the interpretation of the experimental results was also controversial and thus, extensive experimentation has not settled it in the way expected by the two schools. It has rather led them to come somewhat closer, by adopting certain positions of the opposite school.

The debate about imagery began after Shepard's and Metzler's [1971] report of experiments, which showed, or seemed to show, that imagery plays a central role even in simple mental tasks. They found that the time required by a subject to decide whether a geometrical figure is a rotated form of some standard figure or not, increases proportionally to the angle of rotation (according to Shepard [1978], by approximately 60° per second).

This led them to the conclusion that the subjects were mentally rotating an internal representation, similar to the depicted figure, in a manner similar to the one they would use if they perceived an actual object, which they could rotate. However, mental rotation experiments are not always in agreement with this view. The subject's ability to make correct judgements about the appearance of imagined objects under rotations, decreases as their shape becomes more complex [Rock, 1973] and people tend to make systematic mistakes about the number of visible corners of a cube after its rotation in a certain way [Hinton, 1979]. Such mistakes should not happen if the subject were really rotating an internal image of a cube.

Experimental results supporting an imaginal representation and processing of information were also reported by Kosslyn [Kosslyn 1973,1975], [Kosslyn & Pomerantz 1977]. These experiments involved mental image scanning. For instance, he let his subjects memorize a map of an imaginary island showing a church, a lighthouse, a river with a bridge across it and so on. Then he asked them to visualize the map in their minds, to focus on a certain place, e.g. the lighthouse, to imagine a spot moving from the lighthouse to the bridge or the church and to press a button as soon as the spot reached the destination.

The result of this experiment was that the time taken to press the button varied according to the distance between the positions of the lighthouse, the bridge and the church on the map. This led Kosslyn to the conclusion that there must be some image-property that is very much like real distance and some process of changing the focusing point of our attention on an image that is very much like traveling along real space.

In order to investigate the validity of Kosslyn's conclusions, Pylyshyn planned a series of similar experiments on mental image scanning, e.g. moving in an imaginal landscape.

When Pylyshyn asked his subjects to imagine a real spot moving along some course, the time taken to reach the destination increased proportionally to the increase of distance, just as Kosslyn had found out. The same happened when he asked them to imagine themselves walking along that course.

However, when Pylyshyn asked the subjects to imagine shifting their gaze as quickly as possible from one spot of the landscape to another and to press the button when they had done so, the time intervals were not proportional to the length of the routes. Similarly, when he asked them to imagine running rather than walking to a certain spot, they pressed the button more quickly than they should.

Thus, Pylyshyn argued that these experimental results are due to the subject's knowledge, goals and beliefs and not to the scanning of mental images.

People seem to him to adjust their behaviour unconsciously, according to their knowledge. They know that running is faster than walking and they know that shifting one's gaze from one place to another is almost instantaneous, while the time needed to walk from one place to another is roughly proportional to the distance of the two places.

Of course, introspective reports of most people tell us that they often "feel" that they are using mental images in solving certain problems. But experimental psychology has often shown that people's reports of what they think that they are doing, during mental processing, are unreliable.

For instance, they report "scanning" a memorized array of letters and can indeed recall the rows (left to right) and columns (top to bottom). But they are unable to scan the array diagonally, or read it from bottom to top [Fernald, 1912].

However, there are also directly observable indications that mental images are often used during mental processing. Experiments on how subjects recall the verbal description of spatial relations among commonplace objects indicate that mental images are better remembered than propositional descriptions of a phenomenon.

In fact, one of the techniques for memorizing a list of unrelated objects or data is to imagine following a route through a landscape, placing one of the objects or data to be remembered next to each one of the salient features of the landscape. However, another technique is to build a story that connects the various objects or data to be memorized (see [Alan Badeley, pp. 43-45]).

4. The Basic Assumptions of the Two Theories and Their Common Features

Thus, it is not so easy to settle this dispute merely on the basis of experimentation. We rather need a more careful analysis of the basic assumptions of each school.

Imagists generally agree on the following points:

1. The mental processes underlying the experience of an image are similar to those underlying the perception of an object or picture.
2. An image is a coherent and integrated representation of a scene or object from a certain point of view.
3. This image can be transformed, seemingly continuously, in a similar manner as we could change the visual perception of a real object by rotating

it or varying its distance from us. I.e., an image can be rotated or expanded in a continuous way.

4. Images are analogical representations of objects. Structural relations between their parts correspond to the perceptible relations between the parts of the object that is represented.

Propositionalists also agree on certain basic points:

1. Propositional representations are discrete, like those in a digital computer, using syntactic manipulations of abstract symbolic representations, rather than imaginal manipulations of analogic representations.
2. Propositions do not correspond directly either to words or pictures, but have some abstract symbolic form which can be processed and evaluated syntactically, like a sentence. Their structure is not analogous to the structure of the object they represent. Thus, they can be true or false or be parts of a semantic network.
3. In spite of that, they can represent continuous processes by small successive increments of spatial parameters.

Propositional representations are able to perform imaginal processing, because the symbolic propositional representation of an image is similar to the one that represents the perception of the corresponding object.

Comparing these points of view, Johnson-Laird observes correctly that many of the claims about "images" and "propositional representations" are very similar in both schools.

The "propositionalists" do not deny that the mental processes used in solving spatial problems are similar to those used in perceiving an object or scene. They agree that images can be put together by means of the underlying propositional representation, but consider these only a superficial feature of the mental system, since images are supposed to contain no new information but merely to make the stored information easier to handle.

Similarly, "imagists" are willing to admit that the ultimate means of storing parts of images may be "propositional", but they stress that the final imaginal processing is more than mere syntactic propositional processing.

5. The Significance of Imaginal Processing – The Creation of New Information by Means of Thinking

In this last assumption, they are obviously correct. Images do contain new information, even if they are put together by means of propositions. If we deny that, we may just as well claim that the way we reach mentally the solution of an equation is already contained in the equation, inherent in it, so to speak.

However, the solution process is not unique, even if the final solution may be (which is also not always the case). There may be simple or more complex ways of solution.

In order to solve an equation, we have to transform the equation until the solution becomes obvious. But this series of transformations is new information we create or discover! We may transform the equation into equivalent forms, but without a great deal of ingenuity we cannot discover this solution.

How do we find our way in an unknown city, from one point to another, without a guide? Common sense and correct evaluation of all indications, e.g. of the predominant direction of movement of the people or the density of traffic, representative buildings etc. may help us. For instance, if we are seeking our way to a stadium, a short time before an athletic event, we need only follow the main current of moving people on the street.

Such indications are not contained in the name of our goal, but result from other features of the goal, which we know or may infer. In the present case we make the reasonable assumption that we are not the only ones who are interested in the sport event and that a great number of people around us have the same goal.

We thus create information from non-explicit indications of general nature. Something similar to this also happened in the example of finding the value of the double sum, as we reported in **Chapter 4**.

6. The answer of the "Imagists" to the Homunculus Problem

As a representative example of how imagists answer the "homunculus" objection, we may consider Kosslyn's theory as it is presented in his book [Kosslyn, 1983].

In order to avoid objections to his theory, Kosslyn has introduced the notion that a mental image is not a truly spatial, but a quasi-spatial representation, resulting from the processing of depictive information, which is possibly propositionally encoded. This representation is the content of an innate mechanism for processing of visual-spatial information, the "mental matrix". This mechanism is used both during the perception of an image, as well as during its reconstruction from contents of the Long Term Memory.

Kosslyn [1983, p.25] assumes that various sets of neurons (or rather, patterns of neurons firing) in the brain stand for the various portions of an object, while other neurons connect these representations to a coherent whole, so that they give us the information needed to discern a picture.

So we do not need a "mind's eye" or a "homunculus in the brain" observing the internal images, which led Pylyshyn to reject imaginal theories. The "mind's eye" simply consists of the various tests that evaluate and interpret visually the information contained in the memory locations. Also, the "mental matrix", the basis of image processing in the brain, does not have to be a kind of two-dimensional arrangement of neurons. This name is only a metaphor for a functional organization, which results into mental images, although it may not have a two-dimensional form.

In a similar manner, a computer's screen depicts visual information as a two-dimensional array of dots, although this information is not stored in the computer's memory as a two-dimensional array. The computer merely identifies these pixels "in a way that results in their functioning as if they were arranged in a visual array".

As Kosslyn explains: "This model is roughly equivalent to saying that we have a set of neurons in the brain that store the information that goes into an image; to form the image, these neurons are simply activated, which directly results in a pattern of neural activity that is interpreted as a visual pattern in a matrix (i.e. a mental depiction) " [1983, p.93-94].

Kosslyn [1983, p. 96-98] also remarks that images are apparently not stored by memorizing the location of single dots but of organized units, such as lines, enclosed areas and so on. This follows from the fact that people being asked to memorize various composite geometric forms shown to them, take different times to imagine them, depending on how they analyze them into simpler figures for memorization.

The mind seems to analyze and store each image, in "chunks" (as Herbert Simon would call them), accompanied by instructions on how they have to be put together to reassemble the picture. Pedagogues know, for instance, that children gradually learn to recognize words as whole patterns, rather than arrays of letters, and they use this to teach them to read more quickly.

It seems that in the brain there is a function of "memorizing" recurring patterns, which operates for the permanent storing not only of the shapes of single letters, but also of words. The same is also true for much non-pictorial information. It is also processed and stored in "chunks", rather than either word by word or in whole passages [Kosslyn, 1983, p.99].

This is, for instance, obvious from the fact that people may remember very well the plot of a story they have heard, although they can be deceived about its exact wording.

Another indication for this, is that we may remember all details of an article without being able to tell whether we have read it in an English or German journal (at least, this often happens to me). This shows that the specific linguistic form is not kept in memory. It also shows that propositional representations, the kind of "symbolic" representation supposedly used in the mind, are very different from the actual linguistic expressions.

Kosslyn [1983, p.102] observes that the process of reassembling an image has no need of recalling first the names of the parts in order to imagine them, because people need less time to remember a row of letters than to form an image of each letter separately. Rather, nonverbal descriptive information seems to be used (stored possibly by propositional representations), which also explains, according to Kosslyn, why we have no awareness of using it. Since it is nonverbal, we will not "hear" it in our mind.

Let us now return to the equivocal experimental results, some of which favor imaginal processing, while others do not. How can we explain these results?

7. Mental Models as a Way Out of the Dilemma

A possible way out of this dilemma is Johnson-Laird's theory of mental models. They are models that capture structural relations of spatial-visual perceptions or of temporal ones, in an imaginal form, but they are not similar in all details to the objects they represent.

When a sentence referring to artists and beekeepers is processed, artists and beekeepers are represented by "tokens" and not by some sort of images. Mental models are not truly imaginal and they do not have to function in all instances like perceptual images. In certain instances, we may scan them in a way similar to perceptual scanning, while in other instances they may not be detailed enough for us to do so!

It seems thus that, even an initially linguistic description, is basically used in order to compose an imaginal "model". Then, this is scanned by means of the same automated observation procedures which are used during the direct observation of objects (this happens, for instance, during the mental performance of algebraic operations).

However, the imaginal model is by no means an accurate reconstruction of objects, because the brain has limited capabilities of maintaining its various parts in activated mode (on the basis of the experimental evidence on remembering "chunks", perhaps 5 to 7 such parts).

The composition of the imaginal model may be performed partially symbolically and even partially echoic-linguistically, as in the case of separating a telephone number in two-digit parts in order to memorize it. Here, the memorizing is echoic because, if someone repeats the same number separated into three-digit parts, we are confused and cannot recognize it immediately. This happens, obviously, because we do not recognize the series of echoic memories we have stored in our mind.

8. Pylyshin's Central Position

For the sake of completeness, let us also consider more closely Zenon Pylyshyn's central position [Arthur Miller, p.224-225]. Like Herbert Simon and Alan Newell, Zenon Pylyshyn considers the mind as a symbol-manipulating machine and mental images as mere epiphenomena, which play no central part in thought processes.

In this view, the brain plays the role of a computer, a "cognitive virtual machine", as he calls it and the mind is the "software", which runs on this computer.

Pylyshyn defines those mental processes that are explainable only biologically, and not by rules, as "cognitively impenetrable" and hence as a part of the functional architecture of the "cognitive virtual machine", while all other cognitive processes can be simulated with mental algorithms executed on the functional architecture.

This last category includes all the important mental phenomena, because propositionalists practically consider only primitive reflexes as cognitively impenetrable, i.e. as explainable only biologically and not by rules. The execution is supposed to be accomplished in a language similar to the computer's internal language, in which meaningless symbols are manipulated according to formal syntactical rules.

The symbols used in the brain are called "propositions" and have three properties:

a. They are abstract, having no meaning connected to any words or pictures.
b. They have truth-values carried by the rules of mathematical logic.
c. They have rules of formation or syntax.

These abstract symbols can thus convey meaning only by means of the syntactical rules used for their composition. Only in this sense are they interpretable as models of outward objects, functions or abstract concepts like numbers.

9. Shortcomings of the Theory of Propositions – The Advantages of Analog versus Digital Processing

Such a theory gives rise to two objections. First, it is very difficult to conceive what a "proposition" may be. There is nothing like them, especially in the lower levels of the brains mechanism for processing memories.

Second, there are far more "cognitively impenetrable" cognitive processes than propositionalists believe. All efforts to develop algorithms for problem solving or even to elicit information from experts in order to build "expert systems", have lead to no significant results.

An intrinsic reason why some researchers favor propositional encoding is, of course, that such representations make information representation and processing accessible to computers, or even to people trying to follow up mental processes.

In fact, only structural-syntactic processing is completely reproducible by a computer or any other digitally operating machine and only such a kind of processing can also have a precise formal or linguistic description. The various aspects of a thinking process cannot be described precisely if this process is analogical rather than digital.

An analogical computing device merely settles down gradually to a certain mode of operation and has no identifiable steps of progress towards this end. The change of the system is continuous, without discrete intermediate stages. Therefore, there is no adequate description of the solution process itself. Analogical computing is, however, far more effective than digital, exactly because the system has only to settle down to a solution by following its dynamics.

This is possibly the reason why mental processes like permanent formation of memories, e.g. building up of reflex mechanisms, are analogical and not digital. The study of neural processes reveals that beyond the processes leading to transmission of pulse trains, which might possibly be interpreted as digital, there are also electrochemical processes leading to plastic changes of the neuronal network (creation and annihilation of synapses), which could be conceived as discrete only at the level of single molecules. The only rule followed by this process is that if two neurons fire almost simultaneously, then a synapse tends to be formed between them.

10. Final Conclusions

Here we may recognize that the real issue, lying at the bottom of the debate between imagists and propositionalists, is the central question of Artificial Intelligence: can all information processing be reduced into purely structural-syntactical processing or not?

Propositionalists would like to answer, "Yes! It can!" to this question. That is why they are not persuaded by the imagists' efforts to incorporate "propositions" at some levels of their theories.

The "modelist" Johnson Laird [1983, p.151-153] observes that, in spite of the similarities of the two theories discussed above, imaginal processing seems to play a more central part in problem solving.

The internal representation of images is not as significant as their role in mental processing, just as in the case of programming languages, their features and the facility of using them is more important than the sort of internal binary code used in representing each command. We may also note again, that mental symbol processing is usually imaginal.

In order to compute mentally, e.g. the sum or the product of two fractions, we keep in mind an echoic-visual image of them and produce gradually the result, following the same processes we would follow if we were writing on paper and we could see what we have written. We do handle symbols, but we must also compose and keep in mind images of the intermediate transformations (e.g., transformation of the fractions to a common denominator). If we get numerators and denominators mixed we are unable to proceed.

Generally, propositional representations are supposed to be mental representations of verbally expressible thoughts. Thus, Johnson-Laird assumes that mental semantics maps propositional representations into mental models of real or imaginary worlds, in order to process them further. This is done, no matter what their actual physical form may be. I.e., he says that propositional representations are interpretable by means of mental models [1983, p.156], while images correspond to views of mental models [1983, p.157].

Of course, both mental models and images have a disadvantage compared to propositional representations. They are unable to represent a general concept like "dog" or "triangle". They can only form an image of a specific dog or triangle. As we have already mentioned, Johnson-Laird overcomes this difficulty by using a well known theory of empiricist philosophers.

He claims that, although a model is specific, it can easily be used to represent a general class of entities. When necessary, interpretive processes possibly treat a mental model as a representative sample from a larger set.

This position was adopted, according to Bertrand Russell, [Russell, 1912, ch.9], already by Berkeley and Hume. Rejecting the existence of Platonic "abstract ideas", they claim that when we wish to think about whiteness, we build a mental image of something white, taking care not to conclude anything about this thing that is not also valid for any other white thing.

Observing that this is certainly true as an explanation of the way our mind works, Russell remarks also that, when we try to prove something about triangles in Geometry, we draw a specific triangle but take care not to use any feature of it that does not belong to all other triangles as well. In spite of these remarks, Russell favors Platonism, which we discuss extensively in **Appendix 4.5**.

Actually, we may consider both theories as partially true, at a deeper level than the one considered by Johnson-Laird. Concepts are stored echoically, visually, kinesioesthetically (sensori-motorically) and have also a name. However, this representation is interwoven, as we try to show in **Chapter 6**, in many layers of "preconceptual" structures, thus combining analogic and structural elements in a complex way. Every "preconcept" is analogic, since it is formed by empirical associations, while the combination of "preconcepts" is structural, when it is based on rational and linguistic processing.

These complex representations are then activated and used to build up complex images, when it is necessary to build up a scene. This process probably takes place in the same parts of the brain that are used for initial storing and initial processing (comprehending) of incoming sensory information.

For instance, when we are singing, we recall one after the other (serially) the stances of a song we already know and thus recreate in our Short Term Memory a series of images, which are able to raise various sentiments. Thus, we revive in our minds the sentiments or the thoughts we experienced the first time we heard this song.

Obviously, in order to enjoy or appreciate some information we have stored in our mind, we must first activate its remembrance by reactivating the same mechanisms we used the first time we received it.

A Short Review

In this appendix we have considered the controversy between theorists of "imaginal" processing of information in the brain and theorists of "propositional" processing. The first ones believe that a great part of mental information processing is based on some kind of mental images, while the second ones believe that the only possible processing is a syntactic processing in a medium, which is very much like linguistic expressions (without being exactly that). Thus, they call the objects of this syntactic processing "propositions".

A principal objection of "propositionalists" to an imaginal processing is that any mental images require a mental observer, a "homunculus" who will observe these images. To this objection, the "imagists" answer that mental images do not have the form of real images, but are only an activation of the same mental processes used during the observation of images.

The imagists have based their views on a series of psychological experiments, which seem to show that, during the solution of problems of movement in space, people build mental images and process them in the same way in which they would process the images of real objects. The propositionalists, however, repeating the same kind of experiments, somewhat modified, have shown that such kind of imaginal processing is not able to answer more detailed questions correctly. Thus, they have concluded that the supposed mental images are mere epiphenomena. The answer of the imagists to this could be that mental images are never as detailed as the mental presentation of real, perceived, images.

In fact, both theories can be reconciled with each other by assuming that propositions are the medium creating mental images.

Here, we have emphasized that the mental processing of mathematical symbolic expressions, which is mostly done by means of visual manipulations, produces new information which is not necessarily inherent in the original expressions. This shows that the processing of images is not a mere epiphenomenon.

As a way out of this controversy, Johnson-Laird proposes his theory of mental models which, indeed, introduces in mental processing only as much imaginal information as is necessary and not full images.

Considering briefly the views of the imagist Pylyshin, we have seen that the theory about the existence of propositions has two shortcomings:

(a) It has not yet determined what these propositions are.

(b) It overlooks that some vital parts of information processing as, for instance, reflexive reactions, have an analog and not digital-syntactic form.

Concluding, we have proposed that mental processing is done, neither by means of images, nor only at the level of mental models but to a great extent, at the deeper level of preconcepts.

BIBLIOGRAPHY

Arnheim, Rudolf: Visual Thinking, Berkeley, University of California Press, 1971.

Fernald, M.R.: The Diagnosis of Mental Imagery, Psychological Monographs, 14, no.58, 1912, pp.1-169.

Hadamard, Jasques: The Psychology of Invention in the Mathematical Field, Dover Publ., New York, 1954.

Hinton, G.E.: Some Demonstrations of the Effects of Structural Descriptions in Mental Imagery, Cognitive Science, v.3, 1979, pp.231-251.

Kosslyn, S. M.: Scanning Visual Images: Some Structural Implications, Perception and Psychophysics, v.14, 1973, pp.90-94.

Kosslyn, S. M.: Information Representation in Visual Images, Cognitive Psychology, v.7, 1975, pp.341-370.

Kosslyn, S.M. & Pomerantz, J.R.: Imagery, Propositions and the Form of Internal Representations, Cognitive Psychology, v.9, 1977, pp.52-76.

Kosslyn, Stephen M.: Ghosts in the Mind's Machine, Norton &Co., New York, 1983.

Pylyshyn, Zenon: What the Mind's Eye Tells the Mind's Brain: A Critique on Mental Imagery, Psychological Bulletin, v.80, 1973, pp.1-24.

Russel, Bertrand: The Problems of Philosophy, Oxford U. P., 1912.

Shepard, R.N. and Metzler, J.: Mental Rotation of three dimensional objects. Science, v.171, 1971, pp.701-703.

Shepard, R.N.: The Mental Image, American Psychologist, v.33, 1978, pp.125-137.

Appendix 4.5
On Platonism and alternatives
(Philosophical Considerations on Meaning)

As we have seen in the discussion of Johnson-Laird's theory of mental models, some meanings exist in the mind, while others, like "lemon", seem to be in the world. However this does not resolve the philosophical problem of the nature of meaning in all instances. Does an "absolute truth" exist or not? Does the idea of "freedom", for which we are willing to sacrifice our lives, exist outwardly or only in our mind?

Abstract concepts like "'justice", "beauty", "freedom" or even "number", may be interpreted as "subjective", i.e. as constructs of the mind or "objective", existing somehow independently from our thoughts.

1. Plato's World of Ideas

Plato believed that all abstract concepts (e.g. beauty, truth, justice, or the mathematical concepts) have an objective, independent existence, because they are more than extrapolations of appropriate experiences. They correspond to inward impulses. Otherwise we would not use them.

According to Plato, we would not have a sense of beauty, if there were no abstract idea of beauty in the world. Why create the idea of beauty if there is no inward seeking of beauty? And why should such an impulse exist if "beauty" has no outward, somehow "objective" existence?

Without hunger, we are not looking for food. But we would never have such a feeling as hunger, if food did not exist objectively, so that we could seek it. Similarly, we would not develop the ability to count, i.e. create the integers, if the idea of integer did not somehow objectively exist.

2. Formalism versus Platonism in Mathematics

As a matter of fact, mathematical objects were the most outstanding platonic examples of concepts having an objective existence, independent of our knowledge of them (which he called "ideas"). These objects, for instance the integers or geometric schemata and relations are, of course, not physical and material but exist, according to him, eternally outside of space or time in an immaterial world, the world of ideas.

This seems reasonable to most mathematicians, at least with respect to the positive integers: 1,2,3,4,... The eminent German mathematician of the 19th century, Leopold Kroneker, says for instance: "God made the integers, all else (in Mathematics) is the work of man".

These mathematicians are called "Platonists", since they adopt the platonic idea that some abstract concepts are not mere products of our mind, but exist independent of it. Wholehearted Platonists are, for instance the greatest logician of the 20th century, Kurt Gödel, as well as Rene Thom, who discovered Catastrophy Theory.

Platonist mathematicians believe that they do not create mathematics, but simply discover already existing relations, just as a physical scientist may discover phenomena which have always occurred, but escaped our attention.

As a matter of fact, Platonism is what every mathematician wishes to believe, what he "feels" that is true. However, logical inconsistencies in the naive, informal foundation of Mathematics, which have been discovered around the end of the 19th century, make many of them oppose Platonism totally and take refuge in other philosophical interpretations of the nature of Mathematics.

The most prominent of these schools of thought is probably Formalism. A formalist is, for instance, Abraham Robinson, one of the creators of Nonstandard Analysis (see [P.J.Davis-R.Hersh, 1983, pp.318-319]). Formalism claims that Mathematics is only a play with meaningless formulas. It is simply the science of formal deductions of theorems from formal axioms, void of any meaning. All these formulas are supposed to have no meaning until supplied by an interpretation (possibly, a visual one) in everyday language, which interpretation is, however, of no concern to mathematicians, but only to those who apply mathematical formulas to other scientific fields.

From the point of view of Formalism, it may just happen that some formulas are useful for describing physical phenomena. This connects them indirectly with reality, but does not grant them an independent existence[38].

3. The Motives of Formalism

This extreme position is taken because, very often, naive mathematical proofs have been shown to contain tacit assumptions, which creep unobserved into the proof through the language used and most prominently through the visualization of the concepts used. Thus, it seems much safer to let Mathematics adhere to strictly formal proofs, void of any interpretation.

Since the last quarter of the 19th century, we know, for instance, that there are continuous, but nowhere tangible (nowhere differentiable) curves. However, it is extremely difficult or impossible to imagine such a curve.

This happens not only to laypeople, but even to professional mathematicians. As we have already mentioned, Charles Hermite says in a letter to Stieltjes (another great 19th century mathematician): "I recoil with dismay and horror at this lamentable plague of functions, which have no derivatives".

Formalism is unpleasant, as a philosophical theory, to most mathematicians, because it is totally counterintuitive, although intuition and visualization are two

[38] One of the reasons why original Platonism is popular with respect to Mathematics, is their unusual effectiveness in describing reality (see [Hersh, 1998]). We will discuss this relation in a separate appendix, **Appendix 4.7**.

of the most important mental abilities used by a mathematician. However, it seems to be the lesser anathema. As Davis and Hersh [1983, p.321] very nicely put it: "..the typical working mathematician is a Platonist on weekdays and a Formalist on Sundays. That is, when he is doing mathematics, he is convinced that he is dealing with an objective reality whose properties he is attempting to determine. But then, when challenged to give a philosophical account of this reality, he finds it easier to pretend that he does not believe in it after all".

Formalists believe that the natural numbers are mere constructs of our mind, defined by a certain formal system of axioms. Mathematics is, according to them, a logical study of certain initial formulas, the "axioms" of the system, which are arbitrarily chosen without any compelling outward necessity.

4. Are the Natural Numbers Inborn?

However, even today, it is difficult to believe that integers are merely an invention. It is inconceivable what else we could use for counting and measuring instead. That is why Kurt Gödel believes in their objective existence as "a priori" forms of conceiving reality, inherent in our mind, as Emanuel Kant would say. Besides, how else can we explain the fact that child prodigies discover the laws of arithmetic at a very early age, without formal instruction?

Shakuntala Devi, for instance, begins the introduction to her book [1990] with the phrase: "At three I fell in love with numbers. It was sheer ecstasy for me to do sums and get the right answers". At this age, most children do not even know how to count to ten.

Gödel's greatest achievement was proving mathematically that, no matter how rich our system of axioms of Number Theory is, there will always exist statements about integers which are neither provable nor disprovable by the existing axioms, i.e. seemingly independent of them. Nevertheless, according to Gödel they are true, because the way he constructs them attaches to them a true "metamathematical" interpretation. Thus, they are no human inventions and they have to be included as new axioms into the system.

However, the same view, of a priori existence in our minds, was also held about the concepts of Euclidean Geometry, both by mathematicians and by philosophers up to the middle of the 19th century. Then it was discovered that non-Euclidean Geometries (theories replacing the parallel axiom by other statements) have Euclidean models and therefore, are just as consistent (free from internal contradictions) as the Euclidean Geometry itself. Later, Albert Einstein showed that physical phenomena of interstellar dimensions can be much better described by a non-Euclidean Geometry, thus shaking any and all remaining faith that Euclidean Geometry is the only "natural" geometry that exists[39].

---

[39] Actually, both Euclidean and non-Euclidean geometries can be unified into one theory, on the basis of Differential Geometry, which is an extension of Analytic Geometry, the algebraization of classical (Euclidean) geometry. In Differential Geometry, different types of

Could a similar situation also occur for the Theory of Numbers? This is not very likely. The axioms which produce the set of natural numbers are only five (see **Appendix 4.7**). Except for the last one, they seem to tell us merely how to count. The last one is necessary, because it allows us to prove properties of the whole set of natural numbers. None of these seems to be replaceable by something else.

Note that the ability to distinguish quantities of up to four objects, without any training, is present in many animals, e.g. in crows. An interesting story demonstrating this fact is presented by the historian Tobias Dantzig [Barrow, 1993, p.168]. Furthermore, various birds and mammals can be trained to distinguish quantities of up to seven objects [Barnett, 1970, p.200][40].

Thus, the concept "integer" may be a Platonic idea, while geometric concepts do not seem to belong to the Platonic universe.

5. How the Concept "Natural Number" is Formed in our Mind

Here we must not forget, however, that we are speaking about the definitive form of using the concept "number" in common language, whose acquirement by a child is finished, according to Piaget, around his seventh year of age. This concept is then further extended by mathematical training, which replaces "number" by "integer", shows how we may build fractions, etc.

According to Developmental Psychology, the concept "whole number" is not a simple entity, but has various aspects, which are only gradually understood by a growing child (see **Chapter 5**).

A child of five or six years may be able to count verbally, but not yet be able to understand all aspects of "number". In his mind, numerical evaluation is linked for a long time with spatial arrangement. Given a dozen of red and a dozen of blue beads, he has to place them in corresponding pairs, in order to find out whether the two sets have the same number of objects [Piaget-Inhelder, 1973, pp.78-79, 104-105]. In order to count out five buttons, he has to actually touch them as he says each number. Otherwise, he may skip one, counting up to four, although five are in front of him [Singer-Revenson, 1978, pp.117-118].

Geometry are simply distinguished by different values of the curvature of space, corresponding to different metrics. In this sense, a geometry which is in some locations Euclidean and in other locations non-Euclidean, is also possible.

[40] This reminds us of the observation that the brain can keep in the Short Term Memory up to seven "chunks" of information at the same time. Both observations may be due to an ability of the brain's activation center to maintain arousal in no more than seven memory-activation circuits at the same time.

The ability to count is first acquired by placing two sets of objects in an one-to-one correspondence, in order to establish their equivalence. At first we find out that, if we have three friends and wish to give an orange to each, we need a set of three oranges.

At this stage, if we space out one of the two rows of beads mentioned above, the child may conclude that the longer row contains more objects. The ability to build up mentally a correspondence between two sets, without actually moving or touching them, comes later and gradually, starting with small numbers of elements. Only then is the child actually able to count in the usual sense, by tagging objects with number-names and ignoring, e.g. spatial arrangement.

Thus, in the case of integers, Platonism must not be interpreted as an a priori existence of this concept in our mind, but as the ability to gradually acquire or develop this concept. Of course, the same is also true for other Platonic "ideas".

6. The Platonic View about Ethics and Justice

The question, whether an abstract concept is related to a Platonic idea, is important also in fields other than Mathematics. If an abstract idea of "justice" does not exist, why do we feel that some actions are just, while others are not, even if they are lawful? Does our feeling for "fairness" have an objective basis or is it simply due to the values of the society in which we live? Is the seeking of social justice a necessity imposed by Nature or is it a mere subjective demand, a way to defend our personal needs and to maximize our share of the amenities provided by the society?

The whole system of conferring justice is based on the belief that there is an objective moral justice, no matter how closely we can approach it by legal laws. We tend to believe that, in spite of their shortcomings, the lawgivers are guided by an inward feeling of justice and not merely a balancing of economical and social interests. Otherwise, there is no need to respect society and we can follow up any egoistical pursuits as long as we do not get confronted with the Law.

It seems that certain concepts like "justice", referring to ideas or ideals, are goal-setters which direct our actions, although we could not say what exactly we mean by them. So we grant them a Platonic existence, even if we have never thought about it. We may, or should, believe in eternal justice, because otherwise we cannot conduct social life.

Of course, we know that some feelings of justice are merely based on social habits. But there are also many situations in which a moral judgement would, or should, be unanimous in all cultures. For instance, the adherence to the principle that we should not do to others whatever we would hate happening to us, should be unanimous.

So Platonism seems to be a very convenient way to provide a fundament for the immutability of ideas like "justice" or "morality", even if we could not always tell how their dictations should be.

7. What is the "Objective" Existence of Ideas?

But what exactly do we mean when we say that certain abstract ideas exist somehow outside of our mind? Plato does not give a clear answer to this question.

A possible answer to the question, how some concepts or ideas may exist without being products of mental processing, is given by Kant. He considers certain abstract concepts as inherent in our mental system, i.e. as preexisting forms of conceiving reality, preexisting molds ("a priori intuitions" and "a priori categories") in which the mind presses all outward experiences and forms them. "A priori intuitions" ("Anschauungen") are space and time, while quantity is an "a priori category". These concepts are therefore not simply preexisting in some abstract sense, but actually coexisting with our mind.

8. The Evolutionary View of Platonism

A somewhat different answer to this question can be given on the basis of the idea of the developmental formation of concepts, which is discussed in this book.

Concepts are seen here as being gradually formed in our minds by assimilating and accommodating outward experiences, on the basis of inborn drives and instincts. Thus, a certain "a priori" seed seems to be inherent in all basic concepts, although it may not correspond to their fully developed forms. For instance, not only man, but even mice strive for freedom. The instinctive urge to seek new stimuli (see **Chapter 8**) makes all animals try to avoid confinement.

The ideals of "justice" and "fairness" may similarly be traced back to rudiments existing in all animal societies. Thus, they may be founded on our social instincts.

Note that, e.g. altruism, i.e. self-sacrifice for the preservation of the species or the social group, is also present in animals. Many species, e.g. birds or apes, have vocal warning signals, which alarm their kin against an approacing enemy. Such a signal, however, is a disadvantage for the particular animal emitting it, since it calls the attention of the possible predator to it. It is, therefore, a kind of "kin altruism" as S.E.G. Lea [1984, p.45] notes. The evolutionary acquirement of such a pattern of behavior is only understandable, because it may increase the survival chance of the species. However, there are even cases of animals, e.g. dogs, who have sacrificed themselves to save their masters.

From this point of view, the "objective" or "a priori" nature of basic concepts seems to be reduced to the question about the origin of instincts, i.e. to the question whether biological evolution of the various species is a mere product of blind chance, or somehow goal-directed.

This is a question we are still not able to answer and, maybe, will never be able to answer. Thus, if we wish, we may still believe in a developmental variation of Platonism, in which ideas have an independent existence due to our nature, but can only gradually be conceived, depending on our efforts to develop our conceptual system[41] and to solve new problems that we did not face before.

This is valid, not only with respect to concepts of Mathematics or Physics, but also with respect, e.g. to moral concepts. Many dictations of Ethics indeed require a readjustment on the basis of changing outward conditions in each era. The exponential increase of world population in our time raises, e.g. the crucial question whether and to what extent is birth control morally acceptable. The same is true for contemporary Genetics.

The moral values are thus something that we must continuously revise, in order to adjust it to the changing cultural and technological environment. However, this revision, in order to be "persuasive", i.e. to comply with our inward feeling of justice, cannot be arbitrary. It must agree with our inward instinctive drives, which are much deeper and far more versatile than we usually believe. They are not conscious to us exactly because they consist instinctive modes of reaction, which do not require conscious processing. However, systematic observation of how we react in typical situations reveals them.

A Short Review

In this appendix we have considered the platonic theory, that certain concepts, like Justice, Ethics or Natural Number, are not creations of our mind but exist somehow "objectively". I.e., they are parts of a self-existent, immaterial world, the World of Ideas. This view is an answer to the question why we seek justice or moral behavior. It also explains why the ability to count appears in some children without systematic instruction, but can be established in a primitive form even in animals.

Considering more specifically Platonism in Mathematics, we saw that it is not universally accepted, although it is very popular. Geometric assumptions, which were considered to be self-evident properties of space, like the Euclidian parallel-axiom, were proved to be arbitrary assumptions. Also, the view that other mathematical concepts have a self-sufficient meaning beyond the mathematical formalism, has often led to errors, because the mathematician tended to introduce, in this hazzy meaning of the concept, tacit assumptions. Thus, another concept was developed, namely Formalism. Formalism believes that all mathematical concepts do not have an "objective" meaning, but acquire their meaning exclusively from the formal expressions in which they appear.

---

[41] Note that this view is totally different from the often proposed explanation that ideas like "justice", "moral values" etc. are products of rationalism, created by the society for the purpose of self-sustainment or self-improvement.

The faith in a "Social Contract" goes back, at least to ancient Greece. It is very precisely stated, e.g. by Socrates in Plato's "Kriton" (51 d-e). In this section, Socrates personifies the laws and lets them say that, whoever did not like them, was free to leave Athens when he came of age, but whoever stayed there, had undertaken an obligation to respect them.

Exactly how successful such a social contract is, for sustaining the social system it supports, is revealed by the fate of the Athenian Democracy.

Considering, more specifically, the concept “Natural Number”, we saw that it is not so simple as it appears at first and that it does not appear suddenly in our mind. It has various characteristic properties, each of which is understood by a child at another stage of his development. Thus, this concept is gradually assimilated and it is fully formed in the mind around the age of seven years. In this case, it was seen that Platonism must not be understood as a preexistence of the concept in our mind, but rather as the ability to develop this concept gradually.

Considering now how some other platonic ideas eventually appear, we saw that many of them seem to have an instinctive origin and appear even in animals. Thus, self-sacrifice is not an exclusively human virtue. From this point of view, Platonism can be considered simply as an inborn tendency to develop certain concepts. However, it is not obvious whether these concepts have an absolutely constant form, or are eventually flexible.

BIBLIOGRAPHY

Barnett, S.A.: ‘Instinct’ and ‘Intelligence’, The Behaviour of Animals and Man, Pelican Books, 1970.

Barrow, J.D.: Pi in the Sky, Penguin Books, 1993.

Courant, R., Robbins, H., Stewart, I.: What is Mathematics, 2.Ed., Oxford U.P., New York, Oxford, 1996.

Davis, P.J. & Hersh, R.: The Mathematical Experience, Penguin, 1983.

Hersh, Reuben: What is Mathematics Really?, Vintage, London, 1998.

Lea, S.E.G.: Instinct, Environment and Behaviour, Methuen, London and New York, 1984.

Piaget, Jean – Inhelder Bärbel: The Psychology of the Child, Routlege & Kegan Paul, London, 1973.

Shakuntala Devi: Figuring, Penguin, 1990.

Singer, D.G., Revenson, T.A.: A Piaget Primer: How a child Thinks, New American Library, New York, 1978.

Stoll, Robert R.: Set Theory and Logic, Dover, New York, 1979.

Appendix 4.6
The Relation of Mathematics to Reality

One of the reasons why original Platonism is popular with respect to Mathematics is their unusual effectiveness in describing reality[42] (see [Hersh, 1998]). Mathematical relations seem to govern all physical phenomena, although the foundations of Mathematics are independent from all physical theories. Thus, many theorists believe that the basic mathematical concepts are somehow independent from our mind and inherent in Nature or, in agreement with Kant, in the way we can conceive Nature.

1. What is Mathematics?

However, there is also a different philosophical explanation of this phenomenon. This explanation relies on conceiving Mathematics as a formal study of standardized structures or patterns[43].

[42] The physicist E. P. Wigner speaks of "the unreasonable effectiveness of Mathematics in physical sciences".

[43] Some people will object to considering Mathematics as a study of structures, because many mathematical disciplines study rather abstract formal systems than somehow geometrically structured objects. The word "structure" bears a geometric connotation. However, we must not forget here that such theories refer to superstructures or derived structures, imposed on much simpler initial ones and introduce a structural element on the visualization of the formal system, rather than on the visualization of the objects studied by it. Thus, while a derivative is still connected with building a tangent to a curve, a differential operator is merely a mapping of one set of functions into another. Such an operator represents rather a structure imposed on a formal system, the formal system of functions, rather than a directly geometric conception. The formal system of functions is also a formal structure derived from much simpler ones.

All these superstructures are geometrizing or visualizing structures, imposed not on real observable schemata, but rather on formal systems. This is obvious, e.g. when we use two- dimensional arrays, called matrices, in order to describe linear transformations. Here the two-dimensional arrangement is imposed on the mathematical concepts used, the linear operators, and not on the objects studied by the formal system. Similarly, considering various sets of functions we often use the name "space" and pretend that certain types of relations have some geometric meaning or connotation. We speak, e.g. of orthogonality of functions, although functions are graphically not represented by straight lines, which might be orthogonal to each other. We find it convenient to consider functions as infinite-dimensional vectors, which may have a specific angle between them.

Its is not the physical (i.e. material) parts constituting a pattern that are actually studied, but a standardization of them, which allows only a finite number of defining properties to be considered. From any physical object or situation, we maintain only certain of its properties and use these as a basis for a mathematical theory. We may even create different mathematical theories that study the same object, if we use different standardizations.

For instance, we may consider the form of an object by using Geometry, or study the way it moves by using Differential Equations. We may also use Group Theory if the object has a crystalline form and we wish to study its symmetries. All these mathematical disciplines study the same object by looking each time at a different selection of its features.

In the same manner, if we only wish to find out in how many ways we can place a set of similar objects in orthogonal arrays of rows and columns, having the same number of units in each row or column, we need only study divisibility and factorization (If the number of objects is m*n, then we can place them in m rows with n objects in each one).

2. The Example of the Straight Line

From the mathematical point of view, a "straight line" is a reduction of the perceived straight line to the absolute minimum of necessary defining properties. It is not an arrangement of infinitely many points in space, but an abstract object defined by only two points and by the Euclidean axioms. All other relevant properties of "straightness" result from the study of the axioms by means of logic. Psychologically, it may be very important to conceive a line as an unbroken series of points or segments lying adjacent to each other (if there is a gap, people speak rather of two lines than one [Piaget-Inhelder, 1973, p.106]). But mathematically, this whole mental schema is irrelevant.

3. What is the Essential Feature of Mathematics? Mathematics as a Parsimonious Study of Structures

The importance of Mathematics, its essential feature, does not lie in the fact that it simply considers various patterns or structures that supposedly bear some resemblance with reality. It rather lies in the fact that it is a *parsimonious study of structures*. Only absolutely necessary properties are postulated and all other properties result from logical deduction.

A mere observation of patterns in nature presents all these properties together, at the same time. So there is no indication of how they are linked together. There is no indication of which properties are necessary consequences of others. The interrelation of the properties is not obvious. That is what Mathematics is trying to reveal. Not just properties of structures, but the interrelations of properties of structures.

That is also why Mathematics is important for other sciences. It does not merely study patterns similar to some patterns occurring in nature, but rather gives a parsimonious representation of such patterns and reveals the interrelations of their properties. Mathematics is the art of parsimonious explanations.

This is, e.g. obvious in the famous formula which connects the number of vertices, V, the number of edges, E and the number of faces, F of a simple polyhedron, i.e. of a solid without "holes" in it, whose surface consists of a number of polygonal faces. As Euler has shown, it is always:

$$V + F = E + 2.$$

We may observe as many polyhedrons as we wish. Even if we suspect the existence of such a relation, this does not yet give us the certainty that it holds in all instances. This certainty is provided by Mathematics. It makes this relation obvious by mentally cutting out one of the faces of the polyhedron and deforming the polyhedron into a plane figure (for a proof see [Courant-Robbins-Stewart, 1996, pp.236-240]).

4. Why Mathematics Describes Successfully Physical Phenomena

So the case is rather the opposite. It is not true that mathematical theories have the inexplicable ability to describe physical phenomena particularly well, without having been created for this purpose. What is rather true is that physical theories consist of only those aspects of reality that can be described with sufficient accuracy by Mathematics. They are the science of precisely formalized and logically structured conceptual molds.

Consider, for instance, what we do when we describe the flight of an airplane. Usually we give a precise mathematical description of the starting or landing trajectory or of the travel route. However, we do not attempt to describe the movements of the pilots, although they are relevant to the whole phenomenon. We also do not try to give an account of the pilot's judgements about how to steer the plane, taking in account the wind or other weather phenomena. From the whole complex phenomenon, we select only that part which fits into a mathematical model particularly well.

The flow of water out of a tap is another example. Initially, when the tap is opened only a little, the flow is laminar, i.e. the water particles move along almost parallel straight lines. Their motion can thus be described by using the simple, algebraic form of Newton's law of gravitation.

However, when we open further the tap, the flow becomes turbulent. The droplets of water move unpredictably, splashing in various directions. Since we could no more describe this motion by simple dynamics, we used to apply statistical mechanics in this case. Although this description still remains valid as far as it goes, some decades ago another mathematical mold was found, which can describe this motion: Nonlinear Dynamics. This mathematical discipline describes turbulent flow as a chaotic phenomenon, using extended Dynamics and not Statistics. Thus, various mathematical disciplines can describe various standartizations of the same phenomenon.

5. Mathematics as a Mold of Reality

Possibly the most obvious mathematical mold imposed on reality is Euclidean Geometry. Are all shapes around us straight-edged or circular? Obviously, they are not! Why does then Euclidean Geometry consider almost exclusively (with the exception mainly of conic sections) shapes consisting of straight lines and circles? The answer is not that they are somehow the only "natural" shapes, but that they are easy to draw (by using rulers and compasses).

How strong the influence of this mathematical mold is, is revealed e.g. by the fact that Aristotle thought that the trajectory of a shot consists of parts of circles and straight lines. It took 18 centuries until Galileo and his pupil Cavalieri discovered that this trajectory was a parabola. It would probably take many more centuries, if this trajectory did not happen to be a conic section, i.e. an already known curve.

Mathematics is thus a parsimonious study of simplified and formalized structures, which we rather impose on reality than discover in it. Platonism, in its original form, is not necessary in order to explain why we use Mathematics in many scientific fields. Mathematics does not study structures somehow inherent in the objects and phenomena around us, but the physical sciences simply restrict themselves only to such aspects of the phenomena that can be pressed into mathematical molds.

6. Evolutionary Platonism or Kantianism as a Possibilities

However, also in this case, there may still be a psychological-developmental version of Platonism, or rather Kantianism. Although ideas evolve gradually both in our mind and in society, one may claim that there is always an initial seed of them that belongs to the nature of the world or the nature of our mind and not to outward experiences.

As we have seen, even animals are able to distinguish small numbers of objects of the same kind. With some training, they reach the number of seven objects. So the natural numbers are, eventually, not a mere human invention. They seem to preexist in an elementary form as gradually developing perceptual schemata, even in animals. This tendency to develop such forms of perception may be conceived as inherent in the nature of the world or in the nature of our mind, in its hereditary dispositions.

A Short Review

In this appendix we have considered the nature of Mathematics and its relation to reality. More specifically, we have considered its "unreasonable effectiveness in physical sciences", as the physicist E. P. Wigner calls it. However, we saw that it is not necessary to admit any kind of Platonism in order to explain this phenomenon.

It is not necessary to assume that the mathematical concepts preexist and are inherent in the world. It is only sufficient to become aware that Mathematics is, generally, the study of parsimonious models of abstract structures. It considers abstract structures and tries to bring forward the necessary interrelations between their elements, those which must hold even if they are not directly obvious. I.e., Mathematics tries to reveal the greatest possible number of interrelations of the elements of a structure, using the least possible number of assumptions.

Its effectiveness in the physical sciences is simply due to the fact that, in principle, they restrict their attention only to phenomena which can be effectively described by mathematical formalism. They are the sciences of parsimonious physical models and they are helped in their studies by referring them the physical models to corresponding mathematical structures. In this way, they obtain ready-made all the compulsory relations between the elements of the model.

Concluding, we have noted that, if so inclined, we can accept instead of classical Platonism, an evolutionary Platonism or rather evolutionary Kantianism. Concepts like the concept of natural number seem to preexist in an elementary form as evolving perceptual schemata, even in animals. We can thus say that these concepts, although they evolve gradually in our mind, have an inherent tendency to be formed, which belongs to the nature of our mind or the nature of the world.

BIBLIOGRAPHY

Courant, R., Robbins, H., Stewart, I.: What is Mathematics, 2.Ed., Oxford U.P., New York, Oxford, 1996.

Hersh, Reuben: What is Mathematics Really?, Vintage, London, 1998.

Piaget, Jean – Inhelder Bärbel: The Psychology of the Child, Routlege & Kegan Paul, London, 1973.

Wigner, E. P.: "The unreasonable effectiveness of mathematics in the physical sciences" in E. P. Wigner: *Symmetries and Reflections*, Indiana Univ. Press, Bloomington & London, 1967, pp.222-238.

Appendix 4.7
Could there be many kinds of Arithmetic?

This is not likely. The initial axioms which characterize natural numbers {0,1,2,3,....}[44] can be stated in common language (informally) in the following way:

1. 0 is a natural number.
2. 0 is not the successor of a natural number (For each natural number n, sn≠0, where sn is the successor of n).
3. The successor of any number is a natural number.
4. No two natural numbers have the same successor (For all natural numbers m and n, if sm=sn then m=n).
5. (Postulate of Mathematical Induction): If 0 has some property and for each natural number having this property its successor has also the same property, then all natural numbers have this property.

These are the axioms proposed by R. Dedekind in 1888 and G. Peano in 1889 [R.Stoll, pp.58-59]. Except for the last one, they seem to tell us simply how to count. The last one is necessary, because it allows us to establish properties of the whole set of natural numbers[45].

None of these axioms seems capable of being replaced by something else. They do not rely on visualization, like those of geometry used to, but are very simple statements about 0 (or, in other formalizations, about 1) and how to build its successors sequentially.

To these axioms we may add the recursive definitions of the internal operations of addition and multiplication, raising their number to seven:

6. For all numbers m and n: 0+n=n and sm+n=s(m+n), where sk is the successor of the number k.
7. For all natural numbers m and n: 0*n=0 and sm*n=m*n+n.

These last axioms introduce some rudimentary spatial structure (without making any use of axioms of Geometry). The sum of the integers m and n can be visualized as a linear arrangement of m objects and n objects, one after the other and the product of m and n can be visualized as the number of objects contained in a spatial arrangement of m rows with n objects in each row, i.e. in a structure resembling a parallelogram or, more specifically, a rectangle (see [Courant, Robbins, Stewart pp.2-3]).

---

[44] In some axiomatizations the sequence of natural numbers begins with "1" instead of "0".

[45] Note however, that these axioms do not hold only for the natural numbers, but also for the terms of an arbitrary sequence, where the successor of any "natural number" (= term of the sequence) is the next term of the sequence, while 0 can either be included in the sequence as its first term or may be considered as symbolizing the first term of the sequence.

Arithmetic, which is based on the above seven axioms, seems to be the simplest theory that can provide an extension of the basic process of counting. A different "Arithmetic", that does not follow the Peano-Dedekind axioms, stated above, does not seem to be as important as e.g. non-Euclidean Geometry. It seems that any formalization of the properties of natural numbers must be based primarily on a formalization of the process of counting, so that differences between such formal systems would be superficial.

Of course, the initial axiom system discussed here is incomplete, as Gödel has shown. However, its initial axioms seem to be immutable no matter how (or whether) we include undecidable propositions in it, because counting is the prototype of serial ordering. It is merely the basis of ordering things by attaching to each one a "name", determining its position in the serial order: "first", "second", etc.

This serial order is also able to determine quantity, since any interrupted counting-sequence corresponds to a certain number of objects. The number-name attached to the last object in the sequence also indicates how many objects are in it. Numbering is, therefore, in a certain sense equivalent to serial ordering.

If intelligent beings, totally different from us, exist somewhere in the Universe, they will probably also have some system of abstract ordering of objects and therefore a numerical system not different, or not much different from ours.

So the concept "integer" may be a Platonic idea, while the geometric concepts do not seem to belong to a Platonic universe. Note that a counterpart to non-Euclidean Geometry exists in Arithmetic, when we leave the integers and create the sets of rational numbers (fractions) and their extension, the real numbers.

Ordinary positive fractions have the trivial property that any one of them, multiplied by a sufficiently large integer, will exceed 1. The same is true for positive real numbers. This property is called "Archimedean Axiom". However, Nonstandard Analysis considers not only archimedean numbers but also infinitesimals, which do not fulfill the Archimedean Axiom, as self-existent quantities (the nonstandard numbers are composite, having an archimedean and a non-archimedean part).

BIBLIOGRAPHY

Courant, R., Robbins, H., Stewart, I.: What is Mathematics, 2.Ed., Oxford U.P., New York, Oxford, 1996.

Stoll, Robert R.: Set Theory and Logic, Dover, New York, 1979.

Appendix 6.1

Implications of the proposed Modeling for the possibility of constructing "intelligent" machines

Can we create intelligent machines, intelligent robots? We should not maintain self-deceptions that this is easy, even if we will gradually be able to simulate restricted features of the mind. So for instance, we are able to fly, although our flying machines do not have the flexibility, the dexterity of birds.

In the present case, an analysis of concepts into logical components has a restricted range of applicability, while the interactive creation of the conceptual system presupposes the existence of sensori-motor feedback systems, which will be gradually organized into more composite neuronal structures on the basis of experience, as well as certain preliminary, i.e. preexisting, structures.

The evolutionary formation of concepts is necessary for intelligent behaviour. Therefore, the creation of intelligent machines requires continuous education and training, just as baby education takes place continuously from the day of birth.

It is not a shortcoming of nature's "design" that babies learn to handle objects only gradually. Only in this way can all preconceptual levels be formed, i.e. all the wealth of empirical categorizations. The same will, therefore, also be necessary for "intelligent" machines. Among others, a new field of study: "Robot – Education" is needed.

Apart from this, "intelligent" machines must be somehow equipped with drives and feelings (what is more, such that develop on the basis of experiences), since the significance of drives and feelings for the evaluation of new situations is immense (see [LeDoux, 1998] or [Goleman, 1996]). Beyond that, a personal relation with a "parent" of sorts will also be necessary, since it influences, it enhances the sentimental maturation and provides motivation for the development of intelligent behavior.

As an argument against the basic possibility of creating intelligent machines, may be seen the fact that even today's neural networks perform very poorly in tasks of generalizing rules, i.e. tasks of induction. For instance, this can be observed in a three-layer neural network, devised by Hinton to compute family relations, i.e. to deduce relationships between pairs of members of a small, made-up family, after it was trained to reproduce a certain number of them correctly.

As Steven Pinker [1997, p. 130-131] reports, the network was indeed able, after a period of training, to generalize to new pairs of kin. But this happened only for the last four of 104 possible pairs of kin, after the first 100 were used for training. Besides, each of the 100 training pairs had to be fed into the network 1500 times, which makes up a total number of 150,000 training lessons.

Pinker notes, correctly, that children do not learn family relationships in a manner even remotely similar to this training. He also observes that such numbers of training sessions are typical of connectionist networks, because these networks do not deduce rules, but merely interpolate (new cases) between (already known) examples.

However, the "learning" procedure in neural networks is totally different from that in animals, as we have already pointed out in **Appendix 3.1**. One must also keep in mind that a baby does not begin acquiring the meaning, e.g. of the word "mother" the first time he hears this word, but the first time he interacts with his mother. In fact, this is a basic shortcoming of neural networks and certainly one of the reasons for their poor performance: they do not have a multifaceted interface with the surrounding, but receive only one kind of stimuli; the adjusting responses during training trials.

A new approach has been introduced in this field by Rodney Brooks, director of MIT's Artificial Intelligence Laboratory. According to Brooks, the complexity of biological behavior derives, not solely from organisms themselves, but from their interaction with a complex environment. "Gradually, it dawned on me that just because a behavior can be described as deriving from a complex set of rules, does not mean that is how it occurred", he said to Horgan [Horgan, p.220]. "Humans are capable of going through logical chains of reasoning, but mostly it's post hoc rationalization."

Extensive work, based on these ideas, is carried out on the technical level in his laboratory. Dozens of robots have been built, some of which are in various aspects human-like, e.g. the robots called Cog and Kismet. These robots have been equiped, not only by sensori-motor mechanisms, like arms, legs, a visual system, an acoustic system etc., but even with drives and emotions of sorts. They are able to distinguish roughly both facial expressions, as well as a speaker's feelings by his vocal intonation. They are both being educated, rather by interacting socially with their attendants, than being trained in the sense that neural networks are trained.

However, there are many difficulties connected with this kind of work:

(1) The brain is extensively prestructured at birth, although the concepual system develops on the basis of interactive experiences, after this moment. It may take considerable time before these initial structures are satisfactorily approximated.

(2) Even such an elementary concept as the concept "in", is based on a very complex preconceptual mechanism. In order to acquire it, first we have to acquire various aptitudes of inserting our hand or our finger in hollow objects. But this is not enough. What is, for instance the meaning of "in" in "Dressed in black" or "being in a bad mood"?

The first expression obviously refers to an action of covering an object with something having black color. Thus, even covering or dressing is associated with "in". The second actually means "having a bad mood". But why does it actually use the word "in"?

"In" is also associated with immersion. For instance, we often say "sunk into depression" or "sunk into despair". Thus, the concept "in" is also associated with experiences independent from tactile feeling, which do not come to mind instantly, but still are integrated in the concept's encoding.

(3) Intelligent behavior is not merely a coordination of reflexes, but needs also a marking, a tagging system (words) and social drives, which will make concept-standardizing possible. Building in, or teaching a robot to use specific reaction schemata and even providing it with some sort of drives, may bring it,

at most, up to the degree of maturity of an one-year old child. The conceptual system is still missing. The maturation state is that of pre-symbolic thinking. The robot must be provided with some means of transforming specific reaction schemata to non-specific ones, or rather of applying specific reaction schemata non-specifically. This gives rise to “general” concepts of sorts. We cannot teach the initial concepts directly. They first have to be built up by interactive experiences and then standardized by social interaction.

These difficulties may explain why Brooks, in his interview with Horgan, concluded: “The things we build just don’t work *nearly* as well as biological systems” and expressed the suspicion that some vital component, an organizational principle, concept or language is still being overlooked in this kind of work. Brooks also doubted whether computer programs could evolve on their own and create truly intelligent versions of themselves, which is the secret hope of designers of neural networks, genetic algorithms and other alternatives to the old rule-based approach to AI [Horgan, p.222].

A possible first result of such work will be, not a “thinking” machine, but rather a hybrid machine able to do much more efficient pattern recognition and image processing.

Another possibility of creating “intelligent mechanisms” of sorts, may be the one offered by genetic mechanics. The progress of this discipline may make it, for instance, possible to equip lower biological organisms with some effective interface to digital or analog processing devices, thus allowing a better evaluation of imaginal information.

In any case, we should not forget that truly intelligent machines of any type, biological or not, will not be controllable. Unpredictability is inherent in intelligent behaviour. But this is opposite to what we wish our machines to do.

What we require, above all, from any machine, is predictability. It should not be able to do something unexpected. We wish it to be an assistant with precisely determined behaviour and not an independent companion or eventual opponent. The initiatives and the final decisions must belong to us. However, a machine that would be able to think creatively, would be by its own nature unpredictable. We would never know how it would react to a new, unexpected situation. We will, therefore, have to be very careful with all attempts to create intelligent machines, when we reach that stage of technical maturity.

BIBLIOGRAPHY

Goleman, Daniel: Emotional Intelligence, Bantam Books, New York, 1996.

Horgan, John: The Undiscovered Mind, Simon & Schuster, N.Y., 1999.

LeDoux, Joseph: The Emotional Brain, Phoenix Paperbacks, London, 1998.

Pinker, Steven: How the Mind Works, Penguin, 1998.

Appendix 6.2

The transfer of a concept of accuracy from one field of Mathematics to another

The first instance which I am going to describe refers to a very simple mathematical insight. It is rather a matter of constructing a successful algorithm and, as experienced programmers know, successful algorithms do not always rely on intricate mathematical thinking, but rather on dexterous algorithmic thinking.

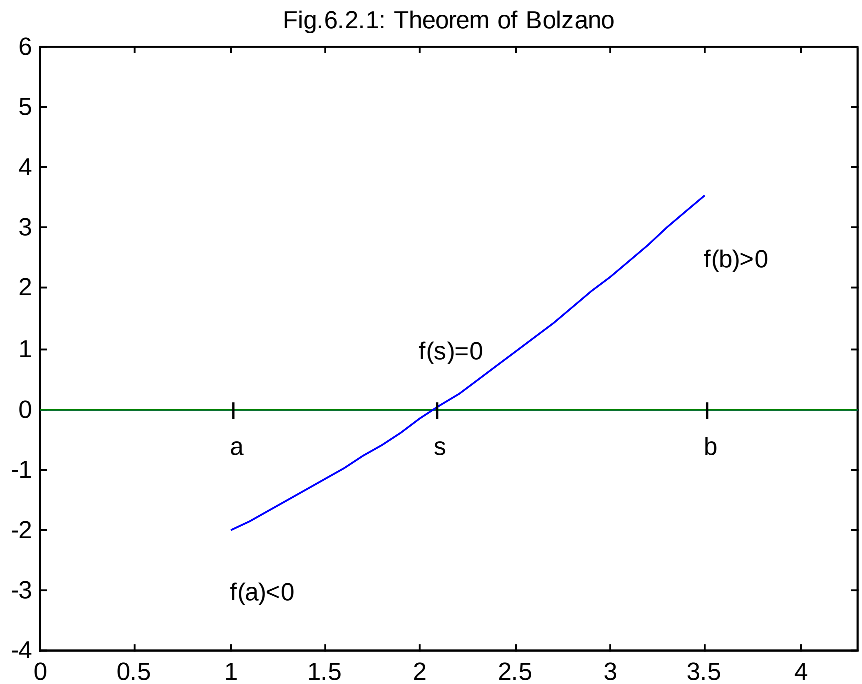
Mathematically speaking, the main idea is elementary, although its generalization, to make it applicable in all instances, needs considerable mathematical elaboration. This main idea is basically only a matter of recognizing the significance of certain magnitudes. However, as Ian Stewart says [1995, p.4], "No one has ever formalized "significance". To recognize what is significant, you need a certain amount of experience, plus that elusive quality: "intuition"."

Numerical techniques for solving nonlinear algebraic equations (finding computationally zero points of a given function), are based on iterative algorithms. These are repetitive procedures which, starting with a very rough initial guess of the solution, produce a usually better approximation with each repetition. Thus, when using such algorithms, one is faced with the necessity of determining how good the approximation obtained, after a certain number of repetitions, is.

This error estimation is usually based on Banach's "Fixed-Point Theorem", which analyses theoretically such iterative algorithms and provides an error-estimation formula. This formula, however, is not easy to apply, because it requires the knowledge of an appropriate, so-called, Lipschitz constant, which is not at all easy to determine. It depends not only on the equation considered and the algorithm used, but also, not in a very definitive way, on the initial guess of the solution.

Trying to show in my lectures how this error-estimation formula can be applied, without getting confused by all its inherent difficulties, I started teaching my students how to reduce the hazards of the initial guess and of the determination of an appropriate Lipschitz constant, by using a very simple theorem of Bernhard Bolzano.

This theorem states that, if a continuous function takes on values with opposite signs, at two points a and b ($a < b$) of the x-axis, then its graph must cross the x-axis at least once, at some point between a and b, which means that the function has a zero point there (Fig. 6.2.1). Obviously, any continuous line going from negative values to positive ones (or vice versa) must cross somewhere in between the x-axis, although the mathematical proof of this is not quite as simple (this is another example of a truth that is visually obvious, but by no means logically obvious as well).



[Figure 6.2.1: Bracketing a zero point by Bolzano's theorem]

This theorem seemed, however, to offer more than some method of making an initial guess of the position of a zero point. It is obviously a very good instrument for error-affirmation. If c is an approximation of a single zero point, s, of some function, then the approximation error does not exceed the value d, provided that:

$$- d < s-c < d,$$

or equivalently,

$$c-d < s < c+d.$$

According to Bolzano, in order to affirm or reject this statement, all we have to do is look for a sign change in the values of the function at the points c - d and c + d.

For some time the question, how I could profitably use this criterion, kept puzzling me. I did not know how to choose an appropriate error bound d. Then, as I was rethinking the whole situation just before starting a lecture, I suddenly realized what the d-values to be tested as possible error bounds should be.

Instead of seeking directly an error-estimation, i.e. an appropriate value for d, I should try a series of error-affirmations by using the values:

$$d_n = 10^{-n} \quad \text{with } n = 1,2,3,...$$

as possible errors. This sequence of possible error bounds quickly confines the error-estimation within the closest decimal order to the actual approximation error.

If an error affirmation is achieved in this way, e.g. with $d_3=10^{-3}$, we know that |c-d| does not exceed $10^{-3}= 0.001$, no matter how much better the error estimation by means of Banach's theorem would be. Thus, we are relieved of the necessity of determining Lipschitz constants at the cost of some additional function evaluations by the computer.

This of course, is a very simple idea, which does not involve any deeper mathematical thinking. From the mathematical point of view, it was more complex to find a way to generalize this procedure to functions with multiple zeros and to systems of algebraic equations (see [Kioustelidis, 1978]).

Discovering the appropriate kind of transformations in order to generalize this error-estimation technique and make it applicable in all cases, was not very difficult. I just had to think visually, using geometric interpretations of all mathematical terms involved. What may seem the difficult mathematical part of creating an effective algorithm was mere application of heuristic techniques, which I had often used.

The moment of inspiration, the "aha-effect", the moment of "illumination", as Poincaré calls it, was not connected with finding the transformation, but with the moment I realized that I had only to use a sequence of error-affirmation tests, in order to restrict the approximation error in a very narrow interval.

The important step for me was to realize the significance of the scale of possible errors d_1, d_2, d_3, … In this way, instead of seeking an appropriate error bound, I simply tried a whole scale, a geometric progression of possible errors, thus inverting the problem. Instead of concentrating my attention on the actual error, I tried to gradually enclose it by intervals becoming rapidly ever narrower.

Now, where did the idea of using the scale d_1, d_2, d_3,… come from? It may seem simple, when one describes it, but it was not at all obvious to me, since I kept puzzling for several months until I got it (it was also, possibly, not obvious to other mathematicians, since I found no mention of it anywhere).

The typical mathematical training, in trying to achieve maximal accuracy, led me to consider an error-estimate, which is, say, ten times greater than the actual error, as rather crude. Thus, I missed the main goal, which is to have a viable estimate of the accuracy of the given approximation. If one needs much higher accuracy, he probably has to find a better approximation by continuing the iterative procedure and not only a better error-estimate.

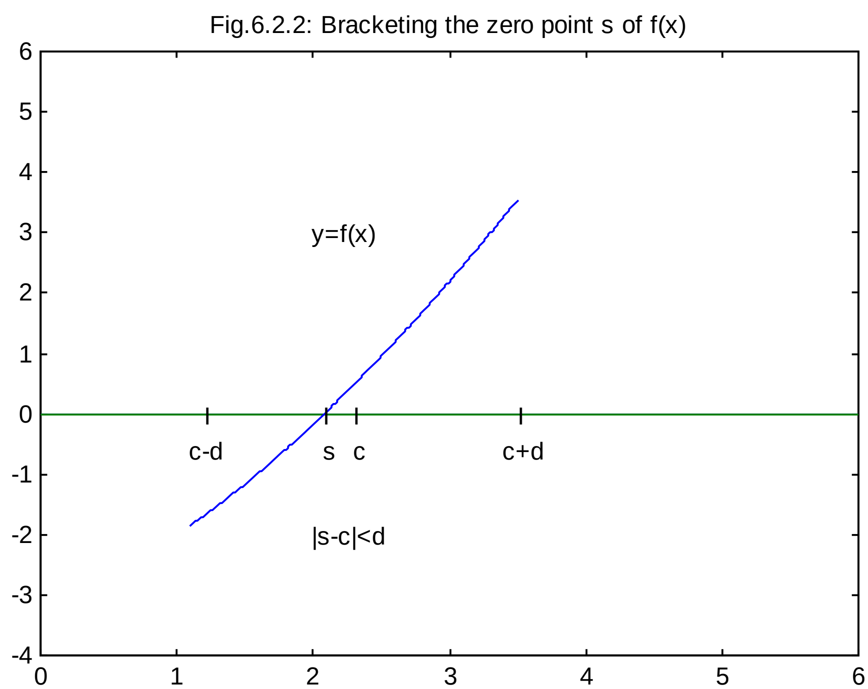
Trying to guess how I had reached the idea to use a geometric progression of negative powers of 10 as possible error bounds, I soon realized that I had met this kind of scale connected with errors in a very different mathematical field, the field of Chebyshev approximations for continuous functions.

The determination of an optimal Chebyshev approximation is complicated. So I had been very much impressed by a theorem saying that we can very easily construct "nearly optimal" polynomial approximations (by interpolation at Chebyshev points, whose definition does not concern us here). What had impressed me in that theorem, was the way it looked at "good", i.e. "nearly optimal", in contrast to optimal approximations.

An approximation was still considered as "good" (or "nearly optimal"), when the approximation error was "only" tens, hundreds, or even thousands of times larger than the best approximation error. The reason for this unusual appreciation of errors was that, for instance, a hundredfold error means only a loss of two decimal places in the accuracy of the approximation (i.e. the "good approximation" approximates the given function only two decimal places less accurately than the best approximation), e.g. up to one thousandth instead of one hundred thousandth. The approximation $\pi \approx 3{,}1416$ is still a viable approximation, although not so exact as $\pi \approx 3{,}1415927$.

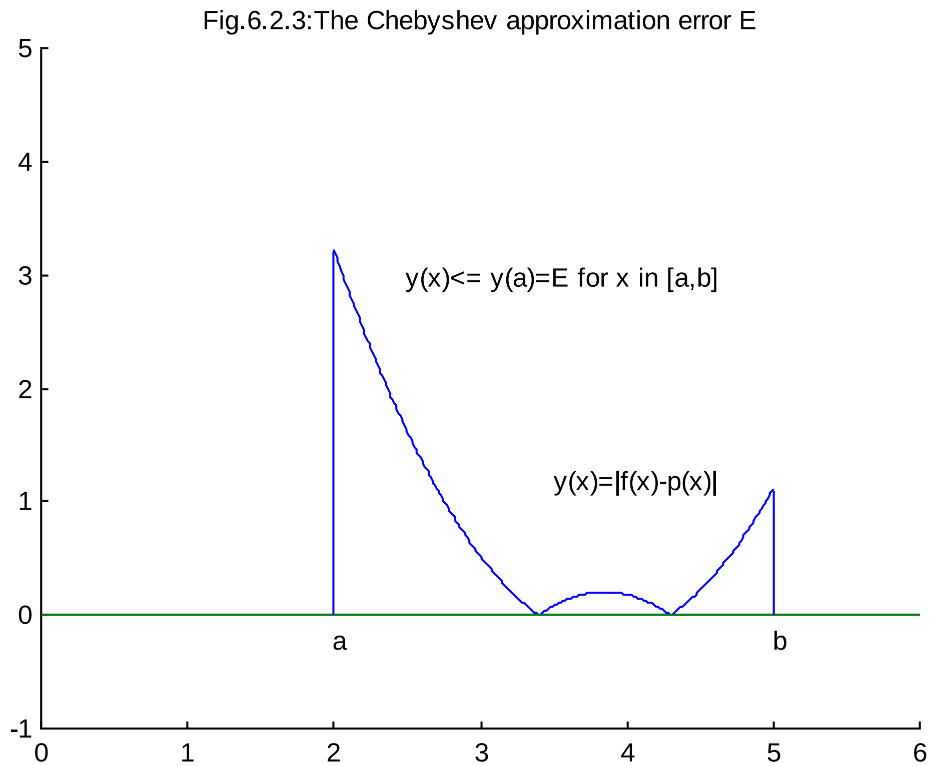
Remembering this theorem, I was led to the guess that somehow, a connection between the error conceptions in the two mathematical fields had unconsciously occurred in my mind. But how did this happen? The visual models of the two approximation problems under consideration are very different (see Figures 6.2.2 and 6.2.3):

The first problem, in the one-dimensional case, is visualized by a curve crossing the x-axis at some point, s, close to which there is another point, c, the given approximation of s. The last point, c, is furthermore enclosed in a symmetric interval [c-d,c+d], which provides the error estimate -d < s - c < d, if s lies in this interval (Fig. 6.2.2).



[Figure 6.2.2: Bolzano-bracketing of a zero point]

The second problem can be visualized as the curve of absolute error values $|f(x) - p(x)|$ (where p(x) is the approximation of f(x)), plotted in a certain given interval [a,b]. This curve never goes below the x-axis, but may touch it repeatedly. The Chebyshev Error Measure is simply the highest point of this curve in the interval [a,b] (Fig. 6.2.3).



[Figure 6.2.3: The Chebyshev approximation error]

The mathematical conceptions of error are thus very different. What is common between the error conceptions, in both cases, is not their mathematical model, but rather the way we tend to visualize the term "error" as a numerical value. Independent of its mathematical definition, "error", as a number, has also the general hazy meaning [numerical discrepancy], often visualized hazily by the schema "0.0...0nn…", where n represents various digits.

The unusual appreciation of good numerical approximations, in the latter theorem, was obviously what led me to the choice of the scale d_n, when looking for an appropriate error scale to be tested. This scale was quickly decreasing and it also gave the exact number of correct decimal places, when the test succeeded (more precisely, this last, happens by convention when we use the scale $d_n/2$).

This connection did not happen consciously, because the above visualization "0.00...0nnnn…" for "numerical error", is a schematization of practical experience with errors and not mentioned anywhere in books on Mathematics.

Such a device may help us work heuristically, but does not constitute "proper" mathematical usage. It is the kind of intermediate stage of thought, which we eliminate carefully, when presenting a proper proof. For this reason, it is not used consciously. It does not constitute a proper part of any definition of error, but rather a supporting semiconscious or unconscious part of the whole complex of ideas, referred to as "error". It gives a schematic expression to experience, without having any relation to the actual definition of the concept. Thus, it constitutes a subliminal, interactive, preconceptual[46] element of it.

The whole process of inventing an algorithmic error-estimation, described here, took place unconsciously and it was certainly neither linguistic, nor "propositional" in any conceivable sense or some kind of scrutinizing of a "model".

On the other hand, a visual model of the method was built up almost instantly in my mind after I got the main idea. It consisted of a curve crossing the x-axis, while pieces of this picture were cut out by vertical (to the x-axis) cuts, made alternatively on each side of the crossing point. At the same time, the picture was magnified by some kind of "zooming" as the cuts got rapidly closer and closer to the crossing.

This kinetic-visual model was certainly not involved in getting the crucial idea, because it is not even correct. The cuts should not be made alternatively, but simultaneously on both sides, in order to reproduce visually the mathematical process.

It seems thus that, while logical thinking (albeit non-formal) may take place at the level of "models", in Johnson-Laird's terminology, intuitive processes may take place much deeper, at levels of what we may call preconceptual connections.

The idea of using a geometric progression as a scale of possible error bounds, which is discussed here, exemplifies an outstanding feature of many instances of the creation of new ideas, new conceptual frames in Mathematics. These new ideas are often due to a mere shifting of motivation.

BIBLIOGRAPHY

Kioustelidis, John B.: Algorithmic Error Estimation for Approximate Solutions of Nonlinear Systems of Equations, Computing 19 (1978) pp.313-320.

Stewart, Ian: Concepts of Modern Mathematics, Dover Publ., New York, 1995.

[46] Preconceptual, in the sense that it constitutes an empirical preliminary stage of the concept's formation, an intermediate perceptual pattern.

Appendix 6.3

A second personal experience of inspiration: Optimizing Segmented Approximations

This second personal experience of discovering a mathematical relation [Kioustelidis, 1980] involves a visual modeling of the premises more directly, but it also uses a kinetic modification of this model.

This was a case of a general investigation with no predetermined premises and no known answer. I was investigating the problem of determining best approximations, of a certain kind, in very general terms. I knew that a simple answer existed for a very specific case, because I had derived it computationally one year earlier. But how could this result be generalized?

This is a typical instance of a general investigation, in which neither the appropriate, the most promising premises, nor the possible answer is known. The whole process was, therefore, not merely a matter of putting together all relevant information into a model and then scrutinizing it to find the answer. Initially, it was not clear which factors were relevant. I did not know whether an answer existed and what form it might have. Possibly, a simple answer would be available, not under general assumptions, but only under restrictive conditions of some kind.

Thus, I formed and scrutinized various models in my mind. I was not merely looking for an answer, but I also tried various modifications of the premises, that might yield a simple and obvious answer. I was not disheartened by not having immediate success, only because of the success I had in the previous year. Therefore, I kept on looking for a simple answer under fairly general conditions.

Some other futile attempts led, finally, to the construction of a visual model, an arbitrary curve having certain specific features that seemed to contain only all absolutely relevant information and nothing more. I studied, tested and modified this model for some time. Then, suddenly, I made a visual-kinetic modification that made the answer obvious. The answer was obtained by means of a sensori-motor evaluation of the situation, represented by the model, and not by any logical reasoning. No preconceptual", i.e. in depth study of the concepts was necessary in this case, but rather their procedural content, the sensori-motor procedures related to their visualization.

The final model that I was studying, transformed the initial problem into the simple geometric question of how I could reduce a certain area. Considering this visual representation in my mind, I suddenly realized that, to do so, I only needed to extend a certain curve. Its extension would "cut away" some of the area. This realization occurred by the change of the initial static image into a kinetic one. Only when I tried to shift one of the two endpoints of the curve, did I see that this would reduce the area.

In this case, a sensori-motor action schema was applied on the visual representation of the problem. The mental procedure that was used seemed like those we apply in order to prepare the action of peeling an orange or cutting a loaf of bread into pieces.

Another important factor involved in the process, was the motivation. I was fairly confident that a simple, directly observable answer to the problem did exist, because of my success one year earlier.[47] So I tried persistently to find a similar answer for the general case.

In fact, the immediate, visual way of proving this result, was characterized as "very surprising" by a specialist in this field, who already knew the result itself ([L.Schumaker, 1979]). He had actually proposed the same problem as a subject for a doctoral thesis to one of his students, who had solved a more restricted form of the problem analytically, i.e. by means of lengthy computations and received his Ph.D. just before my publication appeared. So the result itself was not new to them. What they found impressive, was obviously the visual, almost devoid of words and formal calculations proof of it[48].

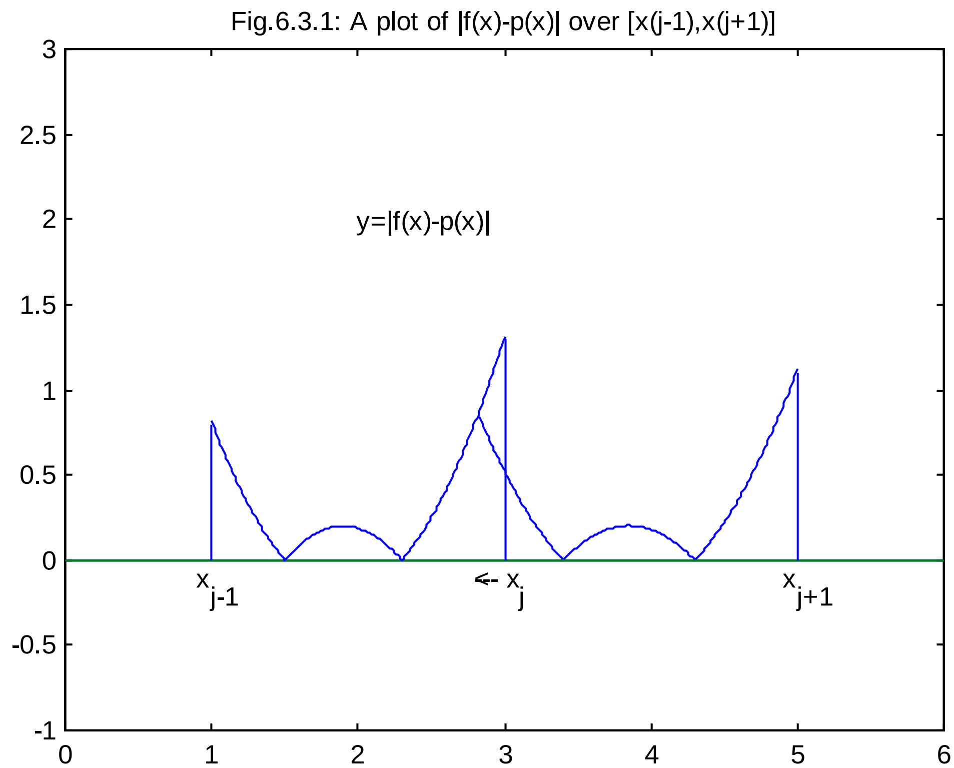
The problem under consideration is how to approximate a given continuous function, f(x) over some interval [a,b] by a fixed number, n, of pieces of a given form, say parabolas or even straight lines.

Suppose that p(x) is the approximating curve, which consists, e.g. of pieces of parabolas. Since there are n such pieces, the whole interval [a,b] is the union of n non-overlapping subintervals, $[x_j,x_{j+1})$, j = 0,1,2,..,n-1, with $a = x_0$, $b=x_n$, so that in each such interval the approximating function p(x) consists of a single parabola $p_j(x)$ (or whatever other kind of approximating curve we have chosen).

The question is now how to choose the coefficients of each parabola and the break points, x_j, in order to have the least possible overall approximation error.

[47] It said that in the case of a convex function and piecewise linear approximations, the optimal approximating curve must be continuous.

[48] Later, I found out that special instances of this problem had been solved analytically by other mathematicians also. The earliest such solution is that of M.J.D. Powell [1968], as part of the proof of some other result. A similar result was obtained also by D.M.Hawkins [1972].



[Figure 6.3.1: A plot of |f(x)-p(x)| over [x_{j-1},x_{j+1})]

We can use various criteria as measures of the quality of the approximation of f(x) by p(x). One possible such criterion is the magnitude of the area under the error curve, f(x)-p(x) (between this curve and the x-axis), over the whole interval [a,b]. However, in order to avoid the case where positive deviations are cancelled by negative deviations, we prefer to consider the area under the curve of the absolute value of the error |f(x)-p(x)|. If we could make this area vanish, then we would have a perfect fit.

I was studying how to minimize this error measure during an evening walk. I visualized the error curve |f(x)-p(x)| as consisting of adjoining non-negative pieces |f(x)-p_i(x)|, each lying over the corresponding interval [x_j,x_{j+1}).

Obviously, these pieces should touch the x-axis at one or more points, so that the area beneath each such piece would be as small as possible. In fact, if the position of the break points x_1,...,x_{n-1} is fixed, it is known how to determine the optimal approximation, the one that minimizes the error area. But since we may choose the position of these points as we wish, how should they be placed in order to minimize the error?

Note that the arcs lying on either side of the arbitrary break point x_j need not meet, i.e., have the same y-value at x_j, since the parabolas used on each side are different. So let us say that the left arc has a higher value at x_j than the right arc.

Looking at this mental image, I suddenly realized that I could reduce the total error area by extending the (lower lying) right arc to the left, until it met the (higher lying) left arc. This would cut away the part of the error area that lies between the two arcs. This would also mean, of course, that the separation point of the two arcs, x_j, would also move to the left and upwards of the position, in which the arcs meet each other. This, then, was a basic criterion for choosing the positions of the separation points, x_j.

This realization occurred by the change of the initial static image into a kinetic one. Only when I tried to shift the point x_j did I see that this would reduce the error area. In this case, it seems that only kinesthetic (visual-kinetic) thinking was involved in the evaluation of the final model, since no change in the conception of the various aspects of the problem was necessary. Furthermore, it was the conviction that a simple, visually obvious answer exists, that led to the creation of the final, successful model; thus, motivation played an important role in the whole process.

This is a case that reveals how important can be the sensori-motor schemata, attached to some abstract concepts. It also shows that the role played by motivation, in making some discovery, can be very important.

BIBLIOGRAPHY

Hawkins, D.M: On the choise of segments in piecewise approximation, J. Inst. Math. Applic., 9 (1972), pp.250-256.

Kioustelidis, John B.: Optimal Segmented Approximations, Computing 24 (1980) pp.1-8.

Powell, M.J.D.: On Best L_2 Spline Approximation, pp.317-339 in ISMN 9, Birkhauser, Basel, 1968.

Schumaker, Larry: L_2-Approximations by Splines with Free Knots, pp.157-182 in: G.Meinardus (ed.): Approximation in Theorie und Praxis, Bibliographisches Institut Mannheim, 1979.

Appendix 8.1

A Discussion on the Feedom of Will

The realization that our behavior depends on inborn action and reaction schemata raises a basic question: How strong is this dependence? I.e., how automatic are the inborn stereotypical behavior schemata we call drives or instincts? How much does an individual control his actions and reactions? How far can he modify his actions and reactions, given that they are certainly "co-determined" by his drives or instincts?

What we consider as freedom of behavior, might simply be the result of increasing differentiation of behavior, due to a different conceptual development of each individual, according to his personal experiences.

The question, whether human beings have freedom of will or not, is very often posed, but the answers given vary according to one's basic philosophical outlook. Possibly, no definitive answer can be reached by rational arguments. From a certain point of view, this question seems to go beyond the proper extension of the concepts involved, so that resolving this dilemma seems to be like trying to find a logical answer to certain logical antinomies, which go beyond the proper extension of logical categories and rules (see **Chapter 1**).

In order to elucidate this problem, let us look at the possible answers, developing arguments both ways, pro and contra free will.

Thesis in Favor of Freedom of Will

All physical theories about a deterministic or quasideterministic (partially probabilistic)[49] functioning of the world, are models built up on the basis of observations. Such models try to organize these observations into a coherent system of ideas, which will allow predictions of what will happen in similar situations. However, they can never claim to be absolute truths, since any new observation may change the model.

No model of reality can be identical with the reality itself, just as the picture of some object is not identical with it. An absolutely true model of reality would have to be as complex as reality itself, encompassing all possible phenomena, not only physical but also mental ones.

Sir Arthur Eddington [1985] uses a related argument in defense of free will. He says that, in order to create Physics, we are obliged to use simplifications, to omit various aspects of a phenomenon and reduce, so to speak, its "dimensionality". If there are too many factors involved in it, we try to keep some of them constant and possibly ignore others, with the excuse that they are

[49] Quantum Theory offers a deterministic model, but only for the probability that a particle will be in a certain domain of space at a given time.

unimportant. But there is no guarantee that this is so. Today we would say: if a butterfly flapping its wings in New York could in certain instances create a storm in China, as a popularization of Chaos Theory says, who is to judge what is unimportant?

Note that this popularization is somewhat exaggerated. Nonlinear deterministic models do not always behave chaotically. This happens only when some parameters of the problem are in a certain range of values. But when a parameter changes slightly, entering such a range of values, the magnitudes modeled turn from a regular, predictable (quasiperiodic) behavior to a chaotic, unpredictable one. This means that insignificant parameters can suddenly become significant for the description of a phenomenon.

From another point of view, modeling requires an observer, who is free to observe and build up the model. If the model refutes free action, then it refutes its own validity. It is no more a result of distinguishing true and false.

Consider more closely the meaning of the logical evaluations true - false: at the most elementary level, we call a proposition "true", if what it describes has a close correspondence with the outward reality as we experience it and we call it "false", if there is a discrepancy between it and the outward reality. This system of evaluations is then extended to more complex situations by inferring true or false on the basis of logical rules.

But these inferences are based on selecting and putting together the relevant data. How can this be done without a free acting mind? Without freedom of will, there is no freedom of selecting relevant facts. Without freedom of will, the mind reaches conclusions, not on the basis of a free search of supporting or negating experiences and correct application of rules of inference, but mechanically, on the basis of compulsive forces, which may drive it in a certain direction, not allowing a correct establishment of facts.

Of course, compulsion cannot be totally avoided. Acquirement of knowledge is the incorporation of new facts, new experiences in our mental system. But these experiences are sought under the guidance of our drives and filtered on the basis of instinctive and acquired evaluation schemata. What we perceive depends very much on our interests, on our personal preferences. But a free investigating mind will always look for contrary experiences and counterarguments. We can never be certain that what we believe corresponds to reality, but we think that at least it has a close relation to it.

Thus, in spite of the importance of the inward driving forces, inborn or acquired, we cannot deny the existence of freedom of will on the basis of any scientific theory. If there is no freedom of will, there is also no freedom of thinking and judging. But such a conclusion leads to devaluation of all scientific theories.

What is important in all these reflections, is the freedom of the mind to make its choises and reach its conclusions. If there is no such freedom, then all theories are worthless, since we can never be certain of their true value. If we have built up our theories solely on the basis of inward compulsions or outward driving forces, guiding our mind in a certain inescapable way, then the validity of all scientific theories is an illusion. Then we act automatically, like the mechanical puppets built by toy makers. There is no process of mental adaptation to reality; there is no seeking of truth. Our judgment has no value.

Antithesis: Arguments Against the Freedom of Will

Determinism is, in a certain sense, INTRINSIC in the search for explanations! It is a prerequisite of Science in looking for causes. To every "why" correspond necessarily one or more "causes". So we have to adopt determinism, unless we are ready to admit the existence of phenomena without natural causes, e.g. phenomena due to the intervention of a "higher will".

This is also the case in the search for motives: Suppose that we admit the existence of a non-causal, non-deterministic part of the brain. What is its nature? How does it operate? If the mind is not moved by inward motives, outward stimuli and by the inherited structure of the body, what else is the cause for the course that the thoughts take? What does "freedom of thought" imply? Random choices? But random choices, if they exist, have nothing to do with freedom. They are merely products of chance. If choices are not random but, in spite of that, "free" what determines their selection?

In reality, the meaning of our belief, our confidence, in a "freedom of will" is different: we feel free, not when what we do does not obey any rules (physical or biological laws), but when we are able to follow the dictations (or suggestions) of our internal drives, our internal motives! Freedom means independence from outward restraint and the possibility to follow our inward impulses, our drives. So, freedom of will is merely absence of outward restraint (so that we can follow our inward drives, the dictations of our instincts).

This, of course, does not mean that we have to sink into fatalism. As Euan Squires [1990] correctly implies, no matter whether we believe in the freedom of will or not, we have to behave "as if" thoughts were free. Otherwise, we do not act "freely", according to our impulses, but become victims of inactivity and passivity.

Even if the world is a big machine, driven by necessity,y we cannot take this view as a guide for our behavior. We cannot simply act on the basis of fatalism, of the view that "whatever will be, will be" and merely sit back waiting for the opportunities to knock at our door or the ideas to come to our mind, without preliminary investigative effort. We just have to "pretend" that we have a free acting mind and try to use it in the best way we can, to achieve our goals

Opposite View (Counterargument)

In the above deterministic way we may try to explain everyday (recurring) situations. But can we also explain unique phenomena like scientific discoveries, which alter permanently many aspects of our social and even physical surrounding?

Deterministic explanation is a reduction of a composite phenomenon to simple ones, occurring indeviably, which are considered as basic principles. But scientific achievement is ultimately unique. Even if many discoveries are made by different people almost at the same time, this probably lies in the maturing of a certain idea in the scientific world. But there is no guarantee of replicability of scientific achievement.

Bernhard Riemann, for instance, wrote that he had a proof for his famous statement, concerning the location of the roots of the, so called, "Riemann's zeta function". He did not publish it, because it was based on a too complex expression for the zeta function, which he was still trying to simplify [G.F.Simmons, 1993, p.218]. Unfortunately, his notes were lost and no one has yet been able to provide a proof for Riemann's conjecture, more than one century after his death.

Deterministic explanations have no value if they are unable to predict and replicate phenomena. Concerning mental phenomena, it is useless to try and replicate them exactly. It is therefore futile to try to reduce free will to determinism.

Final Position: A Synthesis, or rather a Hyperbasis (an Overstepping) of the two Opposite Views.

Actually, there is another consideration, which seems to favor a certain kind of "freedom of will" or, at least, renders the questio whether we have freedom of will or no meaningless: as we have seen, drives and instincts are modified by experience. They are not merely inborn mechanisms, but often only inborn abilities to build up action and reaction schemata ,on the basis of experience, or to acquire certain skills.

For instance, young ducklings tend to follow the first animal that is near them, when they come out of the shell, and consider it as a "mother" (they behave as if it were their mother). Thus, our drives, our instincts are not merely determining factors of behavior, but are also formed by outward experiences, which are partially a matter of personal choices. This means that we are not merely influenced by our surrounding, but rather constitute a unity with it. In this process, there is no unique deterministic direction from cause to effect. We are influenced by our surrounding, but we also influence it. By making choices of actions but also of conceptions, we modify both the actual physical surrounding, as well as our view of it. Of course, we may say that, ultimately, the physical world, or "Nature", determines our choices. But this is merely a choice of starting point in the cycle: man ⟳ world.

A Short Review

In this appendix we have considered the question whether a freedom of will exists. The observation that a great part of our behavior is controlled and directed by instincts raises the reasonable question, whether what we consider freedom of thought and will, is a self-deception. In order to discuss this question, we have successively developed arguments for and against the existence of freedom of will.

We have started with a thesis in favor of the freedom of will. We said that all theories, which reduce the functioning of the world and therefore also of the mind, to a deterministic model of some kind, overlook the fact that a model is like a photograph. No matter how great its resemblance to an object may be, it cannot be considered as being identical with this object. On the other hand, we have noted that the credibility of a theory, of a model of reality, depends on the existence of an uninfluenced observer, who will examine and verify this model. If no free and independent observer can exist, then no guarantee for the trustworthiness of this model can exist. Any model of cognition, which rejects its freedom, refutes its own credibility.

Opposing these arguments, we have said that that determinism is, in a certain sense, inherent in the search for explanations. To each "why" corresponds one or, possibly more, "causes". If the mind is not exclusively moved by inward drives, by outward stimuli and by the hereditary structure of the body, what else is the cause of the direction taken by our thoughts? What does the "freedom of thought" imply? Random choices? But random choices, if they exist, have nothing to do with freedom. They are merely products of chance. We can, just as well, believe that we make free choices when we throw a coin and choose what we consider to be true or false, depending on the outcome of the throw.

Here we have explained that "freedom of will", in reality, means for us freedom to follow the instigation of our instincts, i.e. lack of outward restraint.

To these arguments, we have countered that in the above deterministic way we can try to explain everyday, i.e. repetitive, situations, but we cannot explain unique phenomena. Such phenomena are, e.g. the scientific discoveries, which often change permanently many aspects of our social and even natural environment. Deterministic explanation is a reduction of a complex phenomenon to simple ones, which happen invariably and are considered as basic principles. But scientific achievements are ultimately unique.

Finally, we have ended up with a position, which is not a synthesis, but rather a hyperbasis of the above, opposite views. As we have seen, instincts and drives are modified by experience. They are not merely inborn mechanisms, but often only inborn abilities to build up action and reaction schemata on the basis of experience, or to acquire certain skills. Our drives, our instincts, are not merely determining factors of behavior, but they are also formed by outward experiences, which are partially a matter of personal choices. This means that we are not merely influenced by our surrounding, but rather constitute a unity with it. In this process, there is no unique deterministic direction from cause to effect. We are influenced by our surrounding, but we also influence it. By making choices of actions but also of conceptions, we modify both the actual physical surrounding as well as our view of it.

BIBLIOGRAPHY

Eddington, Arthur: Beyond the Veil of Physics, pp.167-183 in: Ken Wilber (ed): Quantum Questions, New Science Library, Shambhala, Boston & London, 1985.

Rucker, Rudy: Infinity and the Mind, Paladin/Granada, London 1982.

Simmons, George F.: Differential Equations, McGraw-Hill, 1972.

Squires, Euan: Conscious Mind in the Physical World, Adam Hilger, Bristol and New York, 1990.

Appendix 10.1

On the Conception of the Ideas "Truth" and "God". A Mental Indeterminacy Principle

A basic question that naturally arises when we consider such a memory organisation system as the one discussed here is: What is the meaning of Truth as a general conception in such an evolving system?

Is there something we may call "absolute or ultimate Truth" and what is its nature?

This is no futile question, because the belief in the existence of an absolute truth lies behind many endeavours of man. It is a guiding conception that makes our efforts meaningful. We need a guiding interpretation of the world in order to conduct our life, to give it goals[50].

Thus we are lead to the basic question: What is the nature of Truth? What is the nature, not of any specific truth, but of the truth as a coherent set of ideas about the world?

1. The Various Kinds of Reality and Truth

Truth is often defined as correspondence of the outward reality to our internal model of it. However, this definition is incomplete, because a large part of what we call "reality" is an interpretation of what we perceive and not perception itself.

For instance, we often say that a staircase is ascending to the next floor, while somebody staying on that floor will say that the same staircase is descending towards our floor. The staircase is, of course, doing neither of these things. It is merely connecting the two floors, but we choose to interpret this structural element of the building the way we would use it.

Thus, there are various kinds of reality. The actual, the original reality is the biological one. It is what we perceive with our senses. According to it some bodies are solid with no openings or cracks in them, some surfaces are smooth, a strung string follows a straight line etc.

Another "reality" is the physical reality, which we reach when we try to make logical sense of what we perceive. This reality may some times contradict the biological reality. For instance, it tells us that the earth is not basically flat and that the sun does not turn around the earth every day, but rather the earth turns around itself.

[50] In this sense Truth is self-existent, i.e., it is postulated as a Platonic "idea". However, we need this conception of an ideal in order to be able to conduct a meaningful life, no matter what features we finally give to this absolute truth.

Biological reality dissolves when we look at very small or very big objects, so that we have to use some apparatus instead of using only our senses.

Through a microscope we can see that a “flat” surface is not so flat as it seemed. It has cracks and bumps in it. There are also tiny organisms living on it, which we never suspected to be there. Through a telescope we discover that the tiny spots of light in the night sky are distant planets and even “suns”.

These facts or “realities” are implied realities, because they do not merely refer to what we directly perceive, but to an interpretation of it. They rely on trying to make sense of what we see, as when we conclude that the earth is basically round, because when we travel by ship the coast disappears beneath the surface of the sea, before it becomes too small to be seen because of the growing distance. They also rely on the assumption that the apparatus used does not distort what it shows us.

Biological reality has, thus, some times to be revised in order to make better sense of what we see. Physical reality is also time and again revised by making a new synthesis of old observations in order to make them consistent with new ones.

Thus, reality is not merely what we perceive with our senses, but a composition, an interconnection of biological and physical observations and interpretations.

2. The Reality as a Composition of Observations and Interpretations - Examples

Another example of considering rather an interpretation of the observable facts rather than the actual observations as real, i.e. true, is the general admission of the heliocentric system.

Everybody considers nowadays as a fact that the earth turns around itself and moves around the sun, although very few could give persuasive arguments that this must be so. Most people trust in this respect what is claimed by schoolbooks. However, what we believe in this case is exactly the opposite of what we see with our eyes, since every day we see the sun rise and set.

Similarly, all evaluations of our experiences (what is beautiful or ugly, what is just, proper etc and what is not) depend very much both on personal choices and on cultural tradition.

Many aspects of truth are, thus, actually interpretations of experience and express more what we expect to be true, rather than something that precisely corresponds to reality. What is more, even what scientists hold to be true is only a model of reality. It may be based on observations, but does not merely describe them. It interprets them. Often, it interprets them in an way that does not agree with biological reality.

For instance, Newton reduced gravitation to the existence of attractive forces between all material bodies, which are supposedly acting from afar, without any medium that interacts with these bodies. This is, actually, strange from the point of view of biological reality, since our first experience with bodies is that we can not move them unless we push them with our hands or feet.

On the other hand, Einstein claimed, even more abstractly, that there are no attractive forces but merely distortions of the space-time geometry in the vicinity of matter. Finally, quantum physicists say that forces, or rather, force fields are carried by corresponding particles, which they call "elementary" although they have yet no explanation for their diversity[51].

These particles seem to remedy the strange view of action from afar, but they are not less strange themselves, since they are also waves or they are accompanied by "probability waves" (of being at a certain place). These waves are purely mental and have no material existence, since probability is an abstract mathematical concept. What would you think, for instance, about a notion, that there are "beauty waves" somehow moving through space?

All these different views are, obviously, interpretations of the observations, which were accepted as "true" in their time, or are today considered as "true", because they gave or give a parsimonious enough description of all available data.

3. What is Absolute Truth?

Thus, absolute truth is not easy to define. What we consider to be the absolute truth is not something fixed we only have to gradually discover. Well established convictions may suddenly be overthrown by new experiences and the partial models of reality must be extended or modified in order to include new data. The initial model of a, basically, flat earth which is created by the experience of plain stretches of land, is overthrown, for example, by the observations we make when we travel by ship. During the departure, for instance, we see that the land disappears

[51] Today's Physics is like the Chemistry of the 19th century, which distinguished various elements without being able to say why they were different. Similarly, Physicists postulate at present the existence of various elementary particles without being able to say why they are different.

How can we consider such particles as truly elementary, if we equip them with different properties without giving an explanation for this diversity?

We may postulate that all matter has the property of gravity considering this property as an essential feature of matter. But why are some material particles positively charged, while others are negatively charged? What is charge anyway?

Of course any theory must begin with some basic assumptions. It can not proceed by postulating nothing at all. But the mere admission of the existence of a dozen or so different elementary particles seems quite complicated.

These questions are mentioned here, in order to show that Theoretical Physics can still raise, if it wishes, new questions, beyond the, for many decades sought, unification of General Theory of Relativity with Quantum Theory.

gradually, first the coast and then the tops of the mountains, before they grow so small, that they cannot be seen. This means that the surface of the see is curved and thus the various elements of the landscape disappear gradually behind it.

However, we are compelled, in any case, to search for an all embracing "truth", because only then we reach certainty.

However, "Truth" does not mean for most people only a composition of all present scientific views, no matter whether they are actual observations or interpretations of reality. It rather means the integration of all life experiences (including scientific views, but not only those) into a coherent whole. This integration incorporates, for instance, among others one's philosophy of life.

4. Why People keep on looking for Truth

People permanently try to build up such an integral model of "reality", no matter how difficult it may seem. There is a good reason for this: As said before, we need a guiding interpretation of the world in order to conduct our lives.

We keep on trying to reach an all embracing conception of truth, in order to grasp the essence of the world, because we need to know its nature, the actual meaning of its existence, as well as our existence! We strive, thus, for internal consistency in spite of the fact that we also tend to maintain various illusions in our life, as, for instance, with respect to our political ideology or our relations to the opposite sex!

5. Is there a Permanent Truth?

But is there any permanent truth in the world? In view of what we have said, it seems futile to look for such a truth. Our model of the world as a whole is indeed continuously changing, but not arbitrarily. It is evolving.

A basic issue here has been to stress the importance of "evolutionary procedural semantics". This means, however, that the conceptual system is an evolving entity. Grasping of new ideas, e.g., creating new scientific theories, is to a great extent based on a process of concept-evolution.

From this point of view, discovering new truths is also a never-ending dynamic process. The ultimate, the absolute truth is, therefore, something we can never reach. This is not due to any limitations to discovering new facts, new aspects of reality or to the fact that these aspects change arbitrarily. It is simply due to the fact that these aspects are truly infinite. New aspects of what we consider to be real are discovered time and again, while in some instances the whole system of ideas concerning a specific field of investigation needs to be reorganised, restructured, in order to make the concepts compatible with each other.

6. The Evolutionary View of Reality and the Ultimate Truth

Reality, as we know it, seems to incorporate only a finite number of aspects or relations. But reality as an evolving entity (evolving together with our conceptual system) is transcendental and inconceivable as a whole.

The ultimate truth is transcendental, because it lies always just beyond the rational. It is something that we can not rationally grasp in its entirety. It is like the numbers $\sqrt{2}$ or π, whose decimal digits we can calculate to an ever greater extent, without ever reaching an end. That is why many people call this ultimate, transcendental truth "God".

7. The two Views of God – Outward and Inward God

God is not something we can mentally grasp in any way. This ultimate truth is, in some sense, the "resultant" of all laws of nature we can ever discover. We can neither visualise God nor discern in him (or it) various aspects or parts, because he (or it) is the rationally inconceivable. Nevertheless, this inconceivable, evolves in us, just as the mathematical concepts evolve, revealing time and again new aspects of them. This is how our mind works.

The motor of this evolution are the drives or instincts. They motivate our mind and form, thus, our mental system. They are, therefore, the basic creative powers, which guide our lives, i.e. our feelings and actions. They determine and form the system of values, which directs our actions. They are an absolute system of reference, which we can not ignore without being punished by mental suffering, or even mental illness.

However, they are usually unconscious, exercising their influence without any explicit linguistic expression. This influence can be immense, overruling some times our usual self, our usual personality, and upsetting our carefully laid plans. They seem, thus, some times, like independent personalities acting in us and having immense strength.

The composition of all these internal creative powers into one whole eventually manifests itself (is experienced inwardly), therefore, as an immense personality revealing to us unexpected aspects of reality and new values. Mystics call this personality "God".

I.e., "God" is for some people an inward experience and not a mere idea. Apostle Paul, for instance says: "It is fearful to fall in the hands of the living God", (letter to the Hebrews, 10:31), obviously reflecting on personal experiences like his vision on the way to Damascus (Acts, 9: 24&25), while

Socrates spoke of his "daimon", which inspired and guided him (see e.g. [Russel, 1979, p.107]).

8. The Duality of Personal and Impersonal God. The Impossibility of Choosing one of these Views

From an outward point of view, God is the Transcendental, whose aspects we keep on discovering, giving it a rational expression, but which can never be exhausted in this way. From an inward point of view God is the resounding of the entirety of this transcendental truth in our mind, which guides our life and whose inspirations or instigation we try to fulfil.

This may seem to be a mechanistic explanation of the idea of God, but it is not so, because we do not know and can not know what sets this whole mechanism in movement.

In some ways we are in the same position as somebody, who sees on a television screen a landscape like the tops of some trees and the sky above them, which suddenly start being stirred and distorted. He may then realise that he was not shown the landscape directly, but rather its mirror image in the still waters of a lake. When the wind stirred the surface of the lake the image of the landscape appeared also to be stirred.

However, in our case there is no way of telling whether it is the evolution of our mind that creates the change of the whole overview, or something outward is doing it. We have no frame of reference.

We can not sweep around the television camera and find out that we have been looking at the surface of the water. We can never tell, whether there is an immense personality acting in us and on the world, or merely a mirroring, a reflection, of our own evolutionary mental processes on the world that creates this impression.

A Short Review

In this appendix we have discussed the question whether there is something, which we may call "Absolute or Ultimate Truth" and what is its nature.

Initially, we have noted that this question is not futile, because we need a general conception of the world, an interpretation of it, which will allow us to evaluate all situations and choose our goals. Thus, we were lead to the question what is the nature of this Truth, i.e. not of any specific truth, but of Truth as a coherent set of ideas about the world.

Starting with the, often given definition, that Truth is a correspondence of the outward reality to our internal model of it, we saw that it is incomplete. A great part of what we call "reality" is an interpretation of what we perceive and not the perception itself. For instance, the visual reality of tiny light spots on the night sky explodes into millions of "suns" when we look through a telescope. This is a derived reality, an interpretation of the initial one, based on our trust in the fidelity of telescopes. Discussing various examples we have concluded that reality is not simply what we perceive with our senses, but a composition, an interconnection of biological and physical observations and interpretations.

The fact, that all these observations and interpretations are constantly changing, has lead us, then, to the conclusion that Truth as a unified view of the world evolves continuously. The conceptual system, as we have presented it here, is an evolving system and reaching new ideas, e.g. the creation of new scientific theories, is greatly based on a process of concept evolution. From this point of view, therefore, the discovery of new truths is a never-ending dynamic process. Absolute Truth is, thus, something we can never reach. This does not happen, because there are any limitations in discovering new data, new aspects of reality, but because these aspects are truly infinite.

Reality, as we know it, seems to incorporate only a finite number of aspects or relations. But reality as an evolving entity (evolving together with our conceptual system) is transcendental and inconceivable as a whole. The ultimate truth, we steadily seek, lies always just beyond the rational. It is like the numbers $\sqrt{2}$ or π, whose decimal digits we can calculate to an ever greater extent, without ever reaching an end. This is the reason why many people call this ultimate, transcendental truth "God". God is not something we can mentally grasp in any way. This ultimate truth is, in some sense, the "resultant" of all laws of nature we can ever discover. Nevertheless, this inconceivable entity of truths evolves in us together with the conceptual system.

The motors of this evolution are, as we have seen, the drives or instincts. They motivate our mind and form, thus, our mental system. These are the fundamental creative powers, which guide our mind and determine our system of values. We can not ignore them without being punished by mental suffering or even mental illness. Sometimes, we feel them as independent personalities, which act in us overpowering our usual self, sometimes with catastrophic consequences.

However, in some people they manifest themselves as a unique immensely powerful and creative personality, which speaks inwardly to them and reveals unexpected aspects of reality and new values. Such mystics call this personality "God". "God" is for them an inward experience and not a mere idea.

Thus, we ended up discussing the question, whether God is a mere idea, the integration of all possible laws of Nature, or an immensely powerful personality, which some times speaks inwardly to us and reveals new aspects of truth. We saw that there can not be an answer to this question, because we do not have an outward point of observation of our mental phenomena, an outward point of reference. This means that God is conceived by the mind both as an impersonal as well as a personal entity.

BIBLIOGRAPHY

Russel, Bertrand: History of Western Philosophy, Unwin Paperbacks, London, 1979.

Table of Contents