DECISION-MAKING STRATEGIES
IN THE
GENERAL PRACTICE

(BESLUITVORMINGSPROCESSEN BIJ ALGEMENE ARTSEN)

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE

GENEESKUNDE

AAN DE ERASMUS UNIVERSITEIT ROTTERDAM

OP GEZAG VAN DE RECTOR MAGNIFICUS

PROF.DR. A.H.G.RINNOOY KAN

EN VOLGENS HET BESLUIT VAN HET COLLEGE VAN DEKANEN.

DE OPENBARE VERDEDIGING ZAL PLAATSVINDEN

OP WOENSDAG 15 OCTOBER 1986 OM 14.00 UUR.

JACOBUS RIDDERIKHOFF

GEBOREN TE HOORN

PROMOTIECOMMISSIE

PROMOTOR:	PROF.DR.H.J.DOKTER
PROMOTOR:	PROF.DR.A.P.J.ABRAHAMSE
OVERIGE LEDEN:	PROF.DR.P.J.van der MAAS
:	PROF.DR.IR.H.W.van den MEERENDONK

CONTENTS

PREFACE

My first question when I entered medical school as a teacher was: "What shall I teach?" The answer I received was not as clear as I wanted, but wrapped in some (to me) mystical concepts like "medical competence" and "experience", the first one only to be achieved by the other.

"Experience" could only be reached by intensive practicing for several years (and a lot more years thereafter). "But", as I asked myself, "when competence and experience are so thoroughly linked, is it for junior physicians a logical impossibility to reach the same degree of clinical competence as the more experienced doctors?"

The training in clinical competence has not essentially changed throughout the centuries. Formerly, the apprentice turned to a master, a well-known example of his profession, a so-called expert, and received his lessons by closely watching the skills of his teacher. Nowadays, students go to a medical school, which contains a crowd of experts, where they get a basis of practical and impractical knowledge, and are assigned to a teacher who tells them to closely watch his (or her) skills and attitude. The training for these skills and attitudes is performed on patients placed at the students' disposal by the teacher.

Measurement of the clinical competence is accomplished by making a comparison between the knowledge, skills and attitude of the student with those of his teacher (or peer-group).

There is a growing tendency to disagree with this procedure. It does not e.g. take into account the teacher's competence, neither as a practitioner nor as a professional teacher; it assumes problem solving competences invariable over time; and it does not explain - nor test - the mental strategies along which physicians will arrive at a decision.

However, for teaching programs and procedures, we need knowledge and insight into the reasoning processes of experienced (and non-experienced) physicians. Without this insight we may, with a famous word of Mager, "very well end up somewhere else, and not even know it."

The intention of this study is to investigate the methods physicians use in solving the patients' problems. A myth in medicine tells us that there are as many methods in medicine as the world counts doctors. To make up an inventory of all these methods is a Sisyphean labour, appalling and discouraging most investigators. Disagreeing with this myth I wondered why physicians all over the world so easily recognize each other: problems, methods, and contexts. Can it be that physicians have more in common than they want to acknowledge? Moreover, why would physicians differ in their methods of decision-making from other people when placed in comparable circumstances? Although often assumed otherwise most physicians are normal human beings.

If a general framework for reasoning and/or decision making processes could be modelled, I could investigate whether the physicians' practical clinical work fitted into this(ese) model(s). The theoretical framework for this study has been derived from the two types of scientific reasoning: deductive and inductive inference. These two types were rephrased into the usual medical jargon in order to establish congruence.

The discovery of the exclusive use of the inductive inference strategy yields several consequences. The main consequences descend from the typical characteristics of inductivism among which are: the personal character of the physician's medical knowledge; the personal - and, therefore, - incomparable - conception about several common medical items like symptoms, signs, diagnoses, treatments, and epidemiological estimates; the irretraceability of the mental

processes in medical problem-solving, and the myth of enhanced expertise with increasing experience.

It also means the questioning of medicine as it stands, medicine as it relies on its traditional classification of diseases; as it maintains the suggestions of a unified conception of diagnostic entities; as it suggests a joint concept of particular therapies for specific diseases.

The study also challenges the feasibility and the possibilities of normative medical decision making to the routine of medical work.

But, above all, the study pictures the physician as an ordinary human being, as an imperfect decision maker, "satisficing" rather than "optimizing", who is kept within the constraints of uncertain situations, medical ethics, and his personal emotions. These characteristics the physician has in common with many decision makers like economists, politicians, managers, and everybody facing situations of uncertain choices. Insight in these processes may lead to an understanding and improvement of the medical problem solving and decision making.

If this study could in any way contribute to this aim, it might prove to be of value.

Behind the scenes two people especially contributed greatly. Without my wife's shrewd observation of my habitual working pattern as a general practitioner and her astute explanations, this study may have failed from the beginning. In combination with her domestic and relational sacrifices during the eight years of this study she really showed to be an invaluable comrade at arms.

To Hans Geilenkirchen I owe my sincere gratitude for his assistance in the elaboration and ultimate completion of this study. His intriguing questions and his inspiring thoughts were of decisive help for the accomplishment of the task.

I wish to thank all the participating physicians for their sincere co-operation, and furthermore all friends and relations who provided me with their valuable advice.

I am very grateful for the financial support of the "Praeventiefonds" which made this investigation possible.

CHAPTER I

BY WAY OF INTRODUCTION

La preface d'un livre scientifique en est une partie necessaire; elle satisfait a cette question. Quel est cet homme qui vient me faire des recits? (Apres Stendahl).

Physicians solve patients' problems. It is their duty, the goal of their daily work, some even speak of vocation. But how the problems are solved is wrapped in mystery. Some people call it art, others consider it as a scientific process. But the real questions essentially are:

- do patients benefit from their doctor's advice?; and
- do physicians know what they are doing?

The first question is repeatedly stated in politics and in the press. It is noted that in the United States there is about the highest expenditure per head of population on health care but this country ranks about 15th among the healthiest nations in the world. Yet, in Japan, where so much less is spent on health care, there are statistics showing the highest life expectancy, the lowest infant mortality, and fewest deaths from heart disease. It was observed that the Dutch health status was higher than the Belgian, although Belgium counts many more physicians.

To state this question in terms of these statistics is putting facts in a wrong focus. Populations are no patients; physicians treat patients, no communities.

Life expectancy and mortality rates do not satisfy personal conceptions about health.

The discrepancy between the personal conception and the population's concept of health can be illustrated from a biological point of view. From this standpoint civilization implicates the same disadvantages as the domestication of our animals. Lack of physical exercise, inappropriate nourishment, and a number of subsequent causes led to a malfunctioning of several organs of animals and humans. It diminishes the individual's resistance to several diseases and ailments.

In 15,000 years caries developed from zero to more than 98% of the global population. The second factor is the introduction of the doctor. In the Mesolithicum when the human race generally came to choose a fixed dwelling-place and started agriculture, the physician entered the scene. From that moment on the number of diseases increased rapidly. Causes and consequences will be left to your imagination.

The physician's interference gave the diseased people the possibility to advance to the age of procreation and pass the disease to next generations. In natural circumstances this development would be disastrous to a herd of animals, threatening it with extinction by the maintenance of larger groups of defective animals.

Biologically speaking, recovery of the patient is most favourable to the doctor. The morbidity of diabetes mellitus, rheumatism, epilepsy and many other diseases increased 5, 10, 20 and more fold in a couple of centuries. However, what to do about it. Maybe it is human's tragedy that doom is cast over one of human's best qualities, his compassion to the poor and the weak. Or in the famous words of Nietzsche: "An deinen Tugenden wirst du zugrunde gehen" (you'll go under because of your virtues).

This illustration typifies - in an extreme way - the contrast between the

health status of a population (growing worse the more health care is applied), and the health care to the individual citizen (benefiting from a superfluous health care).

The former item typically being the concern of politicians and administrators, the latter one concentrating on the pains and the sorrows, the well-being and the illnesses of the single human being. Both pursue the same goal: health, but, as we have shown, they do not frequently happen to coincide.

In this study we have focused on the individual: the physician solving the patient's problems, which in turn is his, the physician's, problem.

Our question, therefore, is:

- "How do physicians solve problems?", with the related questions:

- "Do they use special methods, strategies, in order to reach a solution?"; and

- "Can these strategies be formalized?"

The question is often raised whether medicine is an art or a science (1, 2). Or, in our context, whether the problems are solved in an intuitive, personal way, or by the sheer force of logical reasoning. Must we allude to Kuhn's opinion that medicine is a craft, like calendering and metallurgy, or do we have to share Braithwaite's opinion that medicine is a science because of its "natural domain"?

Science is a method of acquiring exact knowledge by observation, experiment, and measurement leading to the formulation of a hypothesis that will explain the relationship between the facts observed (3).Art, in contrast with science, is harder to define precisely and concisely. Like science, art is based on experience, but the experience is subjective, incapable of precise analysis, irreproducible, and impossible to measure. But whereas science is completely impersonal, art is intensely personal.

From the preceding lines we may think that the physician, generally, is more attracted to art than to science: preferring the personal touch to the impersonal one. But such thoughts lead to confusion. If medical problem solving is an art, and, therefore, subjective and irreversible, how can we possibly know what the physician is doing? When the activities and actions of the physician cannot be logically justified, can there be any justification for this professional doings?

Generally this question is answered by pointing to the - often huge - experience of the physician. Popper presented a patient case to Adler. Although the case was quite unfamiliar to the great Viennese psychiatrist, he easily explained into details the inferiority complex of the boy. Popper asking Adler how he could be so sure, got the reply: "Because of my endless experience."

Although this answer is very much recognisable to many physicians, especially the experienced ones, it does not lead to a solution because of our ignorance about the real nature of experience. When the physician's knowledge is - largely - based upon experience, then we can ask, with the Scottish philosopher Hume (1711 - 1776), "how do we know that the cases we have not experienced resemble the ones we have?" How could Adler know for sure that the presented case resembled the ones he had met in practice? How can anybody predict from previous cases the succeeding one? How can we conclude from a certain experience anything else than its similarity with some previous cases?

Without proper knowledge of experience and its related method of analogic reasoning, we again face the question: "Do physicians know what they are doing?"

Many people assert the problem-solving in medicine to be a scientific one: the (hypothetico-) deductive strategy (e.g. 4, 5, 6). If deductive strategy is regarded as a method of deducing detailed hypotheses from universal statements, deduced in a strict logical, reversible order, we have some doubt. During my thirteen years of family practice, I have never found myself solving problems in a strict logical, deductive way. The more doubts I had the more curious I was to know and to understand how my colleagues managed to cope with that apparently endless stream of complaints, illnesses, and diseases. How do they solve their

problems?

For obvious reasons I refused to admit that it was I who was the exception.

However, we have to assume some method which enables physicians to reconstruct their reasoning process in case of error. A method which enables them to judge their problem solving superior in cases of correct diagnosis. Denying this construction means that nobody will be able to tell the physician what there is to learn, or even whether there is anything for him to learn. He may learn that what he and/or his colleagues think the correct diagnosis and treatment should be, but this is only learning about what an other person thinks that he knows, (but not about the patients) (7).

Brehmer sees this problem as the old controverse between inductive and deductive reasoning. "Inductive judgment cannot be justified in the way deductive judgment can be justified. The judgment may very well be true, and serve as a guide in action, but it cannot be shown to be true. The fact that it works says little about its truth; it just tells us that it works, and the explanation why it works may be very different from what we think it is."(7).

When our judgments cannot be justified, how can we find out whether the patient benefitted from our advices and prescriptions? We can ask the patient but the odds are favourable when the patient recovered from the disease, and the reverse is true in cases of the opposite effect. These reports are grossly biassed. Nevertheless, we are convinced that our skills and knowledge grow with experience. We tend to believe that diagnostic skill is an innate trait of experienced doctors that is generalized across clinical situations. Physicians are expected to profit from their experience, from their successes and from their mistakes. However, to learn only by trial and error, from one's own mistakes, would be a slow and painful process and unnecessarily costly to one's patients (8).

Medical education is partly based on experience. In the vocational training for family physicians emphasis is predominantly placed on experience. The student or the junior physician is placed at an experienced physician-teacher in order to learn the 'va-et-vient', the 'ins and outs' of medical practice. What can be expected from this teaching when the teacher, in the conception of Brehmer, cannot explain the processes underneath the problem-solving structure? Following Polanyi's Stages of Learning (9), only the third one can contribute to real learning.The three stages are:

A: Trick learning = discovering useful means-ends relationships;

B: Sign learning: Observing sign - event relationships: relies primarily on perception;

C: Latent learning: occurs when the process of reorganisation is achieved not by a particular set of contriving or observing, but by achieving a true understanding of the situation, which has been open to inspection almost entirely from the start.

Thus trick learning, like the performance of human skills, is more completely controlled by purpose than sign-learning, which like connoisseurship, is primarily the achievement of attention. However, the stage of latent learning, which is the purpose of education, can only be achieved when the situation can be clearly defined. If this is not possible, we are almost unable to know if the process of reorganisation has taken place into the direction we wanted. Confronted with the family physician's vocational training I was forced to asked myself: "What do we teach?" and "How do we teach?"

Do we teach them tricks or is our goal a real understanding of contents and methods of family medicine. Asking the question is answering it. The first and principal problem, therefore, is to establish methods or strategies as they are used in the general practice.

Preliminary to this study I proposed an investigation about the the actual behaviour of family physicians confronted with a case of myocardial infarction. Joint cardiologists and family physicians have agreed upon a common procedure in these cases. A study about predicting symptoms and signs of myocardial infarction (10) and a device to weigh these symptoms (11) developed from this

study had been published, respectively distributed at that time. So, with a fixed treatment protocol and a ruler to ascertain the probabilities of the symptoms and signs at hand, it was expected that methods among family physicians were more or less uniform. An investigation team interrogated 10 family physicians in order to investigate the methods the participants employed when confronted with a case of myocardial infarction.

The results were shocking. Although all conditions for a uniform strategy seemed to be fulfilled, any consistency among the methods could not be found. Each physician interpreted the predicting symptoms and signs and their weighings in a very personal way. Half the physicians followed the recommended treatment upon which agreement had been reached; the other half deviated from the plan.

We asked each of the 10 physicians to describe the most characteristic picture of myocardial infarction as they had memorized it in terms of symptoms, signs, tests etc. The results were a complete surprise. The description differed like fingerprints among a population. Each doctor has his own personal conception about this disease entity, and hardly a single attribute covered the description of this disease by an other, equally experienced, physician.

The finding of inconsistency for diagnosis, diagnostic methods and treatments made me curious about the physician's problem-solving behaviour in cases of other, sometimes less well-known, diseases.

It seems the more interesting because we are still unable to unambigiously answer the question: "Do patients benefit from the doctor's advice?"

To investigate the physician's problem-solving behaviour several problems had to be conquered, like

a) the - theoretical - modelling of the strategies;

b) the highlighting of the landmarks within these strategies in order to recognize and to observe their various behaviours;

c) the choosing of a way of observation which enabled the testee to behave as naturally as possible;

d) the elimination of mutual influencing of physician and patient by standardizing the 'patient';

e) the fact that the 'standardized patient' had to allow for free-questioning throughout the 'patient data base' by the physician-participant.

f) that the types of patient-scenarios had to cover a large part of all clinical disease classes;

g) that the number of participating physicians must enable us to draw conclusions.

h) that we had to find a representative sample from the population of family physicians in the city of Rotterdam.

Lasegue stated: "Doctors observe little, and they observe badly". Being a doctor myself, apparently little hope is left for observing the essentials and characteristics of the medical problem-solving process. Fortunately, several investigators of different disciplines, like psychologists, information scientists, statisticians, philosophers, and even physicians, preceded me in outlining the theoretical and - sometimes - practical aspects of the physician's decision-making methods. Statistical and Artificial Intelligence models have been originated and - sometimes - even put to practice with more or less success.

However, these models represent a logical way of reasoning, which at the moment cannot be contradicted, but seem a bit remote from the day-to-day working pattern of the physician. At least, this is how I feel it. Howell complains that too much focus is laid upon the methods people ought to use, not how they use them. "We should concentrate on what people actually do and develop descriptive models to account for decision processes. This, then, has been the trend of late: from normative to descriptive modelling." (12)

This thought may be suggestive for the conception that models, how precious

they may be, which do not fit the practioner's working methods will not be accepted and amalgamated in the daily routine. If we could be able to find a common denominator, or even better, a model that could actually reflect the problem-solving behaviour of physicians, we may be capable to retrace or even to reproduce the methods in cases of failure or success. It may contribute to understanding and insight into that mystical occurrence called the medical process. It may also contribute as a 'trait d'union' between normative and descriptive models.

The landmarks must come from the elements and their nomenclature as they are usually employed in the actual medical process.Features that are alien to the physician and the medical process, like sometimes met in normative decision making models, cannot contribute to the observation and recognition of the methods employed by the participating physicians. This implicates a scenery and a registration which enables the physician to operate in a way as natural as possible. It includes observation and registration in his own familiar office, to interview the 'patient' in a verbal mode, free-questioning in every direction, adequate 'patient-answering', customary time-limits, avoiding normative judgment etc. Briefly, to set the participant at ease, expecting to proceed in his customary way. We shall have to realize that some stress will remain as a normal feature of all medical problem solving.

The doctor has to make decisions in an uncertain problem situation which involves two systems: his own and that of the patient. From this the main problem arises, because studying the functioning of the physician means: observing these two systems which at the same time influence each other in a special implicit way. It is like focusing on a certain point on a turning wheel while sitting on an other wheel rotating in an opposite direction. One observes flashes only partly recognisable.

In order to investigate the problem-solving process, one has to freeze one of the two systems. Studying the doctor automatically includes a "fixed" patient (and vice versa). A "fixed" patient is a simulated patient: a device which could serve as a patient (with or without the help of an intermediary.)

For our purpose of storing a large number of data, these data must be highly flexible, have to cover a broad area of symptoms and signs and should be easy to produce and manage. Besides, it must fulfil the requirements of realism and above all it must be cost-effective.

The various types of patient-scenarios should cover a large part of the clinical disease classes. They have to deal with a wide range of complaints, illnesses, physical, psychical and social dysfunctions. Especially these latter elements represent a large proportion of the total number of patients visiting the doctor. Eventually, we chose for the clinical part of medicine as the part being most clearly circumscribed. The ambiguity of the psychosocial part may bias the investigation. This feature runs the risk of proving the inadequacy of its registration and its classification. "It is hardly possible to distinguish, at least clearly, between illness and illness behaviour, and complaints of psychophysical origin and complaints of social origin. The mixture of all these elements make us talk about psychosocial complaints and problem behaviour (of the patient). But it is the problem of the doctor not recognizing the appropriate parameters for each of the component symptoms and signs. "(13, 14).

We asked and got the co-operation of 60 family physicians and 8 general internists. We strived for a stratification of the family physicians according to:

- age and sex;
- spread over the research area;
- patient population between 1500 and 3000;
- medical school of study.

Because of the total number of family physicians in the city of Rotterdam (approx. 200) we could not strictly adhere to this stratification, but strived for a sample as representative as possible.

Before we come to the actual description of the investigation and its results (chapters VI, VII, VIII), we shall first make a tour d'horizon of the domains which have contributed to the origination and framework of the study.

Chapter II will describe the domains of the cognitive processes involved in problem-solving. Human problem-solving, under which guise the human thinking processes have been studied for over fifty years, have drawn attention especially by the involvement of cybernetics (Wiener, 1948), concept formation (Miller, 1956, Bruner, Goodnow & Austin, 1956) and computer science (Newell, Shaw & Simon: Logic Theory Machine, 1957), into this field of cognitive psychology (giving rise to Artificial Intelligence Science).

The study was complicated by the observation of the varying kinds of human behaviour in the face of problem solving and decision making. Especially in the face of uncertainty and stress, which often accompany problematic situations, humans employ strategies which detract from optimizing solutions. Overload of information may force people to the use of shortcuts which detract from an optimal solution, but enables the problem solver to 'survive'.

We are confronted with several questions still waiting for an answer. We can observe the inputs and the outcomes but the thought processes in between these items are still wrapped in mystery, apart from some tempting theories. Medical problem solving is mainly treated as a particular entity. Whether the medical procedure really differs from the general pattern as we may meet in human problem solving, remains to be seen. The claim of medical problem solving as a discipline comes from the special features and essentials within the medical world. The medical world has not only to deal with the specific structure of its science but also with the ethics of the profession.

The basic issue in medicine is whether health and disease are universally definable conditions or relativistic phenomena, identified and labelled according to generally accepted classifications, cultural values and social norms.

It appears that medical knowledge mainly consists of the joint knowledge every physician in the world has in his mind. This knowledge, however, is not explicit and not generalisable across situations, cultures, languages and societies. Patient and physician appear to create a unique situation, not transferable to other similar situations. It implicates that the real meaning of a "medical problem" cannot unequivocably be ascertained.

In chapter III we return to the question of implicit and explicit inference, the personal, art-like reasoning versus the scientific, or logical way of reasoning. The latter way is generally adapted as the main approach to scientific questions and studies. It is based on a sequence of logical steps descending from a universal statement to consequence-bearing hypotheses which can be verified or - preferably - falsified. The process can be traced and retraced, and its outcome can be generalized across situations.

The former is the inductive way of reasoning. It employs a different strategy. Principally, it accumulates information in order to elicit - mainly by intuition (induction) or by pattern-recognition (analogy) - a solution which may explain a certain number of elements within the acquired information. It results in a personal statement as the act of appraisal of the component elements or features of the solution is a personal one. However, "this person is taking a risk in asserting something, at least tacitly, about something believed to be real outside himself. Any presumed contact with reality inevitably claims universality" (9). The inductive outcome is believed to be generalisable and can contribute to the "body of experience". The introduction of probabilistic concepts to the inductive way of reasoning does not essentially change the process or the outcome.

In chapter IV I shall describe what I shall conveniently call 'the medical process'.

The medical process can be described as the total sum of processes taking place between the entry and the depart of the patient from the health care system (or parts of it). The features of this process are largely described in their customary order: data acquisition and data processing, diagnoses and their classification, prediction and treatment. These elements are bound together by the strategies physicians apply to solve the medical problem. These strategies are the focus of our study. However, they cannot be understood without the description of the elements to which they pertain. Although the doctor entered the scene of health care approximately 15.000 years ago, these elements are still beyond generally accepted clear-cut definitions.

In Chapter V we shall delineate that, when the medical science itself does not provide its standards, medical decision making as a discipline has to supply its own predefined nomenclature. At the same time it had to redefine and remodel the structure of the medical procedure in accordance to the principles of Decision Theory. By doing so normative medical decision making replaces a rather undefined medical taxonomy and structure by a system still to be proven more effective in ordinary daily practice. The newly developed system allowed for the application of statistical or mathematical (sometimes called actuarial) methods in order to process data in a probabilistic way.

It is my impression that the means, the various and varying statistical methods, govern the ends, the decision making model, instead of the other way around.

Thusfar the decision-making systems suffer from at least three drawbacks:

- the systems are highly constrained to limited domains of the health care;
- the systems are largely untransferable to other clinical situations; and
- the systems are largely rejected by the physicians.

A new development in this field, Artificial Intelligence, suffers from the same or similar restrictions.

Chapter VI delineates the different models of medical reasoning. They reflect largely the methods of scientific reasoning as we discussed them in chapter III. The deductive inference method has been adapted in minor details for the 'medical deductive strategy' (the term hypothetico-deductive will be avoided because of its tautology).

The inductive strategy is modelled according to the thought I have perceived and conceived from the literature. It is mainly based on analogical reasoning; analogies between remarkably few data of the present patient and memorized patterns of similar cases. The method is based on the verification of whether some pattern fits the particular data pertaining to the patient.

The end-point of the diagnostic process is constituted by two elements: the weight the physician places on a hypothesis of choice to be a diagnosis; and the minimization of his personal feelings of uncertainty with regard to the solution of the problem. The weight, or the posterior probability, to a hypothesis can be viewed as a proportional value of the number of acquired symptoms relative to the number of symptoms within a memorized pattern.

The physician's appreciation of the amount of stress as experienced within the problem solving process is reflected by his perseverence to search for a solution as well as his relief (minimization) to the uncertainty when he is convinced having arrived a solution.

The landmark of the deductive reasoning is constructed from observing the progression down the hypotheses-levels with increasing sharpness of circumscription of disease entities. The recognition of the inductive reasoning is mainly marked by its typical process of cycling around the construction of a 'base of hypotheses' during the course of the work-up.

Chapter VII. To identify the methods and strategies a number of conditions had to be fulfilled, as we discussed already. To me, the most important condition was the construction of a simulation model that could fulfil most of my wishes: to simulate the routine work of the physician and to create an illusion of reality.

A new simulation (R-model) was created on the basis of symptom configuration and classification. The symptom is defined as "every functionally objective or subjective phenomenon which can be observed directly or indirectly and which is indicative of either an illness or condition of the patient". This classification enabled me to meet the demands of the self-imposed task for reality. This sphere of reality was underlined by the use of the verbal mode for the physician-'patient' interview. The video-taping of the scenery did not really disturb this sphere of reality.

Chapter VIII presents the results of this study. To create a review, the chapter is divided into 6 paragraphs, each representing one or more of the characteristics of the process and their interdependencies. After the introduction paragraph 1 describes the main objective of this study, the deduction of the methods of problem solving employed by the participating physicians. These strategies are matched with the two groups of physicians, family physicians and general internists. They are also matched with the different patient cases in order to discover (or not) presumed problem-orientation of the problem-solving process.

Paragraph 2 elucidates the role and the meaning of the hypothesis in the process. Its relation towards the levels of hypotheses-refinement, symptoms, diagnosis, and cycles will be elucidated and - as far as possible - explained.

Paragraph 3 tells the story about physicians' questions, patients' answers and the time-constraints within which the operations take place.Special attention has been paid to the interdependence of these variables,and to the effectiveness of the physician' interviewing in raising relevant information.

Paragraph 4 shows the variability of the subjective estimates.Special attention has been paid to the estimations of some elements as they are used in medical decision making strategies.The application of Bayes'Theorem in Medical Decision Making requires knowledge of the Prevalence (or Incidence) rates of diseases,and the weighted disease - symptom relationships.However, by lack of reliable values for these items, decision making engineers rely mainly on the subjective estimations of (expert) physicians.

In this study we asked the participants to give their estimations with regard to these elements. In this section we shall discuss the results.

In the same section physicians' estimates of their feelings of uncertainty or stress during the case work-up are presented. As we may remember, our hypothesis of the problem-solving end-point predicted a minimization of the physician's uncertainty.

Paragraph 5 delineates the various kinds of actions physicians take after establishing the diagnosis. The various variables are matched to each other and to several related ones. The finding of a treatment plan already being incorporated into the diagnostic part of the process is a striking phenomenon.

Paragraph 6 concludes this chapter with the presentation of several consequences of the applied strategy(ies). The role of the personal appreciation of diagnoses,their interjudge (in)consistency and the effects of experience, and continuing education are shown in this paragraph. Some prudent conclusions about

the effectivity of the strategies will be given.

We shall try to present a description of methods and features of the processes physicians operate in trying to solve the clinical problems of their patients.

This study cannot provide an ultimate and final statement about physicians' problem-solving behaviour, but it may give rise to criticism, argumentation and discussion; it may even bring new thoughts and new inspiration to penetrate into that often mystical domain of science called medicine.

REFERENCES

1. Forstrom L.A.: The Scientific Autonomy of clinical medicine. J.Med. Philos., 1977, 2, 8 - 19.
2. Munson R. : Why medicine cannot be a science. J. Med. Philos., 1981, 6, 183-208.
3. Pickering G., Therapeutics. Art of science? JAMA, 1979, 242, 649-653.
4. Elstein A.S., Shulman L.S., Sprakfa S.: Medical Problem Solving, Harvard University Press. Cambridge (Mass). 1978.
5. Ledley R.S., Lusted L.B.: Reasoning foundations of medical diagnosis. Symbolic logic, probability and value theory aid our understanding of how physicians reason. Science 1979, 130, 9-21.
6. Medawar P., : Scientific method in science and medicine. Persp. Biol. Med., 1975, 345 - 352.
7. Brehmer B., : In one word: not from experience. Acta Psychologica, 1980, 45, 223 - 241.
8. McIntyre N., Popper K., : The critical attitude in medicine: the need for a new ethic. Br. Med. J., 1983, 287, 1919 - 1923.
9. Polanyi M., : Personal knowledge. Towards a post-critical philosophy. London, Routledge & Kegan Paul, 1958.
10. Does, v.d. E., Lubsen J., : Imminent Myocardial Infarction Rotterdam. Dissertation, Rotterdam, 1978.
11. Lubsen J., Pool J., Van der Does E., : A practical device for the application of a diagnostic or prognostic function. Meth. Inf. Med., 1978, 17, 127 - 129.
12. Howell W. C., Fleishman E.A., eds.: Human performance and productivity. Information processing and decision making. Hillsdale, Erlbaum Associates, 1982.
13. Suchman E.A., : Social Patterns of Illness and Medical Care. J. Health & Human Behavior., 1965, 6, 2 - 16.
14. Suchman E.A., : Stages of Illness and Medical Care. J. Health & Human Behavior, 1965, 6, 114 - 128.

CHAPTER II

PROBLEM SOLVING

Human thinking processes have been subject to many theories and philosophies, from the earliest days of mankind until the present day, from Socrates till Artificial Intelligence. Still, we only have a scattered picture of their true nature. Although the notion that (cognitive) psychology could become a science of observation, and experimentation had been stated clearly and explicitly by the British philosopher John Stuart Mill (Logic, 1843) it required a person who really knew how observation and experiments are made to bring it off. The real founder of this discipline can be considered to be Wilhelm Wundt (1832 - 1890) who opened the first laboratory of experimental psychology at the University of Leipzig. His successors at the University of Wurzburg applied "introspective" methods to higher mental processes. Members of this school argued that much thinking goes on below the level of consciousness, thus giving rise to "imageless thoughts", as the results of such processes emerge into consciousness. This theory was challenged by Wundt and others. It was argued that images were always present in consciousness. This controversy could not be bridged; the argument between the theories could hardly have been settled by experiment. After the demise of the Wurzburg school, thinking processes were studied in psychology under the guise of "problem-solving!"

In this chapter we will delineate some of the ideas and conceptions about problem-solving, as far as they are relevant to some theories about the medical part of this subject. Paragraph 1 discusses the question "what is a problem?" The next paragraph sketches some concepts about human problem solving.

To enter the world of medical problem-solving, firstly we have to make clear our statement about medical knowledge. Much theory concerning medical science seems rather implicit, or in terms of Polanyi, is 'tacit knowledge'. Therefore, we cannot enter the scene of medical problem solving without a paragraph dedicated to medical knowledge.

The next paragraph must, of course, discuss the question of "what is a medical problem?", followed by the discussion about medical problem solving in the last paragraph of this chapter.

Paragraph 1

WHAT IS A PROBLEM?

The English Platonist Weldon said that there are three kinds of elements that people are concerned about regarding the thinking processes of man. There are troubles which we do not quite know how to handle; there are puzzles whose conditions and unique solutions are quite clear to everybody; and there are problems. We invent a problem by finding an appropriate puzzle form to impose upon a trouble.

Before a person can attempt to solve a problem, he must understand or assimilate a description of the problem. The presentation of a trouble or a complaint does not automatically imply the understanding by the problem-solver. There are at least two conditions to this understanding.

a.) it can be a problem only if it puzzles or worries somebody (assimilation) (1), and

b.) a translation of the problem can be made in such a way that it becomes

a familiar type of problem to the problem-solver, for which he has means within his repertoire for solving (acquisition) (2). The first condition refers to the situation that something is called a problem when it pertains to a task for which the respective subject does not immediately perceive a course of finding a solution. It means that every problem has its individual and personal character. This is reflected in the definition of Thorndyke (1898):

"A problem exists when the goal that is sought is not directly attainable by the performance of a simple act available in the personal repertory; the solution calls for either a novel action or a new integration of available action" (3).

Most people prefer to avoid problems, especially the complicated ones. Politicians confronted with complex, multi-faceted dilemmas will try to satisfy their constituents, rather than struggle with basic solutions to the dilemmas. Judges asked to settle a major legal controversy will often make their decision not directly resolving the substantial issue, but evading it, to reach a conclusion based on matters of technical procedures. Facing a problem means facing the highly personal dilemma of making decisions. And people are constantly being reminded of what they already know from personal experience - that making a consequential decision is a worrisome thing, and one is liable to lose sleep over it. We cannot but recognize and smile about the cartoonist's statement in Peanuts: "No problem is too big or too complicated that it cannot be run away from!" (cited in (4).

It needs a translation to a familiar type of problem to lead people to the phase of solving it. It can give rise to an early formulation of the problem. This formulation is largely based on personal and/or presumed images and prior expectations. Chapman and Chapman (5) demonstrated how prior expectations can lead to fawlty observation and inference. It has been demonstrated that although there were no relationships between cues and criterion statements, most individuals erroneously "learned" the cue-criterion links which they had expected to see (6). This element of foreknowledge has a number of serious implications.:

- it directs more or less the problem-solving. Medin et al. (7) saw more pattern-recognition when there was an assumed resemblance to cases presented a short time before;
- it conceals largely the details of the process;
- it inhibits an evaluation of the judgment processes by preventing a careful consideration of the various options;
- it biasses our guesses.

Most of these elements are usually complete obscure. The individual has no idea which factors influence the process. Sometimes even the fact that the process is taking place is unknown to the individual prior to the point that a solution appears in consciousness.

The Gestaltpsychology which flourished between the world wars stressed the importance of structural relations in perception. It was argued that it was a full grasp of the structure of a problem which led to the insight yielding its solution (8). Although the concepts of Gestalt psychology are largely underscored it does not explain the problem-solver's troubles in defining his problem. There is a notable inclination towards familiarizing the presented case. It is translated and sometimes transformed to a task the problem solver thinks he is capable of tackling. This means e.g. that the physician translates the patient's problem into a task he can understand within the context of the medical science and the health care. Every problem solver facing a problem creates his task in terms of such a task environment.

This task environment is defined by the problem solver, in order to be able to attack it, in terms of a problem space. The problem space consists mainly of states of knowledge about the problem and rules relevant to the various states: e.g. a mathematical problem and the relevant rules to solve it, or, a medical problem and the ordering of the relevant symptoms and signs to elicit the diagnosis. This means that the task (and its environment) mainly determines the characteristics of the problem solving (9). Simon & Newell (10) summarize these

characteristics as follows:

1.) a few, and only a few, gross characteristics of the human information processing system are invariant over the task and problem solver;

2.) these characteristics are sufficient to determine that a task environment is represented as a problem space and that problem solving takes place within the problem space;

3.) the structure of the task environment determines the possible structure of the problem space;

4.) the structure of the problem space determines the possible programs that can be used for the problem solving.

Eventually, this will lead to the formulation of a number of conditions for a problem-solving program or simulation.

a.) it should predict the performance of a problem-solver handling specific tasks;

b.) it should explain how human problem-solving takes place;

c.) what processes are used;

d.) what mechanisms perform these processes;

e.) it should predict the incidental phenomena that accompany the problem-solving process.

In the next paragraph we shall discuss some of the elements of the system and the program.

Paragraph 2

HUMAN PROBLEM SOLVING

Animal studies led to the belief that problems are solved by trial and error. This concept led to a preference for observation instead of introspection. However, as Bertrand Russell (1927) remarked, animals tended to display the national characteristics of the experimenters: Animals studied by Americans rush about frantically with an incredible display of hustle and pep, and at last achieve the desired result by chance. Animals observed by Germans sit still and think, and at last evolve the solution out of their inner consciousness (cited in 8).

But trial and error is, in fact, a rather limited procedure for solving problems, and perhaps therefore, performed on a large scale.

Although most psychologists studied the problem-solving processes from the viewpoint of deductive inference (with an eye on logic), tensions remain between these conscious deductions and everyday inferences. The inferences that underly deductive problem-solving are often slow, voluntary, and at the forefront of awareness: they are explicit.

For practical reasons like alertness, quick reactions, workload etc. this type of inference is scarcely used in routine problem-solving.

In contrast, the inferences that underlie the ordinary process of perception and comprehension are rapid, involuntary, and outside conscious awareness: they are implicit (8).

The distinction between these two types of inferences can still be observed between, e.g. the formal programs in medical artificial intelligence and the routine problem-solving of the health care provider.

What processes do people use to solve problems?

Essentially two major conceptions can be traced.

1.) Sequential: the solver always appears to search sequentially, adding small successive accretions to his store of information about the problem and its solutions (10);

2.) Configural (Gestaltpsychology): The mind has the tendency to organize

and integrate and to perceive situations, including problems, as total structures. The insight that leads to a solution, in the Gestalt view, stems from the perception of the requirements of a problem (Kohler, cited in (3). Productive thinking is not accidental success or the mere application of bits of past experience. The problem has a structure of its own that points the way to its solution.

Both theories have in common the creation of a problem space which encompasses the elements of the problem, its related knowledge and rules and the - perception of a - solution. "Only if we believe that a solution exists can we passionately search for it and evoke ourselves heuristic steps towards its discovery. Therefore, as it emerges in response to our search for something we believe to be there, (...) it will always come to us with the conviction of its being true" (1). In Gestaltpsychology the perception of the solution of a problem is like the perceiving of a hidden figure in a puzzle picture. That means that:

- the perception is sudden.
- there is no conscious intermediate stage.
- the relationships of the elements in the final perceptions are different from those which preceded, i.e. changes in meaning are involved. The implication of this latter view is decisive, because, the irreversible character of a discovery or solution of a problem can only be accredited as a discovery if it is not achieved by following the definite rules. For such a procedure would be reversible in the sense that it could be traced back stepwise to its beginning and repeated at will any number of items (1). While the sequential process can be made explicit following the steps, the configurational one is unique, 'einmalig'. Especially this latter claim is most often mentioned by, not to say typically for, physicians.

Meanwhile, this does not signify that we now really understand cognitive processes. The bulk of the research on problem-solving has been carried out with naive subjects on tasks that do not call for much specific subject-matter knowledge, with scarcely any resemblance to real-life processes. The results of many experiments have given rise to pessimistic reactions. Recently, several cognitive psychologists (Mandler, 1975, Miller, 1962, Neisser, 1967) have proposed that we may have no direct access to higher order mental processes such as the ones involved in evaluation, judgment, problem-solving, and the imitation of behaviour.

Simon (1981) speaks of the simulation of cognition rather than the creation of artificial intelligence. Such simulation of cognition or creation of artificial intelligence contrasts with the modelling of the actual processes of human reasoning. A system that simulates the reasoning of an expert clinician needs only have similar input and output, but a system that models the reasoning process of an expert would also have to employ similar operations in generating the output from the input (11).

Trying to model a system on human cognition we must first learn how human cognition works. This is usually investigated by means of using three types of models:

A: Process studies of how human judgments are actually made;

B: Prescriptive models which make use of a formal theory of how decisions ought to be made;

C: Descriptive models which attempt to describe the context and factors which influence judgment.

Our study can be classified within Category A and C.

Problem solving enables mankind to react and respond to needs, to behave adequately and to survive in an environment that represents unpredictable threats and opportunities. Varying problems in changing environments require different processes to find adequate solutions. Are these qualities inherited or acquired?

The excellent studies by Piaget and Inhelder (12) reveal much of the phases of thinking processes in children. The authors distinguish 4 stages of

thought and reasoning:

1.) the sensory-motor stage of early childhood;

2.) the preoperatory stage: the younger child tries to discover the world without being goal-directed;

3.) the concrete operations stage: here we enter the stage of the schoolchild. The child is already capable of certain logical reasoning processes but only to the extent of applying particular operations to concrete objects- or events in the immediate present.

4.) the stage of formal operations. The principal novelty of this period is the capacity to reason in terms of verbally stated hypotheses and no longer merely in terms of concrete objects and their manipulations. This is a decisive turning point, because to reason hypothetically and to deduce the consequences that the hypotheses necessarily imply (independent on the intrinsic truth or falseness of the premises) is a normal reasoning process. Hypothetical reasoning implies the subordination of the real to the realm of the possible, and consequently the linking of all possibilities to one another by necessary implications that encompass the real, but at the same time go beyond it. The individual who becomes capable of hypothetical reasoning, by the very fact will take an interest in problems that go beyond his immediate field of experience. Propositional logic appears to be one of the essential conquests of formal thought. This fourth period can no longer be characterized as a proper stage, but would already seem to be a structural advancement in the direction of specialization. This stage is reached in different areas at different ages according to aptitude and professional specialization. It can only develop after the scheme of causality has been firmly established, approximately around the ages of 12 to 15 years.

The notion of order is developed prior to that of probabilism. The notions of chance and probability cannot develop until a person has reached the stage of formal operations. Probabilism contrasts to the schema of order, of causality. Not everyone enters the fourth stage. Smedslund (13) maintains that Piaget and Inhelder's stages of developement represent different levels of cognitive functioning at which adults operate. Thus, although adults are capable of using disconfirming information for the inferring of relationships, they frequently fail to do so and operate at a lower, more frequently used level, for example, concrete reasoning.At the concrete-operatory level a structure cannot be generalized to different heterogenous contents but remains attached to a system of objects or the properties of these objects. A formal structure seems, in contrast, generalizable as it deals with hypotheses. It cannot be stressed that disregarding of the thought levels of operation experimental studies may lead to largely varying results and subsequent conclusions.

There is no algorithm for choosing which method to adopt in trying to solve a problem.

Polanyi (1), quoting Poincare's "l'Intuition et la Logique en mathematique", recognizes four stages:

1.) Preparation : acknowledge a problematic situation as a problem;

2.) Incubation : considering the problem by tentative discoveries;

3.) Illumination : practical realization of the principle discovered by the insight;

4.) Verification : manipulations to the test of practical realization.

This is in agreement with Polya (cited in 8) who recognizes also four phases of work on a problem:

a.) Understanding the problem: you have to determine the goal, the data that are given to you, and any general constraints or conditions.

b.) Devising a plan: you have to find some method of connecting the data to the goal. Here, useful heuristics may be to consider any related problems with which you are familiar, to try to reformulate the problem, or to try to solve part of the problem. Often there will be many related problems in which case it can help to concentrate on those with the same, or similar, unknowns.

c.) Executing the plan: you carry out each step, checking its correctness

as you go.

d.) Looking back: When you have solved the problem, you should check the result and reconsider your method. You may be able to discover a simpler procedure, or a more effective one.

Both concepts consider the problem-solving processes as sequential and explicit information processing. Nowadays, information processing can readily be done by technical equipment, the computer. Therefore, we shall have a look at some concepts in this field.

Information Processing

Information processing methodology essentially involves the determination of the thought processes or symbolic manipulation necessary to the performance of a task and their precise specification as a model or computer program (14).

Investigations are specially directed towards the process, on how particular human behaviour comes about, on the mechanisms that enable them. Simon & Newell (10) proposed a program with the following items:

a.) Discover and perform a set of processes which results in a system capable of storing and manipulating patterns to perform complex nonnumerical tasks, like those a human performs when he is thinking:

b.) Construct an information processing language, and a system for interpreting that language in terms of elementary operations;

c.) Discover and define a program, written in the language of information processes, which is capable of solving a type of problems that humans find difficult;

d.) Obtain data on human behaviour in solving the same program as the one tackled by the program;

e.) Search for both similarities and differences between the behaviour of the program and the human object.

On the basis of these recommendations a number of so-called "Expert Systems" have been created. "Expert Systems" are Artificial Intelligence programs capturing the special knowledge of experts, e.g. experienced clinicians, in a given subject matter field and encoded or represented in a "knowledge base", which will be computer processable. Especially the last recommendation of Simon & Newell is in focus because sofar the differences outnumber the similarities. Various causes can be identified among which special attention can be directed towards the question as to how much of an expert's performance is due to skill and how much is propositional knowledge. This is the difference between "knowing how" (Polanyi's Personal knowledge) and "knowing that" (Popper's Objective knowledge) with the latter being the sort that is most readily articulated and incorporated into computer programs. If Polanyi is correct that the portion of "knowing how" cannot be translated into explicit rules, and if a significant part of the expert diagnostician's performances derives from this kind of knowing, then attempts at formally describing it would seem to be in for trouble (15).

Human information processing characteristics are described as:

1.) The system operates essentially serially, one process a time, not in parallel fashion.

2.) The inputs and outputs of these processes are held in a small Short Term Memory (S.T.M.) with a capacity of only a few symbols.

3.) The System has access to an essentially infinite Long Term Memory (L.T.M.), which has a fast retrieval rate, but a slow storage (10).

Considering the limited size of the working memory (S.T.M.) and the slow storage capabilities we must set out the problem in suitable symbols and continuously reorganize its representation with a view to eliciting some new suggestive aspects.

This means e.g. that a physician working on a patient's problem always will translate the presented data in to him suitable symbols like symptoms and signs, and rearrange them constantly in varying clusters suggesting different disease-hypotheses. Therefore, he has to ransack his memory for any similar problem of which the solution is known.

To overcome the limitations of the Short Term Memory, some tricks may be used.

Faced with a decision task of challenging complexity, an individual may try to restucture that task into a simpler one (16). This can be accomplished in two ways: plunder his memory for any similar and known cases, or to combine data in known categories. In the former method the total amount of alternatives is constrained to a limited number of alternatives (17)

The latter method can be divided into three strategies:

I.) Restructure of the problem

II.) Limiting the information

III.) Restructure of information transmission. This latter item will be discussed in greater detail.

I.) Restructure of the problem : dismantling the problem to its bare bone and limiting the number of alternatives. The quality of a person's thinking will be poorest of all in multivalued decisions that require evaluating the consequences of alternative courses of actions in terms of a large set of values, not just in terms of one or two objectives (4).

II.) Limiting the information .

a.) Avoiding or disregarding data about possible negative (or positive) outcomes (18). People tend to focus on the number of true positives, they follow the strategy of using only confirming evidence (19).

b.) Coupling of systems to correlations: limiting the search procedure to only certain types of correlated symptoms, symptoms that either have both typical or both atypical characteristics. Occurrence of one typical and one atypical symptom counts less (7).

Human judgments are often based on comparisons between couples of faced data (20). Medin et al. (7) report that subjects (in the experiments) are looking for correlation, even with less typical information. The subject's diagnoses were primarily determined by the correlated symptoms, although it is not clear whether the correlation per se was used or whether entire symptom patterns were matched (and thus coincidentally correlated symptoms). The use of correlations suggests that the underlying decision process involves more than the weighed sum of independent attributes. This finding is strong evidence against independent cue models, which predict that the number of typical symptoms should be the principle determinant of performance. A question of special interest is, assuming that people learn about each attribute independently, how a category of correlated symptoms has come about. Presumably it is based on experience with copies of a category.

III.) Restructure of information transmission.

This method deals with the framing of the Short Term Memory and the (over)load of information. The restriction of S.T.M. can be viewed as a safeguard to man in order to survive in a constantly fluctuating world. It is far better to have a little information about a lot of things than to have a lot of information about a small segment of the environment. In a world of expanding specialisms this conflict causes considerable tension in people. When the amount of information is increased people will begin to make more and more errors. At

first the transmitted information will increase but eventually, it will level off at some asymptotic value. (See e.g. 21). This latter value is called by Miller (22) The Channel Capacity. This Channel Capacity can be defined as "the upper limit on the extent to which an observer can match his responses to the stimuli that are given to him". According to Miller the Channel Capacity poses an observer the limits to distinguish between alternatives given a number of bits. One bit of information is the amount of information we need to make a decision between two equally likely alternatives: 1 bit distinguishes 2 alternatives, 2 bits 4 alternatives, 3 bits 8 etc. The general rule is simple: every time the number of alternatives is increased by a factor two, one bit of information is added. The limitation of the S.T.M. is defined by the number of items, to be distinguished from the amount of information. To overcome the limited span some important devices can be employed:

- to arrange the problem-solving task in such a way that we make a sequence of several judgments in a row. For instance, generating hypotheses each sequentially explaining a relationship between the collected data;
- to combine informational elements. The combination of bits is called a chunk. A chunk is defined as "any structure that has become familiar from previous repeated exposure and hence is recognizable as a single unit."(23). S.T.M. restriction is directly related to a number of chunks but independent of the number of bits per chunk. By increasing the bits per chunk the amount of messages that the operator can remember increases correspondingly (22). There is a great deal of evidence to suggest that experts in any field think in terms of chunks whilst the novice regards the same information in smaller units. In the language of communication theory, this process is called recoding (see e.g. 24). The input is given in a code that contains many chunks with few bits per chunk. The problem-solver recodes the input into another code that contains fewer chunks with more bits per chunk. There are many ways to do this recoding, but probably the simplest one is to group the input events, apply a new name to the group, and then remember the new name rather than the original input events. Recoding is an extremely powerful weapon for increasing the amount of information we can deal with. In one form or another, we use recoding constantly in our daily behaviour. When we want to remember we rephrase it "in our own words": we make a verbal description of the event and then remember our own verbalization (22).

Miller estimates the S.T.M. capacity for chunks of information at 7 plus or minus two. Elstein (25) found 4 plus or minus 1 hypothesis a time, and Kozielecki (26) 3 - 6 hypotheses. The question is that we do not know what the contents and the validity of a chunk may be. Largely, they have a very personal nature often based on only singular samples which may be falsified by tomorrow's happening (27).

The Validity of Inferences.

People commonly view their memories as exact copies of their original experiences.

During reconstruction, a variety of cognitive, social and motivational factors can introduce error and distortion into the output of the Memory Recall.

If people are unaware of the reconstructive nature of memory and perception, one cannot distinguish between assertion and inferences; they will not critically evaluate their inferred knowledge. In general, any process that changes the contents of memory unbeknownst to people will keep them from asking relevant validity questions and may lead to overconfidence (28). Wason & Johnson-Laird (29) have shown subjects having considerable confidence in their own erroneous reasoning. The inferences had not been arrived at as a series of logical steps but swiftly and almost unconsciously.

The validity of the inferences was usually not inquired into; indeed, the process was usually accompanied by a feeling of certainty of being right (28). The strategies of the inference processes seem to differ from subject to subject. The subject's strategy becomes a crucial intervening variable between the task environment in which he works and the behaviour he produces. Some strategies may be more effective than others, but effective strategies may be difficult to discover, and it must not be assumed, particularly in unfamiliar tasks, that all subjects will come equipped with the best strategies or will be able to find them during the problem-solving setting. Subjects tend to use various kinds of heuristics to restrict their domain. The selectivity of search, not speed, is taken as a key organizing principle (10). There is a strong tendency to confirm the hypothesis directly. The strategy of disconfirming alternatives is less automatic. Subjects largely behave in a deterministic way.

When the rules fail, they tend to assume that there is no rule at all.They use some sort of memorization strategy or they give up and guess (19).

Wason and Johnson-Laird question why subjects fail to follow adequate logical rules. They suggest that subjects do not understand the logical relation of implication ("if","then") but tend to consider this relation equivalent to a double implication "if, and only if". Hypotheses were not properly tested although the optimal rule to help them was given. (29).

It was assumed that subjects will employ deductive inferences with an eye on logic; instead they use various and varying heuristic rules according to time, environment and personality. The reason may be that they lack the necessary basic schemata to help them understand and use the information provided by their experiences. Subjects have the intuitive understanding that non-randomness is information, and information can be exploited to search a problem space in promising directions and to avoid the less promising. A little information goes a long way to keep within bounds the amount of search required, on average, to find the solution (10).

When they encounter or detect order or pattern stimuli, perhaps a memorized correlation of a known chunk, subjects feel at ease. It will direct their pathway through the problem space. This is especially interesting when conclusions or patterns are formulated early in the process. They can be misleading and may direct attention to irrelevant features of the problem, cause the subject to engage in a search for inconsequential cues that would otherwise be ignored, or lead the decision maker to refrain from useful search for cues that would otherwise be collected.

Emotional Influences on Problem Solving Strategies.

"Decision-making is reducing uncertainty in a problem situation" (30) A

harassed decisionmaker confronted with a complicated task suffers a decline in cognitive functioning as a result of the anxiety generated by his awareness of the stressful situation. Considerable stress can be evoked in a decisionmaker merely by his trying to cope with the cognitive limits on his ability to work out a good solution to the problem in hand. Next to cognitive complexity, there are major sources of stress in decision making, including profound threats to the decisionmaker's social status and to his self-esteem that intensify decisional conflict (4). Every problem to be solved has a more or less mental impact on the decisionmaker.

Theoretical reflections on problem solving invariably show us how different our concepts of thinking processes are from what actually seems to be the case when confronted with real troubles. The cognitive functioning is affected by psychological stress. The difference between the well-trained versus the non-trained individual is often striking. Well-trained people are often less subject to intellectual deliberation. Problem-solving skills notably diminish when the individuals are confronted with problems outside their (sub)specialty areas (31).

Stress and uncertainty are accompanying elements of problem-solving and, have been, thusfar, hardly subject to investigation.

According to Einhorn & Hogarth (32), our culture does not encourage explicit representation of uncertainty; it tends to promote confusion between certainty and belief.However, the relation between stress and problem-solving is noteworthy. It involves several features in the problem-solving process as well as the choices and applications of various decision-making strategies.

When there is a high level of uncertainty the amount of information search declines. The problem-solver spends less time in the initial examination of the problem and enters the information processing phase much quicker than in a less stressful situation. One can observe that subjective uncertainty is inversely related to time. The greater their uncertainty the less people perceive the studying of the problem as a fruitful enterprise which can lead to a solution (33).

Each person (subconsciously) knows his level of uncertainty which marks the difference between intelligent decision making and panic, between looking for solutions and complete confusion. He tends to avoid cues that can stimulate anxiety or other painful feelings. It may sometimes help a person to avoid becoming completely demoralized. Janis & Mann (34) recognize 5 patterns in coping with stress and uncertainty in emergency situations :

1.) <u>Vigilance</u> : the individual is capable of effectiveness in an uncertain situation. It generally results in thorough information, search, unbiassed assimilation of new information, and effective planning.

2.) <u>Unconflicted inertia</u> : if the oncoming disaster is of an unfamiliar nature, the person is likely to generate alternatives by searching his memory in an effort to remember similar threats encountered by himself or others in the past. An effective action will be taken if the memorized alternative emerges into consciousness.

3.) <u>Unconflicted change to a new course of action</u> : the person's aroused emotional state leads him to a defective coping behaviour if the danger materializes. His search for an effective action largely depends on this assessment of his own internal resources. If he cannot find an answer, the person will pessimistically give up searching for a better solution, despite being dissatisfied with the options that are open.

4.) <u>Defensive avoidance</u> : the subject will avoid cues that stimulate anxiety, uncertainty, stress or other painful feelings. The person becomes selectively inattentive to threat cues and avoids thinking about oncoming danger. Three forms of defensive avoidance are distinguished

-the evasive form : neglecting and ignoring elementary safety precautions, becoming fatalistically apathetic.

-buck passing: depending upon someone else to make the decision. It may take the form of relying on outside agents of dubious reputation if they promise a less painful solution than the genuine expert who insists that the person himself must take responsibility.

-bolstering: ignoring available information and developing rationalizations which argue against evidence of its potentially unsafe features. Typical examples of bolstering are to be found among certain types of cancer victims. Many of them ask no questions and selectively misperceive what their physicians are saying. They also develop rationalizations to convince themselves that their worries will be over after treatment

5.) Hypervigilance : in its extreme form it is popularly referred to as 'panic'. The victim surmises that time is too short to make a (thorough) search for alternatives. He is overwhelmed by his emotions, unable to look for any effective action.

Apart from the first one the other coping patterns are more or less defective with regard to an optimal decision making strategy. As we shall discuss in Chapter V, decision theory is based on the concept of optimizing objectives. The optimizing strategy's goal is selecting the course of action with the highest pay-off. Such a strategy requires estimation of the comparative value of every viable alternative in terms of expected benefits and cost. But human beings rarely adopt this decision making approach. The determination of all the potentially favourable and unfavourable consequences of all the feasible courses of action would require the decision maker to process so much information that impossible demands would be made on his resources and mental capabilities.

As a result of personal limitations, which we concisely mentioned before, and various external constraints, a decisionmaker who does the best he can to use an optimizing strategy is still prone to use such gross miscalculations that he ends up with an unsatisfactory suboptimizing solution, one that maximizes some of the utilities he expected to gain at the expense of losing other utilities (4).

The optimizing strategy can be regarded as an excellent NORMATIVE model, i.e. a set of standards the decisionmaker should strive for. However, objection can be raised against the assumption that the optimizing strategy provides an accurate DESCRIPTIVE model of how people actually do make decisions or solve problems. With a view to the obvious restrictions that underly the routine problem solving processes, we are inclined to assume different strategies people use in their decision making.

Pitz hypothesized a strategy of "best-guess" heuristic (cited in 28). Kleinmuntz & Kleinmuntz (35) focus on inductive strategies. Instead of assuming the problem-solver has perfect knowledge of the task environment, they assume that the decisionmaker's only knowledge is what can be induced from the observation of individual instances.

In the same line, Simon (36) hypothesizes that the decisionmaker satisfices rather than maximizes. According to Simon, people simply do not have "the wits to maximize". The decisionmaker looks for a course of action that is 'good enough', that satisfies his goals and expectations. Actually, he does not bother to compare it with alternative actions. Simon argues convincingly that the satisficing approach fits the limited information-processing capabilities of human beings. "The world is populated by creatures of "bounded or limited rationality", he says, and these creatures constantly resort to simplifications when dealing with complex decision problems. Man's limited ability to foresee future consequences and to obtain information about the variety of available alternatives makes him settle for a barely "acceptable" course of action. He is

not inclined to collect information about all the complicated factors that might affect the outcome of his choice, to estimate probabilities, or to work out preference orderings for many different alternatives. He is content to rely on a "drastically simplified model of the buzzing, blooming confusion that constitutes the real world" (36, 4).

Janis & Mann (4) point to a number of variations in the satisficing strategy. In quasi-satisficing, a moral precept serves as the sole rule when making a decision to help someone in trouble. This person promptly takes action without deliberating about alternatives. He does not believe his choice to be minimally satisfactory, instead he is convinced that his action is the best, that no other course would be morally justifiable. Tversky (37) described his "elimination-by-aspects" approach as a combination of simple decision rules, which can be applied to select rapidly from a number of salient alternatives one that meets a set of minimal requirements. Lindblom's "art of muddling through"(38) does not specify major goals but tries to find ends that are appropriate to available, or nearly available means without searching for better alternatives. By setting its goals at a low-scale incremental improvement the decisionmakers will easily be satisfied expecting that each small change will lead to an optimum.

According to Janis & Mann (4) at least four different variables are involved in the distinction between optimizing and satisficing strategies:

1.) The number of requirements specifying the testing rule,used to determine a certain course of action, is small in the satisficing strategy.

2.) The number of alternatives is small in the satisficing strategy. The satisficer tests each alternative that comes to his attention; if the first one happens to be minimally satisfactory, he terminates his search.

3.) Ordering and retesting of alternatives. When using a satisficing strategy, the decisionmaker typically tests the alternatives only once and in a haphazard order.

4.) Type of testing model used. When testing to see if an alternative meets a given requirement, the satisficing decisionmaker typically limits his inquiry to seeing whether it falls above or below a minimal cut-off point. If there is more than one requirement, he treats each cut-off point in the same way, as equally important.

Some of these variables can be observed within the context of the medical problem-solving process. They will be discussed in Chapter VIII presenting the results of the investigation.

But before we pass to the medical side of problem-solving, we want to know more of what largely creates the medical problem-solver's task environment: the medical knowledge.

Paragraph 3

MEDICAL KNOWLEDGE

Introduction.

The first question we have to face is whether there is such an identity as 'medical knowledge'. There is a medical knowledge as stored in books, papers, patient records etc. and there is the knowledge as it is applied in daily health care. Are they identical? If we assume the latter knowledge as being the actual medical knowledge, we run into an ambiquity.

On the one hand there is the health care as a social commodity available through the organized channels of society (39). In this sense medicine is a

social institution as well as a technical activity, and is shaped by the economic and sociocultural context in which it is embedded (40). On the other hand there is the medical care for the individual; the patient with his disorders and displeasures of life. But when medicine is dependent on understanding causal relations which occur, both in nature and in persons, we have to admit that these relations can change with changing societies, the rise of social and economic wellfare, and the supply and availability of medical care to populations.

Meehl (41) remarks that the ordinary disorders and displeasures of life become symptoms of disease, latent disease, disease in remission, and even the syndrome of 'non-disease'.

This all leads to what Illich calls the "medicalisation of life" (42). The basic issue, therefore, is whether health and illness are universally definable conditions or relativistic phenomena, identified and labelled according to cultural values and social norms (43). Mechanic (44) tells about a study of skin disorder (dyschromatic spirochetosis) in a South American tribe in which the disease was so prevalent "that Indians who did not have (it) were regarded as abnormal and were even excluded from marriage". In my opinion, Sadegh-Zadeh argues this item very clearly in his editorial "Towards metamedine" (45).

> "We in medicine are always arguing the question whether 'there is' such a thing as disease and what 'the nature of disease' is, while we have no concept of disease which can form the intersubjectively controllable basis for such a debate; chaotic ontologies hamper rational discussion and argumentation because everyone has his own explicit idea of disease; yet we do not hesitate to classify this or that phenomenon as a disease, or this or that person as diseased; the boundaries between medicine and other socio-cultural areas remain quite unclear; hence the blind respect that the layman has for medicine is easily abused to permit the uncritical extension of medical interpretations of events to all aspects of human life. We daily arrive at countless diagnoses in numerous social fields, not only in the hospital or in medical practice, but also at the requests of courts, employers, insurance companies, school authorities etc.; we thus affect the lives of countless people and set far-reaching social processes in motion, and yet we lack a clear concept of diagnosis; nor does there exist any explicitly formulated diagnostic method which can suggest to us how individual physicians reach their supposed diagnoses and what they understand 'diagnoses' to mean. We daily generate countless prognoses and etiologic statements which underlie our therapeutic decisions, and yet we lack an understanding of 'prognosis' and 'etiology'. The same holds for treatment itself because no concept of efficiency for its evaluation is available. In short, medicine is in danger of becoming supervisor of society and the controller of human culture, although its foundations are extremely defective.
>
> As a consequence of these unfortunate deficiencies we have 20-60% or more misdiagnoses (see bibliography : Wagner et al. (46,51), which certainly lead to at least as many mistakes in therapy and to malpractice. The responsibility for these failures is usually ascribed to the practioner. He or she is reproached for not applying 'the latest medical knowledge', and it is said that keeping up to the minute would have helped. This is an all too primitive view for at

least two reasons. First, one cannot reasonably require that specialized, complex knowledge must be applied in order to reach a particular goal if one does not state at the same time exactly what the goal is and how the method of application is supposed to work. Such a demand is based, secondly, on the misconception that clinical practice is nothing other than the application of medical knowledge. Third, who can assure us that the 'latest medical knowledge' is best? For in acquiring the 'latest' knowledge central notions like 'disease', 'diagnosis', 'prognosis', 'etiology', and 'efficiency' must again be relied upon, and these notions, as stated above, have an obscure meaning".

Sadegh-Zadeh expresses his utmost concern about medicine as it stands. It must be an alarming thought that within thousands of years medical knowledge did not reach generally accepted definitions about the most fundamental elements in medicine. At the time of Hippocrates disease was considered to be a purely clinical phenomenon. It was observed directly in the patient at the bedside, and was identified directly, according to the observed clinical evidence, with such names as fever, cyanosis and consumption (47). Causes for such 'diseases' were frequently ascribed to supernatural influences. It is probably mystics that hampered investigation and organisation of medical knowledge in the following centuries.

The persistence of medical schoolmen, who preferred theoretic disputation to experimentation, did have its own influence. But it had no far-reaching consequences. If a patient at the beginning of the thirteenth century developed an acute fever, the physician would prescribe some medicinal herb. This herb was said to possess great healing powers, although unproven. When also a Pater Noster and an Ave Maria were recited three times , the process was further dependent on God, the physicial condition of the patient, and chance (48). Nothing more could be done.

During the 17th century Thomas Sydenham founded the discipline of nosology by insisting that diseases were "specific". Before that time physicians had regarded each sick person as having his own specific ailment. Since then the classification of diseases flourished. The nosologists of the 18th and 19th century selected 'diseases' in a wholly arbitrary manner. Each nosologist created his own collection of 'diseases' based on his own interpretations and beliefs about grouping diverse clinical manifestations together into a clustered entity called 'disease' or, more modestly 'syndrome'; the latter being free of causal explanation. (See Feinstein, 47). The innumerable personal names, attached to as many diseases are still a memory of these ages.

But already in the 19th century some philosophers gave rise to their concern about the uncontrolled growth of this arbitrary classification of diseases. In 1822 Gilbert Blane pointed at the impossibility of logical and successful medical reasoning. Oesterlen formulated in his book 'Medizinische Logik' (1852) a number of recommendations such as : "to show, clearly and impressively, the mode in which we must proceed in our observations, investigations, and conclusions, in order that our theorem and problem may become more clearly intelligible, and that we may arrive at experimental truths and definite laws in our department of science, as well as at scientific principles of practice"; and "thus, when we speak of medical logic, we simply mean the application of the science in question to our own special province. We only seek by its aid to explain and develop more clearly the ways and means, the various processes by which the natural philosopher, the physiologist, and the physician are accustomed to proceed in their investigation of the phenomena, occurence, and influences".

Bieganski (1894) put it straightforward into a triad of rules:

1) a rational reconstruction of the processes of clinical diagnosis and decision making, with a view to understanding the nature of securing the

validity and truth of such endeavours;

2) a psychological account of the processes of clinical judgment with a view to understanding the nature and securing the efficiency of such judgments;

3) an epistemologically or psychologically based revision of actual clinical approaches to medical diagnosis and decision making to the end of greater validity, truth and efficiency in clinical judgment (borrowed from Engelhardt et al. (49)).

The claims and indications of Oesterlen and Bieganski are still valid, and their fulfilment is not yet in sight.

In the 19th century the natural sciences came into the scene. They have become the core of the medical science, and shaped the outlook of medicine of today. But the fundamentals of medicine itself did not change. It is mainly based on 'ideal type' patterns. 'Ideal types' which are approximated, or assumed to be, by the majority of disease entities, ultimately defined by their joint pathology and etiology.

Initially a physician called attention to a syndrome, a cluster of signs and symptoms that he had observed in his practice, or learned as a prototype in medical school. Sometimes even a single patient might be exemplar - including some account of the course and outcome, death or recovery with or without sequelae, and the like. Even in a fairly advanced state of knowledge concerning them, disease entities are at the clinical (symptomes, complaints, course) level that complete statistical predictability from any symptom to any other symptom or set of symptoms is beyond our grasp (41).

Reasoning, which considers most physician-patient encounters as unique, has various drawbacks. The limiting case of uniqueness, e.g., would make diagnosis and indeed all deductive extrapolation impossible. And if medicine is built on 'ideal types' and 'unique cases' what about the objectivity and general applicability of medical knowledge? The obvious question arising at this point is the following: has the medical language which describes diseases and syndromes really a logical structure? We have some doubts about it? (50). Does medical knowledge use one common language all over the world? What do we know about our daily judgments? Usually we take the most striking judgments, data and results for pure gold and construct with these elements our daily medical decisions, scientific results and explanations, the reliability of which we hardly have any knowledge of (51). De Dombal bitterly remarks that "What doctors do is moving from anecdote and aphorism into operational analysis" (52). But most doctors consider their work not as a science but as an art. They mean that it is a fundamentally qualitative and intuitive form of thought which is seriously distorted and impaired by the attempt to use the precise models of the natural sciences.

Most models consider it possible to describe precise quantitative techniques which generate differential diagnoses on the basis of specifications of symptoms and signs. The existing tensions between the two areas in the medical world can give rise to rather dogmatic and inflexible standpoints, confused thoughts, and all too firm conclusions. "Medicine, like theology, cannot tolerate ignorance. If it does not know the answer, it must invent it" (53).

The theory of myocardial infarction e.g. states that this disease is caused by acquired anomalies of the coronary vessels. The moment a case of myocardial infarction is found without this anomaly the theory ought to be rejected. Hundreds of cases of the latter type were found, but the theory still exists. A second theory was added to the first one: a hypothesis about a temporarily narrowing of the vessels, which is unproven (see also 54). Especially in the field of medical knowledge we can quote Popper when he says that 'the old scientific idol of episteme, of absolutely certain, demonstrable knowledge - has proven to be an idol.'

Perhaps it can be hypothesized that medical knowledge consists of the joint collection of the knowledge of every physician in the world. It could explain why so many physicians react so furiously when medical knowledge is criticized. It could explain the manifold conceptions about health and disease.

The Concept of Disease.

Since the human body (unlike social institutions) has changed relatively little over the millennia, the functional norms of somatic medicine are relatively conservative. But since, understandably enough, medicine has expanded its purview to include mental health and mental illness, and since medicine in general must subserve, however conservative, the determinate ideology and ulterior goals of given societies, the actual conception of diseases cannot but reflect the state of technology, the social expectation, the division of labour, and the environmental conditions of those populations (55).

These general statements about the concepts of health and disease are even more complicated by the physician's personal concept of a patient and his disease. This concept is derived from an integration in his own mind of the real patient in front of him and an artificial one, embodied in the collection of medical records from various periods over the years. But this recollection in its turn is distorted by psychological, social and environmental changes in time and person. It is hardly possible to distinguish at least clearly, illness from illness behaviour, complaints of a somatic or psychological origin from complaints of a social origin. This mixture causes us to talk about psychosocial or problem behaviour, but it is the doctor's problem of not recognizing the appropriate parameters for each of the component symptoms and signs (56).

The physiologist Selye suggested that what turns symptoms into medical problems are massive failures in adaptation on the part of individuals to manage the stress that accompanies these symptoms. The mere presence of fixed medical indicators of disorder is not as clinically relevant as the overall coping capacity of people to deal with the situation (57). For the physician confusion arises between deviance and pathology, as there are also ambiguities for disease and health within simultaneously biological and social conditions (58).

This makes it very hard for a physician to cope with the patient's problem. Tremendous differences (see e.g. Friedson, Mechanic, Scheff) exist between medical criteria used by the examining physicians and the relativistic sociocultural standards employed by lay people to identify (the presence of) illness. Patients entering the 'sick role' (Parsons (59)) have their own concepts of disease. Because the 'sick role' provides a person with certain privileges, this concept is not free from personal values. The doctor's role is to define illness, confer the sick status on potential patients, establish priorities, and take the initiative in evaluating health status and controlling health problems (60).

Redlich (61) describes two stages, referred as sick role I and II. The patient in sick role I assumes that

a) he is not responsible for these changes (they are not voluntary and not punishable);

b) he is excused from his ordinary social and occupational obligations; and

c) he is expected to strive to get well as expediently as possible.

Acknowledged as a 'patient' by the physician, he enters sick role II, in which it is assumed that

1) he will drop or modify his own system of thinking; and

2) adopt the physician's system, especially if clear professional information is offered to him.

This not only puts the responsibility in the hands of the physician, but it may also lead clinicians to confuse symptoms with sociocultural patterns of behaviour associated with help-seeking. The distinction frequently summoned between illness and disease is between "what the patient feels when he goes to the doctor and what he feels on the way home from the doctor's office" is often denied (41). The doctor is greatly aware of the fact that when symptoms become illness problems, they are like the tip of an iceberg - a bothersome outcropping of a condition which much of the time lies unobtrusively below the

surface (43). The physician maintains all this medical knowledge largely in terms of abstractions; that is, his concept of a particular disease is really a generalization of his past experience with patients he has read about or saw himself. These generalizations are the representation of more or less loose associations of similar patients about which some statements may be made regarding common mechanisms or causes of illness or regarding therapeutic choice, response, and prognosis (62).

The concept of disease, therefore, has a strict personal character composed of various sources. Therefore, physicians claim that every individual case is unique, that individual values are involved in clinical judgment and hence no general approach can be valid (63). But as Scriven (63) stresses this claim is about on a par with "every athlete is unique and hence their performances cannot be compared". When physicians (and society) cannot agree on concepts of health and disease, how can we know what we are doing, how can we avoid being criticized about the expensive health industry.

Thusfar, medicine has determined it goals and strategies almost exclusively with regard to its own interest and assumed interest of the population. But we are obliged to analyse medicine as-it-stands. Most of those who have attempted such an analysis have avoided consideration of a fundamental logical problem - the definition of the basic concept : "A disease". Diseases are regarded as having some sort of an independent existence, though the sense in which they can be said to exist is left conveniently vague by referring to them in such undefined terms as "events" (64), or "entities" (Baron & Fraser (65)) or by regarding them as attributes of the patients (Ledley & Lusted) (62). Scadding (66) set forth the requisite requirements for a formal definition of disease:

1) avoiding of tautologies;

2) definition of relevant populations;

3) establishment of normal standards;

4) universal acceptance of terminology;

5) confidence in establishing the presence of each disease;

6) flexibility to allow for modifying definition by changes in diseases with time.

That means a conceptual openness; the list of indicators is an open list, in that we do not rely upon only one or a few definitive indicators; that we do not claim that the list is complete, because we expect more advanced knowledge to be discovered in the future.

One of the points of pursueing this kind of questions is to improve clinical practice, and one way in which this might be achieved is through our conception of disease and to learn how it effects our ways to diagnosis.

Medical Diagnosis.

Medical diagnosis is a difficult and complex task largely empirically based and poorly understood as an intellectual task. The gap between the information which is present for diagnosis and the information accessible from memory is difficult to close even for a highly trained general practitioner with substantial daily exposure to many disorders (67).

One way to make diagnoses more accessible to understanding and comparison is to dismantle them of emotional and personal values. But according to Whitbeck (68) this naive view of diagnosis is that it is the scientific determination of the nature of the disease, manifest in a given case of illness, that it is value-free, and that the value issues arise only when one turns to subsidiary

tasks of deciding upon treatment and management. But this Aristotelian view which declares natural objects like diagnosis to be existent prior to and independent of human purposes, separates diagnosis from the rest of the decision-making and makes diagnostic reasoning often less accurate. The present difficulty is that if a patient has a problem that is not in the specialty area of the physician, it will not be recognized, and the diagnosis may be missed entirely.

Sadegh-Zadeh summarizes what the diagnosis is not: it is neither an attribute, an abnormality, a dysfunction, a pathological state or a disease, nor the name of these attributes, abnormalities etc.; the recognition of the patient's disease, the recognition of the 'ultimate cause', or the class to which the patient belongs. Diagnosis is a particular statement about an individual and relative to the patient to whom it applies. This particular statement is relative to the physician, his conception of disease, his medical knowledge, and time (69). But this gives diagnosis the status of a highly personal opinion and, therefore, incomparable to similar cases, and not accessible to evaluation.

Newell and Simon (70) viewed diagnosis as a representation of the problem-solving process wherein the problem-solver (in this case, a physician) is required to examine, evaluate and select information in order to reach a goal (a disease specification). It requires :

1) formal language for dealing with complex problems;
2) a precise formulation of the theory;
3) a direct and unequivocal test of the theory.

As we have illustrated before, in most cases these requirements are not all there.

Attaching a diagnostic label to a patient is equivalent to establishing a similarity of the presented signs and symptoms with prototypical patterns. But in the practice of medicine the patient will not have all the described symptoms and signs of any of the textbook diseases, because they are largely collective descriptions which are personally percepted and described.

The individual presents usually but a tattered fragment of the full-blown composite picture of the disease as described in the textbooks. Experienced physicians know the variability of the presented symptoms and signs. There are very few truly pathognomonic signs either in the sense of an inclusion test (the presence of a sign quasi-proves the presence of the disease) or stronger, two-way pathognomicity, in which the absence of the sign excludes the disease (41). Textbooks, therefore, generally do not concern themselves with providing probabilistic information (52). The symptom-diagnosis relationships are usually described by means of labels like never, almost never, very very seldom, very seldom, more or less seldom, not known, and similar or different fancy epitheta to the positive side (71). In practice the incidence of the various symptoms and diseases is unknown. These figures must be estimated from data bases of past patient records, doctor's memory, and the like. Problems arise because:

1) data collection varies in time, per patient, per physician,
2) the conjunctions of combinations of symptoms or symptom-complexes vary in time, with the kind of patients, per physician. Actually one can say that every doctor creates his own data base which enables him to behave in a heuristic fashion.
3) the accuracy of the diagnoses and their incidence is unknown.

Even if the measurement sectors are complete for all individuals, the problem of reliability of the data remains, since there are still plenty sources of gross errors, from patients as well as from laboratory results (72). Physicians know, indeed expect their patients to be unreliable and unpredictable sources of information. Cascaded inference (73) - stepwise inferring one element from the other - can lead to unreliable inference from the side of the physician. This is because observation (of patients) cannot be value-free, without a prior conceptual configuration in mind. The patient presents his case no more value-free than the doctor observes it. Besides, the fact cannot be

stressed that the same symptoms and signs can be seen in several different diseases as they represent altered pathophysiology or anatomy and do not reflect a specific dynamic of the disease process itself (74). And which element is sufficiently deviant to call "abnormal"? Symptoms and signs are sometimes regarded as the fundamental elements of clinical data, but they are also liable to observer variation, because:

- real 'patients are never identical in different situations (with different clinicians);
- possible differing interpretations upon identical information;
- differing answers are possible by the patient, even upon identical questions.
- there might be a different framing of the questions, as well as
- misleading answers by the patient.
- symptoms and signs might change in the course of the illness (75).

The patient's data and the relation between the variables seem to have a probabilistic rather than a deterministic character. But these probabilities are neither fully investigated nor is probabilistic reasoning taught to medical students. A medical diagnosis can rarely be made with certainty; the end result of the diagnostic process almost always gives a "most likely" diagnosis. But to understand the consequences of this probabilistic character of the diagnostic process it is of crucial importance to define signs and symptoms and diseases as exactly as possible.

Medical knowledge seems to be composed of two parts : one part concerns the general and explicit knowledge as it is laid down in textbooks, journals, reports etc. and is accessible to everyone who is interested in this subject matter. Whether this knowledge is fully reliable and can stand scientific investigation is irrelevant to this description.The other part is the personal component of medical knowledge, which is acquired by experience in routine practice.

It is especially this latter component of the physician's medical knowledge that is greatly emphasized in the medical world. Medical education is partly based on experience; vocational training for family physician is largely based upon experience, physician-teachers are selected by their amount of experience, and patients often make their choice by the years of practice of the physician. It is believed that experience plays a major role in the medical problem solving process. However, this belief has never been tested nor proven. Not only personal knowledge cannot be assessed by objective means, but we also do not know the validity and reliability of this professional and obviously widely used knowledge. We do not know whether physicians do learn from experience. More and more it has been questioned.

Lenoir & Chales (76) state: "if, in the happy cases, experience gives evident knowledge (but personal and difficult to transmit) to the physician, among others the unreliability of the memory, the dangers of too personal retrieval, the major role of the environment and the personal emotions they all together may give him a very inexact representation of the underlying pathology and the symptomatology of the illness". Especially in the studies of Brehmer (19) the effect of experience on reliable knowledge is questioned. Brehmer showed clearly that people do not learn optimal strategies from experience even if they are given massive amounts of practice; that people do not always learn from experience, at least not when the experience consists of a series of cases. And if we do not learn from experience, this is largely because experience gives us very little information to learn from.

People overremember their own past successes and underrate their past failures. Physicians do not realize that in their therapeutic action by choosing one direction they exclude the other, perhaps opposite one, without the possibility to reverse. A person's experience will not necessarily tell the truth. We have come to have what only be called a perverse conception of the nature of experience. We remember what has happened to us in the past, how this thing affected us and what we did about it, but it is filtered by what we want

to observe, want to feel, and wanted to do about it, excluding alternatives.

This can lead to a kind of automatism, a "habit of mind", that makes us believe that we will see even more things of this kind. Thus, when we have seen five white swans, we tend to believe that the next swan will also be white (77).

It is this inductive way of reasoning that gives medical knowledge its own peculiar and personal character. Or, to put it in other words, the extensive store of medical knowledge is limited for every physician by personal commitment.

Paragraph 4

MEDICAL PROBLEM SOLVING

What is a medical problem?

Patients do not present 'problems' to the doctor: they present a number of complaints or signs which they observe as deviant from their normal pattern of physical, psychical and/or social functioning. From these features they present only a small part: the part they think to be relevant to the physician. They present the signs according to their conception of the medical world and medical problem-solving: their notion of 'normality' and 'normal functioning'. But what we do not know is the portion of the total picture of signs and symptoms the patient reveals his physician.

Obviously, the patient presents his unresolved problem to an in his eyes, more competent and professional problem-solver: the physician. It is up to the physician to understand what the patient's problem is. Boshuizen & Claessen (78) raise the interesting question whether the physician solves the patient's problem or is occupied with his own diagnostic task. The answer is that each has his own personal problem.

As we have argued previously, a problem can only be perceived as a problem when it is assimilated into a personal puzzle, which needs a translation into a familiar context in order to catch and solve it.This does not imply that the physician immediately perceives a course of finding a solution, but he recognizes it as belonging to his task environment.

The patient-physician contact can be sketched as follows:

1.) A person experiences some troubles which become a problem by some kind of social process. (e.g. Lay Referral System (79) Sick role (59, 61)

2. Based on the patient's appreciation of his problem and his expectation about the medical competency, he may turn to the medical care institution(s), c.q. the physician.

In our opinion, it is noteworthy that only a quarter of the people who experience some signs of health disturbance visits a medical institution (44)

3. The patient presents (pieces of) his unresolved problem to the physician expecting some advice or prescription for some course of action which enables him to solve his problem. However, this ideally typified procedure hardly ever takes place. According to Redlich (61) the patient is expected to adopt the physician's system, in which case "the doctor solves the patient's problem". The implications of this behaviour may determine - for a large part - the attitudes of the participants towards the processes of illness, treatment and health experience.

4.) From the presented data the physician creates his own problem. This means that every medical problem has its own personal character.

A medical problem, therefore, is the highly individualistic translation (by the physician) of a highly individualistic presentation (by the patient) of a complaint.

Attempts to categorize and classify medical problems seem questionable. Different perceptions of health disturbances, different abilities in coping, different social and professional backgrounds, different education, different social levels (to the physician), different culture and language, and different appreciation of the health care system, are all factors creating thresholds in the appreciation of the patient's troubles. The question of who is healthy and who is not is a matter of great inherent ambiguity that is actively negotiated between people with complaints and the other people ranging from members of the immediate family, friends and neighbours to people presumed to know about health matters because of personal experiences, special training, or connections with the medical care system (80).

Not only the presentation of the illness varies, the illness behaviour influences the physician's identification of the problem also (81, 82, 83) The patient's behaviour can direct the physician's attention to certain features while he neglects other ones. E.g. a physical presentation can distract the physician from psychical or social aspects of the presented illness and vice versa.

Another disturbing factor can be the way in which patients and physicians communicate. "Talking to people is a doctor's game that doctors don't play". This exaggeration from the part of the author illustrates the kind of difficulties which may arise in patient-physician contact. The extensive literature on patient-physician communication examplifies this particular issue.

"Doctors observe little and they observe badly", as Lasegue noted more than a century ago. This rather cynical statement points to the question of the precise and accurate observation, by the physician, of the symptoms and signs in the patient, in order to know what the medical problem is. Some studies (84, 85, 86) about inter-observer (recordings of clinical observations between various physicians) and intra-observer (recording of observations of one physician over periods of time) variabilities underscore the difficulties physicians face observing the patient's symptoms and signs. Wulff (48) expresses the impossibility to determine the precision of the questioning and examining procedure as the patient cannot be questioned several times about the same symptom without one answer affecting the other ones. However, the truth (accuracy) of the patient's answers must usually be taken for granted.

The presented health problems represent abstractions of disease manifestations. They are largely generalisations based on experience, and it is not expected that all (or even most) patients with a given disease will present signs and symptoms which exactly match the predefined pattern.

Nevertheless, the chief complaint, its allied symptoms, related physical signs and laboratory and X-ray information must be grouped into a category, a diagnosis, a name of a disease. Having classified the disease the physicians are able to outline a course of action and treatment.

The Validity of Clinical Judgments.

Hippocrates advises us: "From all the symptoms taken together one should form a judgment". According to the Hippocratic view, clinical judgment is considered to be primarily a cognitive, value-free task, applying basic and clinical knowledge to individual cases, to the point where value payments become a factor for making choice of treatment (68) It has been argued before that this schism between a value-free and a value-laden part is an artificial one. Judgment has to be differentiated from productive thought in that typically nothing is produced. Judgment may, therefore, be identified as the evaluation or categorizing of an object of thought. When we want to try to get some insight in the physician's problem solving, clinical judgment can be rated as the foremost attribute desired in physicians. The task of a physician is to render a judgment, usually a diagnostic classification or therapeutic plan, on the basis of a set of data provided by the patient or the experimenter.

Clinical judgment is an important human cognitive activity, typically carried out by a professional person, with the aim of the prediction of significant outcomes in the life of another individual. When the same type of prediction is made repeatedly by the same judge, using the same type of information as a basis for his judgments, then the process becomes amenable to scientific study. Clinical judgments are a major determinant of the quality in medical practice. They enable us to evaluate the accuracy and the intellectual process of the physician's activities (6).

It is suggested that the clinical judgment of a physician is, at the very least, related to his underlying intellectual ability, to the quality of his medical education and to the depth of his clinical experience (87). It can be viewed in two ways: an objective and a subjective way. In the former one clinical judgment can be understood as occuring within an information-processing

system, which has as its input a specification of observed characteristics of the patient and perhaps some laboratory data and as output a differential diagnosis (88). On the other hand clinical judgment is often viewed as being reached by intuition.

Physicians often assert that they can only make adequate judgments on the basis of actual experience and a "clinical view".

The claim is that knowledge gained on the basis of years of clinical experience is not reducible to explicit rules, recipes, or basic principles (49). However, these judgments are significantly influenced by the knowledge of the outcomes. Even some data and the notion of frequency evokes the physician's judgment upon only a few alternative outcomes. Scriven (63) cynically remarks that most judgments are like astrological advice, they are perfectly reconcilable with any outcome. This simple fact explains the vast power of fashion in even the field of the family physician. Goldberg (6) comes to the conclusion that clinical judgments are:

a.) rather unreliable;
b.) minimally related to the confidence and the amount of experience;
c.) relatively unaffected by the amount of information available;
d.) rather low in validity.

While the relatively few investigations of judgmental stability have concluded that judges show substantial inconsistency in their judgments over time, the vast majority of reliability studies have focussed upon judgmental consensus and have come to widely disparate conclusions (89).

Oskamp (89) found that the judges' certainty about their decisions is entirely disproportionate to the actual correctness of those decisions. The increasing feelings of confidence as the clinician works through a case are not a sure sign of increasing accuracy for his conclusions.

So called clinical validation, based on the personal feelings of confidence of the clinician, is not adequate evidence for the validity of clinical judgment in diagnosing or predicting human behaviour (89).

According to Komaroff (54), the rather low validity of clinical judgments may be attributed to the disturbingly "soft" medical data. These data are defined, collected, and interpreted with a degree of variability and inaccuracy which falls short of the standard which (knowledge) engineers expect from most data. Variability in the medical history may stem from the patient's description of his illness, the physician's conduct of the interview, of the dynamic interaction between the patient and the physician during the course of the interview. The physician's conduct may be reflected in his selection of questions to ask. Several 'filters' will affect the data collection and interpretation. Komaroff (54) distinguishes a number of 'filters' like:

- <u>motivational biases</u> : physicians have a tendency to recognize evidence which supports a position already taken, and to disregard contrary evidences;

- <u>cognitive limitations</u> : resulting in bias in data interpretation and decisionmaking. Tversky and Kahnemann (90) listed a number of such properties like :

a.) representativeness : the mind fails to make proper quantitative assessment of immediate experience,
b.) availability : the mind fails to recall pertinent past experience.
c.) anchoring and adjustment: the mind places undue emphasis on its initial estimate.

- <u>Situational factors</u> like pressure of time, exhaustion and other psychological or organizational distractions.

It is because of the rather pessimistic view on the low validity of clinical judgment, and the relation to the optimization of medical care, that some people turned to a search for a more objective way of investigation and assessing clinical judgment. In their view (based on Decision Theory) cues,

preferably independent from each other, can sequentially be processed by a number of statistical procedures. These models are formally analogous to decision models and can be considered as possible process models for clinical decision making (7). These models may serve as examples or standards for physician's problem-solving, because several models do a remarkably good job of predicting human judgments. These researchers, however, forget that although physician's procedures perhaps sometimes fall short because of a number of biases, the medical data to be processed are beyond desirable accuracy also. Here the two concepts split. The clinical decision-makers believe that improving and optimizing the data-processing strategy eventually will lead to an improved clinical judgment. Physicians, however, believe that they use the more or less accurate cues in a variety of - most nonlinear - ways. They suggest that they are accustomed to these "soft" data and are able to ply their strategy to the case at hand.

Researchers on medical decision-making have attempted to capture these ways with more and more complex equations (91). These researchers expected:

1.) judgments to be sensitive to configural information, and

2.) judgment to be based on a weighed additive summation of information (7).

Physicians, however, mould their configuration, of symptoms and diseases, to their own knowledge and opinions. Not only do they translate their knowledge into personal degrees of association between symptoms and diagnoses (71) but they also consistently indicate that in patient descriptions in which there existed a correlation between symptoms it was more likely that disease occurred than in patient descriptions that did not preserve the correlation (7). This latter finding contradicts an independent cue model.

These notions are reflected in clinical practice, where different (groups of) doctors will attach different diagnostic labels to the same patient.

One explanation is to assume differential knowledge between various groups. Another explanation is that different groups of physicians will give differing definitions of the same situation and that this would be the explanation of multiple diagnostic labels (23) The former explanation will find some support by the study of Nobrega et al (92) who investigated the records of 138 patients. They found that:

49% of the items were not listed in the medical history;
43% of the items were not listed in physical examination;
25% of the suggested lab tests were not requested;
30% of the special diagnostic procedures were not done;
21% of the treatments were not prescribed;
86% of the follow-up items were not recorded.

Especially, while physicians were educated to try to explain all clues by one diagnosis, it certainly can be a hard job to reconstruct the physician's judgments. When we assume that physicians reason by analogy of past experiences differential knowledge may be the explanatory result.

The high incidence of social and behavioural problems, especially in primary health care, can lead to confusion in physicians. The lack of a structure for these types of data, forces the physician to choose between data. Thereby it is important to know which he will leave out and which he will include. Frequently this is an intuitive process accompanied by a state of uncertainty.

Why is there so much case-to-case variation even among the more experienced physicians? (93). Why are judgmental accuracy and confidence not related to one another? (6, 89). Why is the amount of available information and diagnostic accuracy not related? (89). Why are experience of the physician and judgmental accuracy not related? (6, 94, 95). Why do physicians not rely on prior probabilities (prevalence or incidence of symptoms and diseases) but on attribute similarity and are judging by representativeness? (93)

Several explanations are given in literature. Not scientific explanations,

but rather assumptions, hunches. Medicine is a good example of ill-structured problems. Ill-structured problems are the ones in which initial and final system states are often not known. Hereby the space of possible solutions is large so that general methods of searching this space are likely to be unproductive. Success in solving ill-structured problems depends on having good problem-solving heuristics and large amounts of well organized, domain specific knowledge (11).

This latter condition contrasts with the domain of more general physicians like family physicians and general internists. They must struggle with many unresolved and difficult clinical and non-clinical issues. They are subject to such things as observer variability, dehumanized data, absence of designation citeria and deficiencies of medical information dealing with patient care (96). They often rely on deficient prototypes, some assumed correlations, or remember relationships because of some common denominator which may be rather obscure on superficial examination of the actual logic used at the bedside.

This is, according to Mason & Bulgren (cited in 97), the experience of which physicians are proud of. This is what appears to be the artistic ability in the practice of medicine and why it may seem to many physicians that they are gifted in making correct decisions without definable data and why they profess that rare diagnoses can be correctly made when no statistician would dare stick out his neck on the basis of the data (97). They overestimate the importance they place on minor cues and they underestimate their reliance on a few major variables.

Barrows and Bennet (74) were amazed that by almost all physicians the judgment was reached before all the available data from the patient had been obtained. But how can a physician in a real life situation know that he has obtained all data? How can a physician distinguish relevant from irrelevant data? How can he memorize these data? Kleinmuntz (94) demonstrated that data not related to the physician's mental hypothesis or diagnosis are totally forgotten by the physician. During physician-patient interviews doctors often ask the same questions twice or thrice, and cease to inquire further when a stereotype pattern had come to his mind.

It is this type of deterministic reasoning that has intrigued many investigators. They argued that owing to the inherent uncertainty of a diagnostic task, the problem solving task should be performed in a probabilistic way. Since the late fifties a lot of effort has been invested in constructing probabilistic diagnosis rules. Regrettably, much less has been done to revise rational tools for evaluating them (98). These efforts could only be of normative origin, because nobody taught students and physicians really "how we do what we do". There are rules of thumb, there are options and there is "experience", and "intuition"; but how all this works in practice is rather uncertain (99).

If probabilism should play a part, it had to be invented before it can be detected. Besides, the advantage of a deterministic way is that it can restrict the problem space quicker than the probabilistic way. A positive answer can be most rewarding. In 27 different diseases the symptom nosebleeding is mentioned, not in another 3225 ones. In positive finding, assuming that it is pathognomonic, you select 27 diseases, in negative, 3225 (15). Moreover, man is very bad in making probabilistic estimates (100).

People largely rely on estimates from inferred frequency of occurrences or assumed resemblances in their memory of "experience", usually based on striking successes of gross failures. Every physician has anecdotes about brilliant examples of clinical insight, but which of us has a logbook with validated counts on our own or anyone else's long-term track records? (63). When people have to learn about the validity of their judgment, they will not have acccess to all four outcomes (true positive, true negative, false positive, false negative (See Chapter V). Then, it may not be surprising that people have not learned the optimal way of coping with tasks of this sort.

It illuminates an important difference between the expert in practice and the expert as often pictured in literature or folklore. The epitome of the

expert in fiction is the detective who, through superior deductive powers and by sheer force of logic, organizes the facts at hand in a way that lead to a single, inevitable conclusion. By contrast, the realworld physician seems to rely much more heavily upon "guessing", his initial hypothesis being typically based on precious little data. These "guesses" are apparently prompted by patterns of clinical findings. The physician then tries to demonstrate the correctness of his "guesses", moving to new hypotheses only if his initial impressions prove untenable (101).

These two conceptions of reasoning processes in medicine raise the question of strategies, methods or procedures in medical problem-solving.

Strategies in Medical Problem Solving

Implicit in the definition of the diagnostic process is the assumption that there is an underlying structure, preferably a logical one. As sketched before, however, this structure is hardly known, and, to say the least, uncertain. The way in which doctors think is thus a matter of some speculation. Perhaps physicians behave "heuristically" (102) or in an "inductive algorithmic fashion" (103); perhaps by the pay-off's of the presumed outcomes (102), perhaps they think in mathemathical sets and symbolic logic (104). Several speculations about physicians' thought processes have been made. They almost all have one thing in common: the conviction that the clinician's description of what it is like for him to make clinical diagnoses is a rationalization (88).

Generally, we have no access to higher order reasoning processes. Retrospective descriptions are largely flawed by hindsight verification. In search of a theoretical structure of clinical medicine, and of diagnosis in particular, most models try to imitate a process the nature of which we are largely ignorant (105). One can roughly distinguish two types of investigation:

a.) an information processing approach: formalized modelling of rational and logical medical processes;

b.) the "real-life" process approach: observation of patient-physician encounter and/or physician's thought processes in a real or simulated situation.

The former approach encompasses studies in the field of statistical and artificial intelligence ways of decision-making. This matter will be discussed in chapter V.

The real-life studies encircle a variety of studies ranging from interviewing physicians to descriptions of pathophysiological processes. Especially these studies are confronted with the question: 'what is medical problem solving?' Two definitions were found in literature, both by Sadegh-Zadeh (69):

Medical problem- solving is,

1.) "The sum total of the mental and non-mental activities undertaken by the physician in helping a patient who consults him; and

2.) A reconstructive and constructive inquiry into the foundations, structure and processes of clinical judgment".

The latter definition forces us to a more precise and detailed circumscription of the features of the medical process. Some elements of the foundations and structure of clinical judgment are described concisely. We are now interested in its processes. The number of investigations into the processes is rather small. "Most aids to decision-making and human thinking are devised without understanding the process they aided - the thought process itself. The prospect before us now is that we shall understand that process." (106). "We need to first search for and then analyze the individual physician, preferably the physician that can outperform normative formulae (63). The problem to clinical diagnosis resides in the unclarity of what should in fact be diagnosed.

If one really thinks diseases differentiable by means of individual symptoms or groups of symptoms or even contrasting then one must attempt to say by which criteria one can recognize a particular disease (Proppe,cited in 49), because:

1.) only identical criteria allow a comparison both with the knowledge stored in one's own experience and the one in literature.

2.) only identical criteria allow an increasingly important communication with colleagues, especially about the inference for prognosis and therapeutical management.

3.) only identical criteria, acknowledged by all persons concerned, about the concept(s) of disease allow judgments in the medical and social world (107).

Gross (107), in fact, asks for the recognition and description of those elements in the diagnostic proces upon which unequivocal consensus is reached. He asks for the "atoms" of the clinical process.The question is if we can find and recognize these atoms. Newell & Simon's theory (70) emphasizes that rational

problem solving is characterized by the adaption to the problem to be solved. Since it is adaptive, the behaviour of a person solving a problem tells us more about the structure of the task than about the process or the personality dynamics of the problem solver.

For real-life studies, therefore, there are at least two conditions:

a.) a clearcut circumscription of the task defining its 'atoms'; and

b.) a description of the characteristics of the medical process in order to recognize these features in content analysis.

The most important and ambitious 'real-life' study on medical problem-solving is, unquestionably the one by Elstein, Shulman & Sprafka (108). In a 'high fidelity' simulation 24 physicians interrogated three "patients", actors, each of them playing a different role. The physicians were allowed to collect data in any sequence they wished. The actor had no fixed script beyond the opening statement of the chief complaint. Data beyond the verbal mode, e.g. physical examination, laboratory results etc. were provided by an assistant. Between medical history and the provision of the other data a "natural break" was provided. The scene was videotaped for reasons of "stimulated recall", a method in which the candidate is confronted with a playback of his procedure(s) and is stimulated (by an interviewer) to think aloud about each step of the replayed videotape.

The structure of the task, the elements and characteristics of the three diseases were only briefly communicated in the book. The determination about the relevancy of a certain symptom to a disease or a hypothesis (so-called 'cue') was delegated to a judging-committee of experts. Neither explicit criteria for the relevancy (weights of a cue) of the symptoms nor the number of items in the task structure were described.

Elstein et al. used three units of analysis of the process: information search units, cues and hypotheses. Information search units were all elements that seek information or instruct patients regarding their present of previous illness(es) of psychical or social background. Symptoms asked by a physician which were - apparently - not related to any diagnosis or hypothesis in the scope of the investigators were listed in a different manner. Hypotheses, although not defined, served as stepping stones which could be manipulated by specific questions towards a satisfactory diagnosis. The process was supposed to proceed along a logical, sequential, deductive line. This hypothetico-deductive method is "a nearly universal characteristic of human thinking in complex, poorly defined environments" (108).

In our opinion the paucity of results partly reflects the limitations within the research model. Nevertheless, one of the major achievements of the Elstein c.s. study is the approach, the development and the methodology of a simulation study of medical problem solving. Their 'high-fidelity' simulation model proved to be a sound basis for this type of investigation.

Another major achievement is the insight in the course of hypotheses during the work-up. First of all there is the hypothesis generation early in the work-up of the case. These hypotheses have, as was expected, not a broad and vague character but are closely circumscribed, which must be attributed to an associative process with materials of past knowledge stored in memory.

The other hypotheses generated during the work-up of the case did not always follow the logical, sequential system of the first one. They give rather the impression of hunches induced by some data. Although this observation does not substantiate the foundations of deductive reasoning, Elstein et al., nevertheless, speak of a hypothetico-deductive method in a 'hypothesis-driven' process. The investigators found the performances of the physicians to be case specific, which is a substantiation of the Newell & Simon theory about the adaption of the individual to the problem.

The propositions of Newell & Simon were, however, explicitly made for rational human problem-solving, in the frame-work of 'objective' or 'accessible' knowledge as contrasted to 'personal' or 'tacit' knowledge. In the former type of reasoning one can expect a deductive outcome of the inference process, in the

latter type an inductive one. The personal and subjective character of an 'inductive outcome' poses a dilemma. On the one hand the problem orientation of the solving process, on the other the personal task concepts and subjective appreciation of the problems.

Feinstein (47) notices that different problems have different structures and demands. People can learn to solve them all, but a theory of problem-solving must firstly be a description of how different kinds of problems are solved.

Descriptive studies of physician's thought processes are small in number. The first attempts to elicit diagnostic decision-rules were based on personal interviews. Kleinmuntz (94) investigated inference-rules for patient's profiles by using the Minnesota Multiphasic Personality Inventory (MMPI). Wortman (14) interrogated a neurologist for his thought processes on solving patient's problems. His description of hierarchical structures in memory formed the basis for various decision models. Goldberg (109) studied psychologists on MMPI profiles, and constructed a model of linear inferences.

However, the biases inherent to verbal reports on mental processes have been vividly reported by Nisbett & Wilson (100). To our knowledge Barrows & Bennet, (74) were the first to use simulated patient scenarios in studying the problem solving processes of physicians. Smith & McWhinney (110) studied family physicians and general internists for some striking elements in their problem-solving behaviour as confronted with simulated "patients". Gale & Marsden (111) confronted 22 final year medical students and 44 physicians with a real patient. The interviews were videotaped and reviewed for content analysis to reveal the specific thinking processes involved.

Gerritsma & Smal (112) studied 12 family physicians and 12 general internists by means of a written simulation on a limited number of features in problem-solving behaviour.

Along a different line some investigators tried to relate physicians' opinions to pathophysiologic disease-processes. These models make use of knowledge of the patho-physiological inter-relationships in the disease or disease-class under study and the physicians' opinions with regard to these relationships. They try to link symptoms to the disease in a logical system. Examples of these studies are:

Edwards (103) studying a logical system for dysphagia, de Vries Robbe (113) nephrological diseases and Van der Werf (114) about physician's opinions on gastric symptom-disease relationships during the last century.

In our opinion, both lines have to be amalgamated in order to come to reliable observation of physicians' thought processes. To observe these processes a model is needed. A model which enables the investigator to perceive features and landmarks in the problem-solving process thusfar unknown. A model can also prevent wishful thinking of the experimentalist.

To our knowledge, only a very few models have been developed to map the physician's problem solving. Apart from the models of formal decision making, there are some models like the ones from Hull (115) and McWhinney (116) in family medicine and more general ones from e.g. Kleinmuntz & Kleinmuntz (35) and Reichertz (117).

The existing models are largely based on the traditional view of clinical medicine, or upon decision theoretical concepts. The developments in human problem-solving theories are scarcely met in models on medical problem-solving. A similar lack can be observed with regard to the foundations of scientific reasoning.

From general sources it has been indicated that most physicians reason in an analogical way. Several analogies springing from the physician's memory as a result of past experiences are related to the target problem. The incomplete mapping between the aspects of the memorized case and the present one may harmfully influence the inference process. However, according to Gick & Holyoak (118), subjects often fail to notice the relevance of a strong analogy to a target problem when a hint to use the story is not provided. Do people just

'stumble' upon an analogy, or do they use a specific search program? How complete or defective are case characteristics stored in memory, and how complete or defective are the mapping processes? In analogical reasoning processes we may distinguish three interrelated activities:

1.) comparing one instance to another;
2.) deriving a schema for a class of instances;
3.) comparing an instance to a general schemata.

We discussed that schemata, and especially general schema, are hardly provided in medicine. It may be concluded that the analogical reasoning processes must have a highly personal character. Not a logical or algorithmic way of reasoning is present, but an inductive and heuristic way. Kleinmuntz & Kleinmuntz (35) described their viewpoint into a model of two types of strategies:

a.) Heuristic Decision Strategy: it uses heuristics to arrive at a diagnosis and then picks an acceptable treatment. This strategy considers people as satisfiers, who indulge in limited amounts of search, until a satisfactory rather than an optimal solution is reached (see Simon's Satisficing theory). The general procedure is to develop a hypothesis about the patient's disease and to try to confirm the hypothesis. It assumes that only positive evidence is sought;

b.) Generate & Test Strategy: it is a concise itinary of the former strategy. The problem-solver does not even bother about symptoms and signs but chooses a treatment that happens to come by.

We do really hope that this latter strategy does not stand for the ordinary physician. But we can surely fall behind Lusted (119) who wished: "One could hope that a small group of problem solving strategies would underly all medical diagnosis and treatment. However, if such strategies exist they are not yet understood."

REFERENCES CHAPTER II

1. Polanyi M.,: Personal knowledge. Towards a post-critical philosophy. London, Routledge & Kegan Paul, 1958
2. Mettes C.T.C.W., Pilot A.,: Over het leren oplossen van natuurwetenschappelijke problemen; een methode voor ontwikkeling en evaluatie van onderwijs. T.H. Twente, Proefschrift, 1980.
3. Scheerer M. Problem-solving. Scientific American, 1963, 208, 118-128.
4. Janis I.L., Mann L.: Decision making, New York, Free Press, 1977.
5. Chapman L.J., Chapman J.P.,: Genesis of popular but erroneous psychodiagnostic observations. J. Abnorm. Social Psychology 1967, 72, 192 -204.
6. Goldberg L.R.,. Simple Models or simple processes? American Psychologist, 1968, 23, 483 - 496.
7. Medin D.L., Alton M.W., Edelson S.M., Freko D.,: Correlated symptoms and simulated medical classification. J. Exp. Psych., 1982, 8, 37 - 50.
8. Johnson-Laird P.N.,: An introduction to the scientific study of thinking, in: Thinking, readings in cognitive science. Cambridge University Press, 1977.
9. Newell A., Simon H.A.,: Human problem solving. Englewood Cliffs, New Yersey, Prentice Hall, 1972.
10. Simon H.A., Newell A., : Human problem solving. The state of the theory in 1970. Am. Psych., 1971, 26, 145 - 159.
11. Whitbeck C., Brooks R., : Criteria for evaluation a computer aid to clinical reasoning. J. Med. Phil., 1983, 8, 51 - 65.
12. Piaget J.: Intellectual evolution from adolescence to adulthood, in: Thinking, readings in cognitive science. Cambridge University Press, 1977.
13. Smedslund J., Note on learning, contingency, and clinical experience. Scandinavian J of Psychology, 1966, 7, 265 - 266.
14. Wortman P.M., Medical diagnosis: an information processing approach. Comp. Biomed. Research, 1972, 5, 315 - 328.
15. Blois M.S.: Conceptional issues in computer aided diagnosis and the hierarchical nature of medical knowledge. J. Med. Phil., 1983, 8, 29 - 50.
16. Wright P.: The harassed decision maker: time pressures, distractions, and the use of evidence. J. Appl. Psych., 1974, 59, 555 - 561.
17. Rogers W., Ryack B., Moeller G.,: Computer aided medical diagnosis: literature review. Int. J. Biomed. Comp., 1978, 10, 267 - 289.
18. Balla J.I.: Logical thinking and the diagnostic process. Meth. Inf. Med., 1980, 19, 88 - 92.
19. Brehmer B.: In one word: not from experience. Acta Psychologica, 1980, 45, 223 - 241.
20. Sanchez, E. : Inverse of fuzzy relations. Applications to possibility distributions and medical diagnosis. Fuzzy sets and systems, 1979, 2, 75 - 86.
21. Dombal de F.T., Horrocks J.C., Staniland J.R., Guillou P.J.,: Pattern recognition: a comparison of the performance of clinicians and non-clinicians - with a note on the performance of a computer-based system. Meth. Inf. Med. 1972, 11, 32 - 37.
22. Miller G.A..: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psych. Review, 1956, 63, 81 - 97.
23. Dowling A.F., Medically oriented computer based information systems. Med. Care, 1982, 20, 253 - 254.
24. Wortman P.M., Greenberg L.D.,: Coding, recoding and decoding of hierarchical information in long-term memory. J. Verbal Learning and Verbal Behaviour, 1971, 10, 234 - 243.
25. Elstein A.S., Kagan N., Shulman L.S., Jason H., Loupe M.J.: Methods and theory in the study of medical inquiry. J. Med. Educ. 1972, 47, 84 - 92.
26. Kozielecki J.: A model for diagnostic problem solving. Acta Psych., 1972, 370 - 380.
27. Weltman R.: Clinical chemist takes hard look at medicine's information explosion. JAMA, 1977, 238, 2465 - 2469.

28. Fischhoff B., Slovic P., Lichtenstein S., : Knowing with certainty: the appropriateness of extreme confidence. J. Exp. Psych., Hum. Perc. Perform. 1977, 3, 552 - 564.
29. Wason P.C., Johnson-Laird P.N.: Psychology of reasoning. Structure and contents.
30. Slovic P., Fischhof B., Lichtenstein S.,: Behavioral decision theory. Ann. Rev. Psych., 1977, 28, 1 - 29.
31. Palva J.P.: Measuring clinical problem solving. Br. J. Med. Educ. 1974, 8, 52 - 56.
32. Einhorn H.J., Hogarth R.M.: Confidence in Judgment: Persistence of the illusion of validity. Psychological Review, 1978, 85, 395 - 416.
33. Driscoll J.M., Lanzetta J.T.,: Effects of two scources of uncertainty in decision making. Psych. reports, 1965, 17, 635 - 648.
34. Janis J.L., Mann L.: Emergency decision making: a theoretical analysis of response to desaster warnings. J. Human Stress, 1977, 35 - 48.
35. Kleinmuntz D.N., Kleinmuntz B.,: System simulation. Decision strategies in simulated environments. Beh. Science, 1981, 26, 294 - 305.
36. Simon H.A.,: Administrative Behaviour. A study of decision making processes in administrative organization. New York, Free Press, 1976.
37. Tversky A.: Elimination by aspects: A theory of choice. Psychological Review, 1972, 185, 1124 - 1130.
38. Lindblom C.E.: The science of "muddling through". Public Administr. Review, 1959, 19, 155 - 169.
39. Eisenberg J.M.: Sociologic influences on decision making by clinicians. Ann. Int. Med. 1979, 90, 957 - 964.
40. White, K. L.,: International comparisons of medical care. Scientific American, 1975, 233, 17 - 25.
41. Meehl P. E.: Specific etiology and other forms of strong influence: some quantitative meanings. J. Med. Philos., 1977, 2, 33 - 53.
42. Illich I.,: Medical Nemesis: the expropriation of health. London, Marion Boyars, 1976.
43. Kellert, S.R.: A sociocultural concept of Health and Illness. J. Med. Philos., 1976, 1, 222 - 228.
44. Mechanic D.,: Medical Sociology. New York, Free Press, 1978.
45. Sadegh-Zadeh K.,: Toward Metamedicine (Editorial). Metamedicine, 1980, 1, 3 - 10.
46. Wagner G., Tautu P., Wolber U., : Problems of medical diagnosis. A. Bibliography. Meth. Inf. Med. 1978, 17, 55 - 74.
47. Feinstein A.R.: An analysis of diagnostic reasoning. II. The strategy of intermediate decisions. Yale J. Bio. Med. 1973, 46, 264 - 283.
48. Wulff H.R.,: Rational Diagnosis and Treatment. Blackwell Scientific Publ., Oxford, London, Edinburgh, Melbourne, 1976.
49. Engelhardt H.T., Spicker S.F., Towers B.: Clinical judgment, a critical appraisal. D. Reidel Publ. Co., Dordrecht, 1979.
50. Tautu P., Wagner G. : The process of medical diagnosis: routes of mathemetical investigation. Meth. Inf. Med. 1978, 17, 1 - 9.
51. Wagner G.: Fehlerforschung als Aufgabe der Medizinischen Dokumentation. Meth. Inf. Med. 1964, 3, 93 - 94.
52. Dombal de F.T.: Medical diagnosis from a clinician's point of view. Meth. Inf. Med. 1978, 17, 28 - 35.
53. Pickering G.: Therapeutics. Art of Science? JAMA, 1979, 242, 649 - 653.
54. Komaroff A.L.: The Variability and inaccuracy of Medical Data. Proceedings I.E.E.E., 1979, 67, 1196 - 1207.
55. Margolis J.,: The concept of disease. J. Med. Philosophy, 1976, 1, 238 - 255.
56. Medawar P.,: Scientific method in science and medicine. Persp. Biol. Med., 1975, 345 - 352.
57. Selye H.: The Stress of Life. New York, McGraw-Hill, 1956.
58. Gutman Rosenkranz B.: Causal thinking in erewhon and elsewhere. J. Med. Phil., 1976, 1, 372 - 384.

59. Parsons T.,: The Social System. New York, Free Press, 1951.
60. Brody D.S.: The patient's role in clinical decision making. Ann. Int. Med. 1980, 93, 718 - 722.
61. Redlich F.C.,: Editorial reflections on the concepts of health & disease. J. Med. Philos., 1976, 1, 269 - 280.
62. Ledley,R.S., Lusted,L.B.: Reasoning foundations of medical diagnosis. Science,1979,130,9 - 21.
63. Scriven, M.: Clinical judgment. in: Engelhardt,H.T., Spicker,S.F., Towers,B.: Clinical Judgment, a critical appraisal. D. Reidel Publ. Co. Dordrecht, 1979.
64. Warner H.R., Toronto A.F., Veassy L.G., Stephenson R.A.: Mathemetical approach to medical diagnosis. JAMA, 1961, 177, 177 - 183.
65. Baron D.N., Fraser P.M.,: The digital computer in the classification and diagnosis of diseases. Lancet, 1965, ii 1066 - 1069.
66. Scadding J.G.: Viewpoint. The semantics of medical diagnosis. Biomed. Comp., 1972, 3, 83 - 90.
67. Eddy,D.M.,Clanton,C.H.: The art of diagnosis:solving the clinico-pathological exercise. N.Engl.J.Med.,1982,306,1263-1268.
68. Whitbeck C.,: What is diagnosis? Some critical reflections. Metamedicine, 1981, 2, 319 - 330.
69. Sadegh-Zadeh K.,: Foundations of clinical praxiology. Part I: The relativity of medical diagnosis. Metamedicine, 1981, 2, 183-196.
70. Newell A., Simon H.A.,: Human problem solving. Englewood Cliffs, New Yersey, Prentice Hall, 1972.
71. Adlassnig, K.P..: A fuzzy logical model of computer assisted medical diagnosis. Meth. Inf. Med., 1980, 19, 141 - 148.
72. Zentgraf R., Victor N.: Some problems arising in the statistical treatment of diagnosis. Meth. Inf. Med. 1978, 17, 10 - 14.
73. Schum D.A., Du Charme W.M.: Comments on the relationschip between the import and the reliability of evidence. Org. Beh. Human Perform., 1971, 6, 111 - 131.
74. Barrows H.S., Bennett, K.,: The diagnostic skill of the neurologist. Arch. Neurol., 1972, 26, 273 - 277.
75. Gill P.W., Leaper D.J., Guillou P.J., Staniland J.R., Horrocks J.C., Dombal de F.T.,: Observer variation in clinical diagnosis. A computer aided assessment of its magnitude and importance in 552 patients with abdominal pain. Meth. Inf. Med., 1973, 12, 108 - 113.
76. Lenoir P., Chales G.: Pourqoui et comment aider les medecins a porter des diagnostics. (Helping physicians arrive at medical diagnosis: how and why). Med. Inf., 1980, 5, 281 - 289.
77. Popper K.R. : The logic of scientific discovery. London. Hutchison, 1959.
78. Boshuizen H.P.A., Claessen H.F.A.,: Problems of Research into medical problem solving: some remarks on theory and method. Med. Education, 1982, 16, 81 - 87.
79. Erickson K.T.: Patient Role and Social uncertainty: A Dilemma of the mentally ill. In: Advantages in Psychiatry, M.B. Cohen ed. New York, Norton, 1959.
80. Twaddle A.C.,: Sickness and the sickness career : some implications. In: Eisenberg L., Kleinman A., : The Relevance of Social Science for Medicine, Dordrecht, D. Reidel Pub. Cy. 1982.
81. Korsch B.M., Negrete V.F.: Doctor - Patient Communication. Scientific American. 1974, 7, 66 - 74.
82. Mechanic D. : Social psychologic factors affecting presentation of bodily complaints. New Engl. J. Med., 1972, 286, 1132 - 39.
83. Zola I.K.,: Problems of communication, diagnosis, and patient care: the interplay of patient, physician and clinic organization. J. Med. Educ., 1963, 38, 829 - 38.
84. Cochrane A.L., Chapman P.J., Oldham P.D.,: Observer's errors in taking medical histories. Lancet, 1951, 1007 - 1009.
85. Koran L.M.,: The reliability of clinical methods, data and judgments, N.

Eng. J. Med., 1975, 293, 642 - 646.
86. Koran L.M. : The reliability of clinical methods, data and judgments (second of two parts). N. Eng. J. Med., 1975, 695 - 701.
87. Bear,R.,Schneidermann,J.: Decision Analysis in Clinical Medicine. Can.Med.Ass.J.,1976,115,833-834.
88. Sober E.,. The art and science of clinical judgment : an informational approach, in: Engelhardt H.T.. Spicker S.F., Towers B.,: Clinical judgment, a critical appraisal. D. Reidel Publ. Co. Dordr., 1976.
89. Oskamp S., : Over-confidence in case-study judgments. J. Consult. Psychol., 1965, 29, 261 - 265.
90. Tversky A., Kahnemann D.: Judgment under uncertainty: Heuristics and biases. Biases in judgment reveal some heuristics of thinking under uncertainty Science, 1974, 185, 1124 - 1131.
91. Slovic P , Lichtenstein S.,: Comparison of Bayesian and Regression approach to the study of information processing in judgment. Org. Beh. Human Perf. 1971, 6, 649 - 744.
92. Nobrega F.T., Morrow G.W., Smoldt R.K., Offords K.P.: Quality assessment in the hypertension: analysis of process and outcome methods. New England. J Med., 1977, 296, 145 - 148.
93. Balla J.I., Elstein A., Gates P., : Effects of prevalence and test diagnosticity upon clinical judgment of probability. Meth. Inf. Med. 1983, 22, 25 - 28.
94. Kleinmuntz B.: The processing of clinical information by man and machine. in: Formal representation of human judgment : pp 149 - 186. Ned York, John Wiley 6 Sons, 1968.
95. Szucko J J., Kleinmuntz B.: Statistical versus clinical lie-detection. Am Psych., 1981, 36, 488 - 496.
96. Feinstein A.R.: The haze of Bayes, the aerial palaces of decision analysis and the computerized Ouija Board. Clin. Pharm. Ther., 1977, 21, 482 - 496.
97. Gorry G.A.: Modelling the diagnostic process. J. Med. Educ. 1974, 45, 293 - 302
98. Habbema J.D.F., Hilden J., Bjerregaard B.,: The measurement of performance in probabilistic diagnosis. I. The problem descriptive tools and measures based on classification matrices. Meth. Inf. Med. 1978, 17, 217 - 226.
99. Biorck G.,: The essence of the clinician's art. Acta Med. Scand. 1977, 201, 145 - 147.
100. Nisbett R.E., Wilson T.D.,. Telling more than we can know: Verbal reports on mental processes. Psych. Review, 1977, 84, 231 - 259.
101. Pauker S.G., Gorry G.A., Kassirer J.P., Schwartz W.B.,: Towards the stimulation of clinical cognition. Taking a present illness by computer Am. J. Med., 1976, 60, 981 - 996.
102. Dudley H.A.F., Pay-off, heuristics and pattern recognition in the diagnostic process. Lancet, 1968, 723 - 726.
103. Edwards,I.A.W.:Diagnostic Procedures.History and symptoms of esophegeal disease.,in:Vantrappen,G.,Hellemans,J.:Diseases of the Esophagus Berlin,Springer,1974
104. Feinstein A.R.,: Clinical judgment. Baltimore, Williams & Wilkins, 1967.
105. Card W.I.,: The diagnostic Process. J. Roy, Coll. Physcns. London, 1970, 4, 183 - 187.
106. Slovic P. : Toward understanding and improving decisions. In: Howell W.C. & Fleischman E.A. Human performance and productivity: Information processing and decision making. Hillsdale, Lawrence Erlbaum, 1982.
107. Gross R.,: Was ist eine Krankheit? Gibt es stabiele Krankheitsbilder? Metamed 1, 1977, 115 - 122.
108. Elstein A.S., Shulman L.S., Sprafka S.,: Medical Problem Solving, Harvard University Press. Cambridge (Mass). 1978.
109. Goldberg L.R.,: Man versus model of man: A rational, plus some evidence, for a method of improving on clinical inferences. Psych. Bull, 1970, 73, 422 - 432.
110. Smith D.H., McWhinney I.R.: Comparison of the diagnostic methods of family

physicians and internists. J. Med. Educ., 1975, 50, 264 - 269.
111. Gale J., Mardsen P.,: Diagnosis : Process not Product, in: Sheldon, Brooke & Rector: Decision making in General Practice, London, McMillan 1985.
112. Gerritsma J.G.M., Smal J.A.,: De werkwijze van huisarts en internist. Een vergelijkend onderzoek m.b.v. een interactieve patienten simulatie. Utrrecht,W.U.Bunge,1978.
113. De Vries Robbe,P.F.: Medische Besluitvorming, een aanzet tot formele geneeskunde, Groningen,Dissertation,1978.
114. Werf v.d. G. Th.: Geneeskundige Oordeelvorming, Amsterdam, Rodopi, 1984.
115. Hull F.M.: Diagnostic Pathways in General Practice, J. Roy, Coll. Gen. Pract., 1972, 22, 241 - 257.
116. McWhinney I.R.,: Problem-solving and decision-making in Family Practice. Can. Fam. Physician, 1979, 25, 1473 - 1477.
117. Reichertz P.L.,: Computerdiagnostik: Moeglichkeiten und Grenzen, in: Verhandlungen der deutschen Gesellschaft fuer innere Medizin. Muenchen, Bergmann, 1978.
118. Gick M.L., Holyoak K.J.,: Analogical Problem Solving. Cognitive Psychology, 1890, 12, 306 - 355.
119. Lusted L.B.,: Clinical decision making, in: Dombal de F.T. & Gremy F.,: Decision making and Medical care. Can information-science help? Amsterdam, N.Holland Pub. 1976.

CHAPTER III

WAYS OF REASONING

In science we can distinguish two main ways of reasoning : deductive and inductive. The use of scientific methods in medicine seems obvious, because 'the power of science is to make contact with reality in nature by recognizing what is rational in nature! (1). (The question of metaphysical aspects is beyond the scope of this study). We can regard as scientific any purposeful human activity designed to provide tentative and refutable hypotheses about the nature of events, which means, in clinical medicine, the name of a disease explanatory to a number of observations to the patient. According to Feinstein (2) the diagnostic reasoning in modern medicine is a process of converting observed evidence into the names of diseases. The evidence consists of data obtained from examining a patient and substances derived from the patient; the diseases are conceptual medical entities that identify or explain abnormalities in the observed evidence. The diagnostic reasoning also comprises the strategy and tactics to evaluate evidence, select manifestations and adjunctive tests, evaluate them, and ultimately choose the name of the diagnostic entity.

But these strategies and tactics as inference processes in a problem solving situation are largely unknown. Inference processes, especially the everyday inference, may be viewed as a reflection of a more general contrast that can be drawn between explicit and implicit inferences. Inferences that underlie a scientific way of reasoning are often slow, voluntary, and at the forefront of awareness; they are explicit. The inferences that underlie the daily process of perception and comprehension are rapid, and outside conscious awareness: they are implicit (3).

Most patient-physician encounters seems to be of an implicit character. Physicians appears to dig into their bag of cognitive tricks, and use whatever strategy seems most appropriate in a particular situation at a given moment. This varied repertoire used in medical inference has not yet been decomposed into sets of formal procedures, nor has it been shown to be decomposable in such a manner (4). Because of this implicit character physicians are largely unable to accurately break down their diagnostic thought processes into explicit, understandable steps. Neither can they appeal to the fact that there is a generally accepted problem situation, for there is no such thing as the one fact that is generally accepted.

It seems that the whole process of diagnosing and confirmation ultimately relies on our own accrediting of our own vision of reality.The physician is merely drawing a portrait of his conceptually prefigured conclusions, while he does not know how he arrived at them. This is the usal process of unconscious trial and error by which we feel our way to success and may continue to improve on our success without specifiably knowing how we do it, for we never meet the causes of our success as identifiable things which can be described in terms of classes of which such things are numbers (1). We can e.g. assimulate a symptom as a clue if we believe it to be actually useful to our diagnosing. The act of personal knowing can sustain this relation only because the acting person believes that it is apposite; that he has not made them but discovered them. The effort of knowing is then guided by a sense of obligation towards the truth: by an effort to submit to reality.

Moreover, since every act of personal knowing appreciates the coherence of certain particulars, it implies also submission to certain standards of coherence. All personal knowing appraises what it knows by a standard set to itself (1). This reappraise that we credit ourselves with much wider cognitive powers than an objectivist conception of knowledge would allow, but at the same time reduces the independence of human judgment far below the extent claimed traditionally for the free excercise of reason. It is also useless to accumulate

more evidence unless we can first master what has been given so far (1).

We still continue to feel that there is some consistent relation between our beliefs and the factual evidence presented to us. We can regard this (with Hume) as a mere habit, without acknowledging any justification of the convictions expressed by this habit. The reflecting person is then caught in an insoluble conflict between a demand for an impersonality which would discredit all his commitment (to the case) and an urge to make up his mind which drives him to recommit himself (1). In this sense there will always be some discrepancy between someone's belief and the truth. Bertrand Russell defines 'truth' cynically as a coincidence between one's subjective belief and the actual facts; and it is impossible to say how the two could even coincide.

The answer to this dilemma may be that the actual facts are accredited facts, as seen within the commitment situation, while subjective beliefs are the convictions accrediting these facts as seen non-committally by someone not sharing them. Most of the time physicians do not reconsider their assertions (diagnoses, therapies), because of their personal commitment to the statement they have made. Reconsidering the statement means denying one's personal belief in it. The utterance "I believe p" expresses more aptly the heuristic conviction, while "p is true", will be preferred for affirming a statement taken from a textbook of medicine. So we would never use a hypothesis which we believe to be false, nor a policy which we believe to be wrong.

But this means to Polanyi (1) that any act of factual knowing presupposes somebody who believes he knows what is being believed to be known. This person is taking a risk in asserting something, at least tacitly, about something believed to be real outside himself. Any presumed contact with reality inevitably claims universality. And universality opens possibilities for creating a system, a rational one based on scientific logic, or a system based on subjective beliefs. The remarkable stability of the latter one is based on two factors: objections to the (system of) beliefs can be met one by one, because of the circularity of such systems : any contradiction between experience and one mystical notion is explained by the reference to other mystical notions.

The second aspect of stability arises from an automatic expansion of the circle in which an interpretative system operates. It readily supplies elaborations of the system which will cover almost any conceivable eventuality, however embarrassing this may appear at first sight. Some scientific theories possess these self-expanding capacities. Nevertheless, the inborn urge of every human being toward universality, the claim of subjective experience to be applicable as a law of nature to all people, or at least a large group of people, is obvious. It is partly based on the conviction that we applied solid and accurate observation, and that we correctly inferred the conclusion from the data.

It is tempting to think that accuracy of observation automatically yields improved certainty of predictions. Accurate observations are essential to accurate predictions, but certainty is another matter. The certainty of a conclusion is the certainty of the weakest link in the chain of deduction, i.e., of the least certain premise (5). The rules of inference, largely implicit, are like maxims: they can serve as a guide to the practical knowledge of an art, but they cannot replace the knowledge (1).

Practical wisdom is thought to be more truly embodied in action than in rules of action. The physician's action is considered more authentic than what he says he is doing. Without explicit knowledge of the rules, the practioner is more like a connoisseur. And connoisseurship, like skill, can be communicated only by example, not be precept. It would be science's prime concern to elicit these rules. But, as Medawar (6) says, "one cannot but wonder at the fact that scientists themselves are so seldom deeply interested in scientific methodology."

A scientific theory must refer to the mechanisms of nature, not just to the quantitative results obtained by studying those mechanisms in action. (5). It is assumed that in logic we study the ideal form of reasoning. This form includes

trying to find the rules for valid and sound arguments, rules which can direct our thought in the pursuit of truth. What we want to avoid above all else is the drawing of false conclusions from true evidence. The point where one starts in a piece of reasoning will be those statements or principles which one supposes one knows, or for the moment pretends one knows in order to see what are their consequences (5).

It also tells us that a physician, in the supposition a hypothesis or conclusion based on observed evidence to be true, must be able to predict a number of elements or positions to be correct; he then forms his law and tests it against the evidence. The active perceiving physician can fall, at least, into possible traps with regard to the presented evidence: subjectivity and error, which can give: 1) a correct satisfaction of normal standards; 2) a mistaken satisfaction of normal standards; 3) action or preception satisfying subjective, illusory standards; and 4) mental derangement issuing in meaningless action (1).

Building his diagnostic system from observed evidence and knowledge, the physician can go astray with regard to the application of the inference rules. We can distinguish:

1) correct inference within a true system;

2) erroneous conclusions arrived at within a true system (like an error committed by a competent physician);

3) conclusions arrived at by the correct use of a fallacious system. This is an incompetent mode of reasoning, the result of which has subjective validity;

4) incoherence and obsessiveness: an erroneous reasoning in a system of systematic delusions (1).

There are several traps which can lead the doctor to fallacious prediction. It requires a firm conception of explicit rules, to elicit the medical reasoning process, or, as Popper says, 'in order to deduce predictions one needs laws and initial conditions: if no suitable laws are available or if the initial conditions cannot be ascertained, the scientific way of predicting breaks down.

Paragraph 2

HYPOTHESES

Levine's hypothesis theory states that the basic assumption is that, in solving any problem, the subject samples hypotheses from a universe of hypotheses until the problem is solved (7). According to Harre (5) science has its origin in a cloud of conjectures about how things go on. This body of hypotheses is progressively whittled down by the work of experimentalists whose results falsify certain of its components. This view contradicts the so-called Bacon's myth: a scientific method starts from observation and experiment and then proceeds to theories. Bacon's method rested on the induction by simple enumeration. Collecting all sorts of cold, warm and intermediate temperatured bodies and elements he hoped by listing all the qualities of these bodies and elements to find (induct) universal laws of nature with regard to heat.

It must be said that he himself abandoned this method and adapted another inductive process: induction by elimination, without changing the fundamentals of his theory. Contrasted to this conception is the deductive way of reasoning which starts with theories as "bold ideas, unjustified anticipations, and speculative thought as our only means for interpreting nature" (8) and from theory a downward explanation to the facts. Every experiment to the explanation is planned action in which every step is governed by theory. Obviously we deal with two types of hypotheses: inductive and deductive ones, according to the ways of reasoning.

Inductive hypotheses cannot be regarded as scientific as in an

hypothesis-deductive reasoning process (5). In the inductive case many facts can lead to many laws, all equally correct, because they are derived from the same source, the acquired facts. One of the postulates in induction is leading to a solution according to the Principle of Simplicity: only the simplest law should be accepted, a law that proves itself upon the acquired facts, which in their turn give rise to a law etc. (circular reasoning). In medicine the improving or the deterioritation of a patient is based upon a prediction which covers only a (limited) number of facts. Every alteration of the facts can generate a series of hypotheses each justified and verified by the underlying facts, especially when they are so vaguely worded that they might with equal facility explain a dozen other disabilities as well. Their complete lack of logical immediacy deprives them of any serious explanatory power (9).

In the deductive way a hypothesis is like the premise of an deductive argument, and as soon it has been explicitly formulated it can be elaborated into a theory and tested deductively. If the predicted consequences do not follow, the hypothesis can be rejected as an explanation. This is the essence of the hypothetico-deductive method (7).

The crucial question, however, is where the hypotheses come from in the first place? The traditional answer is opposed to the deductive method: they come from the method of induction. Reaching a hypothesis, let's say a bold statement, including predictable consequences, demands a certain amount of information, although it need not to be derived from systematic observation. For problem definition an insight in the case is needed. Without this information it is only "jumping to conclusions". Popper's emphasis on bold conjectures necessarily disregards the kinds of waywards thinking in which conjectures may have their origin.

We believe that there is a personal universum of hypotheses from which a (or some) particular hypothesis(es) arise(s) when confronted with a (recognizable) problem. In this way, a hypothesis can be thought of as a draft, and the everyday scientific process consist essentially of examining the logical consequences of hypotheses to see if what results from them or what they predict does in fact correspond to real life. In this process falsification - in effect disproof - is logically a stronger process than corroboration, which always falls short of completeness, so that no general hypothesis, however solidly established it may seem to be, is even proved beyond the possibility of further critical examination or questioning (6).

Opposite to this conception of falsifiability is the view that hypotheses can be refashioned in the light of the degree of correspondence with reality of their implications and logical predictions. This conception couples a universal statement (hypothesis) to a probable correspondance to (finite number of) the facts. According to Popper, probability hypotheses do not rule out anything observable; probability estimates cannot contradict, or be contradicted by a basic statement; nor can they be contradicted by a conjunction of any finite number of basic statements; and accordingly not by any finite number of observations either.

Let us assume that we have proposed an equal-chance hypothesis (50-50% being right or wrong) for some alternatives ; e.g. that we have estimated that tosses with a certain coin will come up '1' or '0' with equal frequency, so that $F(1) = F(0) = 1/2$; and let us assume that we find, empirically, that '1' comes up over and over again without exception; then we shall no doubt abandon our estimate in practice, and regard it as falsified. But there can be no question of falsification in a logical sense. For we can only observe a finite sequence of tosses. But this means that probability hypotheses are unfalsifiable because their dimension is finite (8). Every hypothesis may not be falsifiable but it might be in a unilateral way falsifiable within a probability segment, say - a certain amount of time. When a certain observation does not come up immediately, it can be said that it falsifies a probability statement unless the boundaries (of time) are infinite, which they aren't practically.

In these cases the test to the hypothesis can be temporarily defined as verified; a verification made for the existential consequences. It stands to the

basic statement in the relation which appears to be characteristic of probability statements. But this also means that any critical verification of a scientific statement requires the same powers for recognizing rationality in nature as does the process of scientific discovery and deduction, even though it exercises these at a lower level. Unfortunately, when philosophers analyze the verification of scientific laws, they invariably choose a specimen of such laws as are not in doubt, and thus inevitably overlook the intervention of these powers. They are describing the practical demonstration of scientific law, and not its critical verification (5).

But can a hypothesis itself form an element of a sequence of hypotheses? One way of interpreting it would take, as elements of such sequences, the singular statements (constituting the hypothesis) which can contradict, or agree with, the hypothesis. The probability of this hypothesis would then be determined by the truth frequency of those among these statements which agree with it.But this would give the hypothesis a probability of 1/2 if, on the average, it is refuted by every second singular statement of this sequence.There are two expedients to that:

- an estimate of the ratio of all tests passed by it to all tests which have not yet been attempted.But here this estimate can be computed with precision, and the result is always that the probability is zero;
- an estimate upon the ratio of tests led to a favourable result to those which led to indifferent result - which produce no clear decision (8).

The reason why this attempt fails is that the suggested definition would make the probability of a hypothesis hopelessly subjective; the probability of an hypothesis would depend upon the training and skill of the experimenter rather than upon objectively reproducible and testable results, as defined for deductive reasoning.

Paragraph 3

DEDUCTIVE REASONING

A special technique in trying to solve certain sorts of problems is deductive inference. For medicine this method has been claimed to be the only possible one (Elstein et al. 10). A valid deduction is one in which the conclusion is a logical result of the premises: if the premises are true, then the conclusion is necessarily true. It is usually stated in the form of a syllogism:

All Greek are mortal (Major Premise)
Socrates is a Greek (Minor Premise)

Socrates is mortal (conclusion)

Or in a more abstract form :

$\forall x\ (Px \text{ -- } Qx)$
Pa

Qa

Deductive reasoning means: from theory to facts. This can be examplified physically and medically. The physical example is borrowed from Braithwaite, describing the Galilean system.

The system has one highest-level hypothesis:

I Every body near the earth freely falling towards the Earth falls with an acceleration of 32 feet per second per second.

From this hypothesis follows:

II Every body starting from rest and freely falling towards, the Earth falls 16 t2 feet in t seconds, whatever number t may be.

From II there follows in accordance with the logical principle permitting the application of a generalization to its instances, the infinite set of hypotheses :

IIIa Every body starting from rest and freely falling for 1 second toward the Earth falls a distance of 16 feet.

IIIb Everybody starting from rest and freely falling for 2 seconds toward the Earth falls a distance of 64 feet. And so on.

In this deductive system the hypotheses at the second and third levels (II,IIIa, IIIb) follow from the one highest-level hypothesis (I); those at the third level (IIIa, IIIb) also follow from the one at the second level (II).

The hypotheses in this deductive system are empirical general propositions with diminishing generality. The empirical testing of the deductive system is effected by testing the lowest-level hypotheses in the system. The confirmation or refutation of these is the criterion by which the truth of all hypotheses in the system is tested. The establishment of a system as a set of true propositions depends upon the establishment of its lowest-level hypotheses. E.g. when the distance in hypothesis III is different from 16 feet, the hypothesis - and, therefore, the whole system - must be rejected (9).

The medical example is derived from Scandellari & Federspil (11). They describe the 'Thyreoid' System:

1. which patient with an autonomous thyreoidal adenome has tachycardia? (Explanandum);

2. the thyreoid gland produces hormone capable of augmenting the cardiac pulse rate, if not otherwise diseased. This patient presents a thyreoidal adenoma. This adenoma produces a raised amount of hormones(Explanans) ;

3. this patient has tachycardia because of a raised serum hormone-level (Conclusion).

Although the form is the same, the differences with the former example are striking. E.g. the highest-level hypothesis cannot be a universal one. Not every patient with a thyreoidal adenoma has tachycardia. Besides, the question rises whether the patient is an exemplar from a particular population. The second-level hypothesis can only function as explanatory when the condition "if not otherwise diseased" is added. As the exact figure for the raised amount is not defined, this hypothesis can only be a probabilistic one. From the second and the lowest-level hypotheses several propositions can follow, all to be submitted to severe tests.

Although the form looks like a deductive one, the methodology certainly is not deductive. The various level hypotheses do not follow unequivocally from each other. Instead, a supernumerary hypothesis (heart otherwise diseased) is added. But to do this would be to make the observed facts evidence for a set of hypotheses which includes one which played no part in their deduction from the set, and would then make them indirect evidence for the supernumerary hypothesis and for its consequences. Since the supernumerary hypothesis might be any generalization whatever, this would have the undesirable result that any observable fact would be indirect evidence for any generalization whatever.To avoid this result one has to stipulate that each of the highest level hypotheses in a system must be necessary for the deduction of the lower level hypothesis in the system. Similarly, two systems whose sets of highest level hypotheses have no hypothesis in common, must not be conflated (9).

Symptoms indicating different diagnostic hypotheses, especially differing organ systems, can only be explanatory to any of the specific hypotheses considered in the particular case. Where in the lowest-level a general proposition (of the II-level) is applied to a special case (the application principle), the deduction of II from I is made by using integral causal relationships as laid down in medical science itself. The derivation from I to II is implicit to the medical knowledge base. Regrettably, there are only few of this kind of logical derivations in medicine. Struggles in this area can be examplified by the attempts of Betacque & Gorry (12) and de Vries Robbe (13) in nephro-pathology. The derivation from II to III comes from knowledge on the basis of statistics from specific populations of this particular disease. It is the conjunction of the two knowledge bases which give rise to many medical problems and shows the trifling results of medical decision making.

The general characteristic of a deductive system is that the logical strength of the hypotheses increases with the height of their level. Sometimes, although each of the hypothesis at a certain level is weaker than the one hypothesis at the next higher level (from which they are all deducible), yet the conjunction of them all is equivalent to the highest level hypothesis. This will happen when there are a limited number of special cases of the higher level hypotheses, each of which is asserted by one of the lower level hypothesis (9).

One of the main purposes in organizing scientific hypotheses into a deductive system is to make direct evidence for each lowest level hypothesis indirect evidence for all other lowest level hypotheses; although no amount of empirical evidence suffices to prove any of the hypotheses in the system, yet any piece of empirical evidence for any part of the system helps towards establishing the whole of the system. The ultimate testing of a hypothesis is against a particular case. The hypothesis system consists out of a number of levels, the highest being the most general one, the lowest the most detailed. (The levels in the medical diagnostic system will be discussed in chapter VI). The system can be described as: (with a slight modification to Braithwaite)

1) according to the evidence the underlying disease is of origin A (the pathophysiology of an organ or organ system).There is e.g. a lung disease, or there is a hyperfunction of the thyreoid gland). From this hypothesis follows:

2) the observed evidence (symptoms and signs) can only be explained by the

B explanation of the A origin.The former disease can e.g. only be explained by an inflammation of the lungs, the second by nervousness, weakness, sensitivity to heat, sweating, restless overcapacity, weight loss, tremor, palpitation, stare, lid lag, and exophthalmus (the Merck Manual, 12th ed.).

3a) in case of B explanation there must be a C test of D level confirmative of the B (e.g. the demonstration of pneumococci in the sputum of the patient, or a raised blood level of thyreoid hormone corresponding with a certain weight loss).

3b) In case of B explanation there must be a C test of E level ($E \neq D$) refutative of B (e.g. when microorganisms of the origin Bact. pneumococci are demonstrated there cannot be a raising antibody titre for a different microorganism, or, a raised blood level of thyreoid hormone is not a consequence of glycosuria in case of weight loss).

Tested against this particular case:

f1 : in all cases of A disease the level D of test C is within the range R (e.g. above a certain number of microorganisms, or, more than x pounds of weight-loss.)

g1 : in this case the level D of the test C is within the range R.

The conjunction of f1, and g1, which conjunction will be called an instance of a hypothesis, support the hypothesis.But it may be clear that this one piece of evidence is insufficient to prove the hypothesis. It would only do so if the hypothesis were a logical consequence of the conjunction g1 with f1. This is not always the case. It is perfectly possible for the hypothesis to hold in this instance, but to be false as a general proposition. This is the case however many times the hypothesis is confirmed. However many conjunctions of g1 with f1, g2 with f2 etc. have been examined and found to confirm the hypothesis, there still will be unexamined cases in which the hypothesis might be false without contradicting any of the observed facts.

In suitable cases we may say that the hypothesis is established by the empirical evidence of its instances, but it never proves it in the sense that the hypothesis is a logical consequence of the evidence. The situation is different if g1 is observed false. For the conjunction of not-g1 with f1 is logically incompatible with the hypothesis being true. Calling this conjunction a contrary instance of the hypothesis, we may say that a hypothesis is proven false, or refuted, by one known contrary instance.

From the examples it may become clear that where in the pneumonia case there is some logical system, in the thyreoid case this internal logic is absent. Not only there is no unambiguous explanation of the symptoms to the disease, there is also no obligate weight loss in all cases of thyreoidal hyperfunction. So it seems that a deductive system is not only case dependant but also subject to the state of medical knowledge. Where several underlying pathophysiological processes are unknown or partly known, it is hardly plausible that deductive systems can be built or function in medical problem solving practice.

Paragraph 4

INDUCTIVE REASONING

We often say: "Well, it is just plain logic!" explaining some problem to another person, without realizing that our explanation is only our 'logical' interpretation of the problem. Obviously this type of 'logic' is different from the 'logic' in the deductive system. The common, daily 'logic' is a 'logic' stripped from nearly all conditions, consequences and precautions as attached to the deductive form. It expresses an insight, a firm belief in one's own statement, being not only casual but of general importance also.

Inductive inferences are supposed to be based on observations, and they are

more general than the observations on which they are based. E.g. "red sky in the morning, shepherd's warning". From having frequently observed the conjunction of two events (colour and weather) we feel compelled to assert a connection between them. As Hume pointed out, there is nothing else than 'habit' or 'custom' to justify induction. It has no logical warrant. In the 17th and 18th centuries, especially in the Anglosaxon sphere, inductive inference was supposed to provide the method of the empirical sciences. Philosophers such as Francis Bacon (1561-1626), and John Stuart Mill (1806-1873) formulated a variety of methods to secure its validity. Bacon formulated two types of inductive inference:

1) induction by simple enumeration. In order to find the real character of heat, Bacon listed all sorts of cold and warm bodies, objects etc., hoping that one day a "happy thought", would present itself. From this "happy thought" he expected to find various laws which stepwise could be constructed to some sort of 'logical system'.Later, Bacon himself castigated induction by simple enumeration as "childish and puerile". This kind of induction draws conclusions by counting positive instances of the association between events. Bacon himself proposed another type:

2) induction by elimination. This induction must analyse nature by proper rejections and exclusions; and then after a sufficient number of negatives come to the conclusion on the affirmative instances.

Inductivism treats laws of nature as general statements of correlations among phenomena, and evidence as particular statements of correlations among observables(5). Unlike the hypothetico-deductive, it merges the distinction between acquiring a hypothesis and submitting it to test. As Medawar put it: "Inductive logic embodies both a rite of discovering and a ritual of proof". Karl Popper points out that answer can only be given to questions; they do not emerge in some mysterious way from the assembling of data. When we collect data we do so with an implicit or explicit question in mind. Popper makes this point for his students: "My experiment consists in asking you to observe, here and now. I hope you are all co-operating and observing! However, I fear that at least some of you, instead of observing, will feel a strong urge to ask: what do you want me to observe?" (14).

Observation is not an unprejudiced principle, but it conveys an a priori conception of nature. Inductivism is a complex of beliefs of which the salient points are: Truth lies all around us in Nature, so that the scientist's main task is to discern and record matters of fact and then to classify and appraise them according to certain more or less well-defined rules - whereupon the Truth will certainly reveal itself (6).

Now it will become obvious from a logical point of view that induction leads to circular reasoning. Hypotheses arise from experience which in its turn becomes evidence to test the hypothesis that has generated from this evidence, and so on. To justify an inductive hypothesis we should have to employ inductive inference; and to justify these we should have to assume an inductive principle of a higher order,etc. These attempts to base the principle of induction on experience break down, since it must lead to an infinite regress (8). Indeed, inductive judgment cannot be justified in the way deductive judgment can. The judgment may very well be true, and serve as a guide to action, but it cannot be shown to be true. The fact that it works tells little about its truth, it just tells us that it works, and the explanation why it works may be very different from what we think it is (15). But why, you will ask, so much attention to a method which seems scientifically inferior? The answer is that inductive reasoning is the most frequently applied way of thinking in daily life.The weather forecast, the politician's utterances, as well as the physician's predictions, almost all are based on inductive reasoning. We have got acquainted to it. We have learned this way of reasoning implicitly from parents, at school, from newspapers and television etc. It has, as Hume said, become a 'habit of mind'.

And, indeed, what is more appealing than looking about and observe, and while observing the hypotheses as universal laws of nature come to you all by themselves. But the 'habit of mind' also make us believe that we can predict

things on the mere ground of assumed foreknowledged similarity. It is far from obvious, from a logical point of view, that we are justified in inferring universal statement, from singular ones, no matter how numerous; for any conclusion drawn in this way may always turn out to be false, no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white (8).

Nevertheless, the prediction that the next swan we shall see is a white one, can momentarily fulfil the demands of plausibility, our reasonable belief. It is in this direction that philosophers have created conditions to strengthen the method and conclusions of induction. An important exponent of the inductive method is John Stuart Mill. He created "Mill's canons": (as described by Harre (5).

I The canon of agreement : If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause (or effect) of the given phenomenon.E.g. if patients suffering from an infectious disease have the same bacteria in common,the particular bacteria is assumed to be the cause of this specific infectious disease.

II The canon of difference : If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur have every circumstance in common save one, the one occurring only in the former,i.e. the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause of the phenomenon. E.g.if two depressive patients, both from the same types of family and social background, differ in their marriage relation, this relation is assumed to be the cause or the effect of the depression.

The method of agreement stands on the ground that whatever can be eliminated is not connected with the phenomenon by any law. The method of difference has for its foundation that whatever cannot be eliminated is connected with the phenomenon by a law. These two main canons Mill offers as the principles or among the principles of inductive reasoning, for, having found the cause we have found the law, or so he thinks.Here are principles by which we can pass from facts to general laws (5).

These canons illustrate the typical character of inductive reasoning: the first canon stands for a heuristic search program eliminating those symptoms and signs which are not consistent with an a priori statement. The second canon tests this statement for internal consistency in order to confirm it. When there is more than 1 causal factor, one has to do another experiment, with the exclusion of one (or more) of the factors involved.

So every stage or cycle in the physician's reasoning process is an experiment in including or excluding a (number of) factors. The dilemma is when there are actually two or more factors equally involved in the case under investigation. One can assume a ranking order, according to probabilities or to define some ranking order. But this again leads to a series of new experiments. This is one of the reasons why Popper postulates the inductive reasoning as to be leading to infinite regress.

And if we find another, a third, fourth etc. factor, how can we be sure that these are the explanatory factor? We judge by force of certainty, whether we are sure or not. Our belief in this theory will depend on how certain we are that we have uncovered the true mechanism. Whether any particular application of Mill's canons yield information of value is determined by how good a theory we have to explain the processes we are investigating. It looks as if Mill's canons are at best a preliminary to deeper studies of scientists (5).

We cannot assume laws produced by inductive reasoning as true. Suppose the only reason for believing an inductive law is the experiment one did. But there is nothing in (most) experiments that leads us to think that we would get similar results in another year, or in another place, with another person etc.

And yet enunciating the results as laws certainly suggests a strong expectation that the process under investigation has worked according to this pattern and will always do so (5). Doctors follow tracks of past successes or failures with the supposition of repeating patterns in similar circumstances. And inevitably, the "laws" based upon coincidences or faulty experiments, almost always turn out to have exceptions. Harre listed the arguments of inductivism into three principles:

1) the principle of accumulation : augmenting knowledge to an existing knowledge base does not alter the latter.

2) the principle of induction : inference of laws from accumulated singular facts can lead to the inference of true laws. The laws of Nature are the codified and generalized facts. As Mach put it: "They are the mnemonic reproduction of facts in thought". In modern science the operation of this principle is often seen in the effort to obtain numerical data and then find (mathematical, statistical etc.) functions to express them;

3) the principle of instance confirmation : our belief in the degree of plausibility of (or our degree of belief in) a law is proportioned to the number of instances that have been observed of the phenomenon described in the law.

Braithwaite (9) constructed a range of degrees of belief and the possibilities for sufficient criteria for the validity of an inference. Braithwaite recognizes three grades of belief: Belief, reasonable or rational Belief, and Truth. Belief is taken generally: the subjective, intuitive feeling of being right. We speak of reasonable Belief when the Belief forms part of the believer's rational corpus, or when that belief in the conclusion will inferentially be supported by the belief in the premise.

These grades can be combined with three features of the inductive inference process which are:

a) the premise, the assertion of the evidence for the inductive hypothesis, i.e. the successful past application of induction;

b) the assertion of the effectiveness of the 'inferential policy' leading from the premise (a) to

c) the conclusion of the inductive inference.

Combination leads to ten possibilities for the validity of an inductive inference, which can be grouped into two headings:

Subjective validity of the inductive hypothesis , when the person subjectively believes the effectiveness of the inferential policy;

Objective validity of the inductive hypothesis , when the effectiveness of the inferential policy is true.

But whatever corollaries have been described, it does not alter the fact that the inductive statement remains part of the private and personal thinking processes individuals use to solve their - scientific - problems. The facts upon which the hypotheses are based, are not genuinely independent. We can believe them, or reasonably believe them, but they are private and do not form a public domain of knowledge. If they are public facts they are affected by all sorts of influences, particularly from previous knowledge, and upon which their exact form, and our confidence in them, depend. Inductive hypotheses, therefore, are not 'strictly valid', but they can attain some degree of 'reliability' or of 'probability'.

Are there any possibilities to establish this 'degree of probability'? Can we scientifically corroborate the inference andd the conclusion of our reasoning process? Several people suggest that "Probabilistic Reasoning" can help us in the dilemma between our natural inclination to confirmation and the scientific demand for refutation.

PROBABILISTIC REASONING

The number of techniques for analysing data has increased enormously over the past decades. These techniques are almost all based on the probability theory, and they have been developed so as to end up with some kind of inference, some statement about the real world. The concept of probability, however, is rather abstract and hence the question arises: what is the interpretation of the notion of probabiliity in real world situations? In answering this question we reach the foundations of statistics and, as is the case in many sciences, it is precisely here that differences in opinion appear. The foundations of this concept have been object of much discussion. Different interpretations have resulted in different axiomatic systems.

If one asks a scientist whether the term "probability" as used in science always has the same meaning, one will find a curious situation. Practically everyone will say that there is only one scientific meaning: but when one asks that it to be stated, at least two different answers will come forth. The majority will refer to the concept of probability used in mathematical, or 'objective' statistics as it is generally applied in the natural sciences. There is, however, also a minority of those who regard a certain subjective, non-frequentistic, concept as the only scientific concept of probability.

Usually we distinguish three types of probability:

1) Objective, or mathematical, or statistical probability;
2) Inductive probability; and
3) Subjective, or personalistic probability.

1) In the simplest cases, probability in this sense means the relative frequency with which a certain kind of events occurs within a given reference class, customarily called "population". This theory of probability originates from gambling problems. In such situations the following definition of Laplace applies: " The probability of an event occurring is equal to the ratio between the number of cases favourable to this event and the total number of possible cases, provided that all these cases are mutually symmetric". This latter requirement places a serious restriction to the general applicability of the statistical probability as we shall discuss later on.

2) Inductive probability is ascribed to a hypothesis with respect to a body of evidence. Inductive probability measures the strength of support given to a hypothesis H by E (evidence) or the degree of confirmation (or credence) of H on the basis of E. The hypothesis may be any statement concerning unknown facts, say, a prediction of a future event: tomorrow's wheather, the outcome of a planned experiment, or the therapeutic plan for a patient. Any set of known or assumed facts may serve as evidence; it consists usually of the results of observations which have been made.

3) Subjective probability can be viewed as a subsequence of the inductive probability. It introduces intuitive weighings into our analysis. The personalistic probability model is characterized by a number of consistency requirements. It is just in this respect that this model has been criticized. We cannot expect that our intuition will automatically obey such requirements. For repetitive events, it is quite possible that the observable relative frequency will be completely different from our original intuitive feelings.

We shall now discuss these concepts more in detail.

Statistical or objective probability, or probability of kind of events,

characterizes an objective situation; it is a translation of statistical laws. It is the mathematical approximation of the conceptions of events governed by chance. It is a theory of certain chance-like or random sequences of events or occurrences i.e. of repetitive events of similar character and in similar circumstances. The second part of Laplace's definition refers to this requirement of randomness. In a sequence of events each event has a certain character or property e.g. a number on a die, head or tail of a coin. In a sequence of die-casting (n) the property six comes up a certain number of times (a). The ratio between the number of occurrence of the property and the number of casts is the Relative Frequency a/n, e.g. in case of the die-casting 1/6.

It is easily understood that when the sequence of casts becomes smaller and smaller, the "n" approximating to 1, the outcome of the calculation of the relative frequency will change respectively. The presumption of a relative frequency of 1/6 for the die-casting can only be made assuming the sequences to be infinite. So there is another element we have to know about a sequence of events: the way the property is scattered within this sequence. We shall assert that the property six will come up a number of times in a random fashion. When we denote the property six as 1 and the non-six as 0, the sequence of events can be represented as follows:

1 0 0 1 1 1 0 0 0 0 1 1 0 1 0 1

The scattering can be noted as:

1	1	1	2	3	4	4	4	4	4	5	6	6	7	7	8
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

number of casts.

Here we have a new sequence, which is called after Popper, The Sequence of Relative Frequencies (8). The two estimates, Relative Frequency and Sequence of Relative Frequencies, or property probability, can be calculated under the assertion of a number of conditions.

Von Mises made a substantial contribution by defining the conditions relevant to the frequency approach in probability theory. These conditions are known as Von Mises' axioms which run as follows.

1) "Random" experiment in finite sequences. Definition: A random (or chance) phenomenon is an empirical phenomenon characterized by the property that its observation under a given set of circumstances does not always lead to the same observed outcome (so that there is no deterministic regularity) but rather to different outcomes in such a way that there is statistical regularity. By this is meant that numbers exist between 0 and 1 that represent the relative frequency with which the different possible outcomes may be observed in a series of observations of independent occurrences of the phenomenon. Thus the empirical law of stabilizing relative frequencies holds for this phenomenon.

2) A "Random" experiment implies "stable" relative frequencies. but these frequencies depend on n (as mentioned before). The empirically established fact that relative frequencies stabilize in long sequences of trials is of utmost importance because it has become the basis of a frequency definition of probability. To be more precise, the frequency (or objective) interpretation of the probability formalizes the property of stabilizing relative frequencies.

3) The impossibility of a winning gambling system. To put it differently, we study sequences which are "impredictable" or "random".

People often deny this requirement. People are very much inclined to assume orderly patterns where, actually, randomness is the case. Most gamblers assume some kind of order in roulette, many people predict an observed trend to

continues in the future. All this implies a reference to a type of orderly patterns which such events can simulate only by coincidence. To test the probability of assuming that they have taken place, there is the method of Sir Ronald Fisher for establishing a contrario to the reality of an orderly pattern. This raises the question whether orderliness is merely a coincidence or a man-made assumption. If our appraisal of order is an act of personal knowledge, exactly so is the assessment of probability to which it is allied. This is, of course, quite evident when the ordered pattern is contrived by ourselves (1).

If gambling is continued long enough (infinite sequence), the relative frequencies in the sequence of the cases supposed to be favourable will approach the same limits as those in the sequence of all casts ($\lim n_i/n = P_i \quad n - \infty$).

Probability, according to Von Mises, can, therefore, be defined as the limit of relative frequency in a collective (this term will be defined in the next condition). Polanyi's statement about orderliness refers to this axiom, but differs in the sense of personal imposed orderliness versus the coincidental randomness to be overcome by assuming unlimited boundaries.

These conditions lead to the following definition of a collective.

4) A <u>Collective</u> is a long sequence of identical observations, each observation leading to a definite numerical result, provided that the sequence fulfils the two conditions: existence of "stable" frequencies and "randomness".

As can be inferred from this definition, Von Mises replaces the infinite sequences condition by a finite one. Within the finite sequence of reasonably long extension it also defines the limiting frequencies and the randomess. It sets the conditions for samples as they are drawn from their parent populations. Probability calculus teaches us to compute the probability distributions in derived collectives from given distributions in the collectives from which have been derived. By means of four operations (place selection, mixing, partition, and combining) the new collective can be derived from a given one (19).

However, one can raise several objections to the Von Mises' theory, which fall under two headings: the axiomatic structure and the applicability of the theory. The objection to the first heading is put into words by Jeffreys who stated that "a definition is useless unless the thing defined can be recognized in terms of the definition when it occurs. The existence of a thing or the estimate of a quantity must not involve an impossible experiment". This objection is very serious because the notion of probability as defined by Von Mises is completely hypothetical, being a limit in an infinite sequence. It is an abstraction; it can never be verified nor falsified.

The applicability of a model requires that the model has its counterpart in reality. This implies that there are two relevant aspects for the connection between model and reality. Firstly, we need input for the model, i.e. certain theoretical variables should be given values derived from reality. Secondly, conclusions derived from the model should be translated back to the real-world statements. In Von Mises' probability model this means that from a finite number of observations we should be able to derive the probability values in the collective, and from the probability values in new collectives we should be able to derive properties of finite samples (19).

Our question should be whether there is any possibility to conclude anything logically about the relative frequency of occurrence of a particular event in a finite sequence of elements otherwise than in case of the relative frequency will be close to the probability p of that event. But what shall we do when that frequency is not approximately equal to p? To this end, in our opinion, implicitly two rules have been introduced.

a) An event with a very low probability of occurrence will not occur at an individual trial; and

b) All probability properties (e.g. distribution functions) are introduced <u>a priori</u> .

In a frequentistic, or statistical, sense the first rule is often,

implicitly, applied. This situation can be compared with a lottery: if the number of winning tickets is small in comparison with the total number of tickets (low relative frequency), it is thought to be impossible for any ticket to be a winning one. Nevertheless there must be some winning ticket. Von Mises' axioms restrict the domain to the application of mass phenomena where experience has taught us that rapid convergence exists, i.e. the frequency limits are approached fairly rapidly. It might be clear that this situation is not often met in medicine.

Sometimes a statement of statistical probability refers not to an actually existing or observed frequency, but to a potential one, e.g. the probability of obtaining a six when casting a die. There are two possible outcomes:

where sufficiently long series of casts were made the relative frequency would be approximately 1/6 (objective probability);

as a guess whether the next cast will be a six then the only conjecture can be, a six or not, so the probability is 0.5

The second outcome refers to the probability p of a die falling six, where six is a property in a sequence. This probability is of course a probability in an appropriate collective. What we need is the probability distribution in that collective which is the theoretical counterpart of the casting of a six, noting its p-values and putting it back in the urn (sequence). Here we see how a probability distribution, which is often called the prior distribution of p, is of importance because the required probability depends on it. Unfortunately this distribution is usually unknown. The theory can be saved by the remarkable fact that the influence of the prior distribution decreases with n and finally becomes negligible. This is the fundamental reason why inference for large numbers becomes possible without knowing the prior distribution exactly, which follows from Von Mises' theory. Probability statements about single observations or events are excluded by the theory. This implies that it is meaningless to say that the probability of a die falling six in the next cast is equal to 0.5 or to speak of the probability of a certain individual dying next year without the specification of the distribution which is, as we already know, in contradiction with Von Mises' third condition.

Bayes' theorem, as a law of large numbers, seems to solve the problem because roughly speaking, it says: if, in one set of observations, n being a large number, the relative frequency of success (casting six) is equal to r_n, then we expect <u>with great certainty</u> that p is equal to r_n. This means that in the derived collective $\{p_i, r_n\}$ the majority of elements possess an inferred p_i value close to n_i/n. But logically nothing concluded about a specific element out of the collective. For a specific element the theory admits the possibility that p_i deviates substantially from n_i/n. We should, therefore, realize that when a prior distribution is unknown we have to determine how large n should be. Probability statements only make sense if we are able to define a collective. The existence, however, of a collective is purely hypothetical. It can be said that the idea of a collective is an assumption. But if we do make this assumption we run into the following problems:

1) all probability properties (e.g. distribution functions) must be introduced a priori;

2) the number of observations is usually so small that many assumptions cannot be verified;

3) it is usually impossible to increase the number of observations because of structures changing, time is running out, or because of ethical obstacles etc.

4) in this way we construct procedures which lead to solutions consistent with the a priori assumptions.

With the latter problem we run in a circular reasoning process. And there is no way of knowing whether we are far from reality or not.

In medicine we meet a number of obstacles with regard to the objective probability approach among which are: the acquisition of "sufficiently" large numbers and the unequivocally defined similar properties. Moreover, physicians are actually interested in the question whether the next patient has a particular disease given a certain amount of evidence. This situation is ascribed as Inductive Probability. Inductive Probability measures the strength of support given to a hypothesis H by E (evidence) or the degree of confirmation (C) of H on the basis of E, which takes the form of

$$C(H,E) = r$$

or the quantified subjective probability $C(H,E) = r$ is the degree of credence an individual X believes a hypothesis H being true given the evidence E.

The hypothesis may be any statement of a future event. Any set of known or assumed facts may serve as evidence; it consists usually of the results of observations which have been made. In this sense, inductive probability theory, as it is developed by Carnap, is a principle of learning from experience which guides, or rather ought to guide, all inductive thinking in everyday affairs and in science (16). It expresses in quantitative terms our confidence in the outcome of a particular process(16).

The Degree of Confirmation (C) is sometimes reflected in a fair betting quotient, and studied as such. To say e.g. that the hypothesis H has the probability p (say 3/5) with respect to the evidence (E), means that to anyone to whom only this evidence but no other relevant knowledge is available (the principle of "fair bet"), it would be reasonable to believe in H to the degree p or, more exactly, it would be unreasonable for him to bet on H at odds higher than $p / (1-p)$ (in the example 3 :2). A probabilistic diagnosis is such a "fair bet system".

In most cases in ordinary discourse, even among scientists, inductive probability is not specified by a numerical value but merely by the fact as being high or low or, in a comparative judgment, as being higher than another probability. Every inductive probability judgment is related to certain evidence. Probability as understood in contexts of this kind is NOT frequency. If we want to know the outcome of a proposed action, or the plausibility that the next cast of the die will be six we have to assume an orderly trend which contradicts the objective probability theory. The same can be said about the prediction whether a hypothesis in a scientific system may be true or false. This type of (inductive) probability belongs, strictly speaking, not to science but to the methodology of science i.e. the analysis of concepts, statements, theories and methods. It does not occur in scientific statements, concrete or general, but only in judgments about the strength of support given by one statement, the evidence, to another, the hypothesis, and hence about the acceptability, the credence, of the latter on basis of the former (16).

Inductive probability can be defined as a way of judging hypotheses concerning unknown events. In order to be reasonable this judging must be guided by our knowledge of observed events. More specifically, other things being equal, a future event is to be regarded as the more probable, the greater the relative frequency of similar events observed so far under similar circumstances (16).

It seems plausible in daily practice, that from the observation of what has in former instances been the consequence of a certain cause of action, one may make a judgment that is likely to be the consequence of it another time; and that the larger the number of experiments we have to support a conclusion, the more reason we have to take it for granted. Our confidence that a certain therapy will work in a present case of a certain disease is the higher the more frequently it has worked in past cases.

From the viewpoint of Logical Positivism Carnap considered an objective concept as possible: "if a certain probability value holds for a certain hypothesis with respect to a certain evidence, then this value is entirely independent of what any person may happen to think about these sentences, just as any relation of logical consequences is independent in this respect. Consequently a definition of an explanation for probability must not refer to any person and his beliefs but only to the two sentences and their logical properties within a language system" (Carnap). In the form C(H,E) the expressions H and E are logical statements, void of any interpretational meaning. While in deductive logic expressions like "if E then H" occur, in inductive logic these types of expressions will have the form " if E, then H with a degree of confirmation C(H,E)". This conception can only apply in those circumstances that the hypothesis H "if and only if" can be inferred from the evidence E. This situation, however, is rarely met in medicine.

When we rely on experience as source of evidence it must be clear that we cannot determine, at least not with any certainty, to what degree repeated experiments confirm a conclusion (17).

Experience largely reflects the contents of thought not its process, as we have described in Chapter II. In reflecting on the way in which we are performing we may feel to establish rules for our own guidance. But we are, at least for a large part, guided by feelings of intellectual satisfaction, of a persuasive desire, and a sense of personal responsibility. This personal act can only partly be formalized. Such formalization is likely to go too far unless it acknowledges in advance that it must remain within the framework of a personal judgment. But, being a personal judgment, is there any justification for assigning numerical values to the probability judgment?

Principally, we can distinguish two main 'sources' of our hypothetical estimates of frequencies (8):

1) equal chance hypothesis: a hypothesis asserting that the probabilities of the various primary properties are equal: it is a hypothesis asserting equal distribution. Equal chance hypotheses are usually based upon consideration of symmetry. Making this kind of hypothetical estimates of frequency we are often guided solely by our reflections about the significance of symmetry, and similar considerations. This is comparable to Polanyi's conception about the assumption of orderliness.

2) estimate based upon an extrapolation of statistical findings e.g. mortality rates. It is based upon the hypothesis that past trends will continue to be very nearly stable. We 'derive' estimates of probabilities - that is, predictions of frequencies - from past occurrences which have been classified and counted. But these 'derivations' are not logical. What is done is to advance a non-verifiable hypothesis: the conjecture that frequencies remain constant and so permit extrapolation.

The inductive probability theory points out that any hypothesis having a finite initial probability will, if it happens to be true, be confirmed by subsequent evidence until its degree of probability approaches complete certainty. Testing in these cases means that we shall eventually come to believe all hypotheses which are true with a degree of probability approaching certainty. This theory rules out the deductive system which postulates that however many hypotheses are verified when there is one logical impossibility the whole set of hypotheses must be rejected.

But in probabilistic reasoning there is no question of logical impossibility of consequences but a plausible explanation for the hypothesis

given the evidence. Just as the rejection of a statistical hypothesis by observation is not a matter of deductive logic but is a matter of a special rule of rejection (k-rule of rejection), so the choice between alternative statistical hypotheses is not a matter of deductive logic but a matter which requires first choosing a policy before making such choices (e.g. choosing cut-off points on a curve). And it is by no means obvious which policy is best, nor indeed what is meant by calling a policy a 'best' policy for this purpose (9).

Sadegh-Zadeh (18) made clear with the help of a logical example of the co-occurrence of two disease entities in a patient that there cannot be a deductive system because of logical external and internal contradiction. The internal consistency in the deductive system is based upon the logical independence between premises and the conclusion. External inconsistency is due to the attachment of the probabilistic character to the conclusion (the probability of a patient of having a disease), and not to the steps in the inference process.

In a descending hierarchy of hypotheses in a deductive system when each hypothesis is a logical consequence of the preceding one, the rejection of one of the hypotheses involves rejecting all the hypotheses of higher level. Sadegh Zadeh comes to the conclusion that in the probabilistic approach in medicine there is not a deductive but an inductive process.

This raises also the question if it is possible to assign numerical values to the estimates. Carnap states that it is not frequency. John Maynard Keynes thought it highly improbable to assign a numerical value to the probability of any hypothesis. He believed that this could be done only under very special, rarely fulfilled conditions (cited in 16).

According to Popper (8), a numerical probability value cannot be attached to a singular hypothesis. According to this interpretation, the statement "The probability of the next cast with this die being six equals 1/6" is not really an assertion about the next cast; rather it is an assertion about a whole class of casts of which the next cast is merely an element". It can be compared with the statement "the probability of the next patient having a flu according to the presented symptoms", which can only be the assertion about a class of patients with influenza presenting similar, or supposed similar, symptoms. No one can really predict. Yet it is widely concluded that the calculus of probability can be applied.

But the main reason for introducing the concept of probability is the incompleteness of our knowledge (Waisman, quoted in 8). Acknowledging the incalculability for future singular events, how can we draw conclusions which we can interpret as statements about empirical frequencies, and which we thus find brilliantly corroborated in practice? (8). The answer to this question is that it is just the corroboration (Popper), or confirmation (Carnap, Polanyi), or credibility (Hempel), or reasonableness (Braithwaite), that seemingly justifies our inference. The peculiarity of a probabilistic hypothesis is that it is neither conclusively refutable nor it is convincingly verifiable by any experience. Empirical criteria can be given for rejecting the hypothesis, but the rejection will always be provisional and liable to cancellation.

In (medical) decision making, nevertheless, the calculation of the probability estimates may direct the choice between alternative hypotheses and choosing among a number of possible actions. The choice is to be determined by a comparison of the desirable effects of believing one hypothesis if it is true and the undesirable effects of believing it if it is false with the desirable and undesirable effects of believing the other hypothesis if it is true or false; commonly described as true and false positives and negatives. The criteria for this choice will be in terms of the arithmetical relationships of four values - the gain or loss obtained by choosing one hypothesis if it is true or false respectively, the gain or loss by choosing the other hypothesis if it is true or false respectively. The different assignments of these four values will give different choices (9).

The numerical values come from extrapolation: the calculation of

probabilities which are not given from probabilities which are given. However, this type of calculation only gives rise to subjective interpretations. It treats the degree of probability as a measure of the feelings of certainty or uncertainty, or belief or doubt, which may be aroused in us by certain assertions or conjectures. In connection with some non-numerical statements, the word 'probable' may be quite satisfactorily translated in this way. But an interpretation along these lines does not seem very satisfactory for numerical probability statements (8).

In our opinion, every attempt to employ mathematical statistics to this type of inductive reasoning will fail in the end. Popper (8) expresses his opinion as: "I do not think that it is possible to produce a satisfactory theory of what is traditionally called "induction". On the contrary, I believe that any such theory - whether it uses classical logic or probabilistic logic must for purely logical reasons either lead to an infinite regress, or operate with an aprioristic principle of induction, a synthetic principle which cannot be empirically tested".

The degree of confirmation only expresses the degree of belief, the degree of certainty a physican has in his own inference and conclusions. This belief is, by lack of an objective reference, mainly based upon his personal experience and his personal conceptions about health and disease. His conclusions only resemble the proximity of the singular statement to his personal knowledge and conceptions but tells us nothing about the real situation. There cannot be a plausible reason to treat inductive conclusions (e.g. medical diagnoses) otherwise than preliminary hypotheses which can only be accepted as objective knowledge (and, therefore calculable and teachable) when they are submitted to severe tests. This raises two questions: "does the medical process permit such testing?", and "will the conclusion of a medical process always be a provisional expression of the doctor's belief about a patient?".

We have considered several axiomatic systems, all of which pretend to give an interpretation of the probability concept. We have seen that the frequency approach is only suitable for applications when large numbers of observations are available. Carnap's approach has some features of the frequency appraoch if large numbers of observations are available, while at the same time it tries to allow for treatment of statements about unique events. The main objections to this theory are that all relevant experience must be made explicit (intuitive weighing of different kinds of experience is not permitted); and the irrefutability and unverifiability of the probabilistic hypothesis.

If one wants to introduce intuitive weighing into one's analysis, one must use the subjective, or personalistic, probability models (e.g. Savage; Pratt,Raiffa,Schlaifer). Such models are characterized by a number of consistency requirements. It is particularly in this respect (consistency) that medical nomenclature and definitions often cannot fulfil these requirements (see also chapter II,paragraph 3). Moreover, we cannot expect that our intuition will automatically obey such requirements, but one of the advantages of the model might be that it may detect and abandon any inconsistencies in our original intuitive feelings (19).

For repetitive events it is quite possible that the observable relative frequency will be completely different from our intuition or estimation of chances. Thus it is necessary to extent the subjective probability model by including the possibility of learning by experience. This can be done in several ways. One could redefine the probabilities after each observation, in a purely personalistic way. This is, however, quite unsatisfactory; one would like to have some consistency in the changes of subjective probabilities. This implies that we need a methodology that tells us how to learn from experience (19). Carnap's methodology can, unfortunately, only be applied in a rather limited (and mostly artificial) context.

In the world of the subjective probability, it is quite common to introduce prior distributions and to learn from experience by means of Bayes' theorem. It is not, however, a logical consequence of the postulates that the process of

learning from experience should take place in this way. (The question whether people do learn from experience is still a matter of debate.) This essential part of the theory is left in the open.

Subjective probabilities are always used within the framework of a model. It is, however, quite possible that the facts will not justify the choice of the model. It has proved necessary e.g. to assume uncertainty superimposed on the models, either because models are too simplified to reflect the complex reality or because of random factors like human free will.

If only a limited amount of information is available, as is often the case in particular medical situations, it is somewhat impossible to develop theories that are objectively true. In practice, this difficulty means that models in medicine have usually a rather weak background. Advanced statistical methods, sometimes used for various models in medicine, result in an accuracy that is misleading if it is considered against the background of the possible absence of a well-founded theory for such models. This causes an extra inference problem because a suitable theory of inference and reasoning should specify at what moment the model should be rejected.

REFERENCES CHAPTER III

1. Polanyi M.,: Personal knowledge. Towards a post-critical philosophy. London, Routledge & Kegan Paul, 1958.
2. Feinstein A.R.,: An analysis of diagnostic reasoning. II. The strategy of intermediate decisions. Yale J. Bio. Med. 1973, 46, 264 - 283.
3. Johnson-Laird P.N. Wason P.C.,: An introduction to the scientific study of thinking in : Thinking, readings in cognitive science. Cambridge University Press, 1977.
4. Blois M.S.,: Conceptional issues in computer aided diagnosis and the hierarchical nature of medical knowledge. J. Med. Phil., 1983, 8, 29-50.
5. Harre R.,: The philosophies of science. An introductory survey. Oxford University Press. London, Oxford, New York, 1972.
6. Medawar P.,: Scientific method in science and medicine. Persp. Biol. Med., 1975, 345 - 352.
7. Goldberg L.R.,: Man versus model of man: A rational, plus some evidence, for a method of improving on clinical inferences. Psych. Bull, 1970, 73, 422 - 432.
8. Popper K.R.,: The logic of scientific discovery. London, Hutchison, 1959.
9. Braithwaite R.B.,: Scientific explanation. A study of the function of theory, probability and law in science. Cambridge University Press, 1968.
10. Elstein A.S., Shulman L.S., Sprafka S.,: Medical Problem Solving. Harvard University Press, Cambridge (Mass.) 1978.
11. Scandellari C., Federspil G.,: La scelta delle ipotesi in clinica. Med. Secoli, 1977, 14, 161 - 181.
12. Betaque D.E., Gorry G.A.,: Automating judgmental decision making for a serious medical problem. Management Science, 1971, 17, 421 - 434.
13. De Vries Robbe,P.F.: Medische Besluitvorming, een aanzet tot formele geneeskunde, Groningen,Dissertation,1978.
14. Johnson-Laird P.N., Wason P.C.,: A theoretical analysis of insight into a reasoning task, in: Thinking, readings in cognitive science. Cambridge, University Press, 1977.
15. Brehmer B., : In one word: not from experience. Acta Psychologica, 1980, 45, 223 - 241.
16. Carnap R.,: Statistical and inductive probability, in: Madden E.H.

(ed.): The structure of scientific thought and introduction to philosophy of science. London, Routledge and Kegan Paul, 269 - 279.

17. Bayes T.,: An essay towards solving a problem in the doctrine of chances. Studies in the history of probability and statistics, IX, 296 - 315.
18. Sadegh-Zadeh K.,: Subjektive Wahrscheinlichkeit und Diagnose. Meth. Inf. Med., 1974, 13, 97 - 102.
19. Leede,E.de, Koerts,J.: A Notion of Probability: a survey. Erasmus University Rotterdam.

CHAPTER IV

THE MEDICAL PROCESS

As we have discussed in Chapter II there are no uniform definitions or general agreements upon the basic elements of medical knowledge. But in practice, physicians can understand each other perfectly on a number of reference points in what we will conveniently call: The medical process which can be described as the total sum of procedures and actions taken by the physician between the entry of the patient in the health care system and his leaving it. This in disregard of all the activities of subsidiary health care workers because of our focus on the physician. Within the medical process a number of characteristics is perfectly recognisable to each physician: data collecting and data processing, making a diagnosis, determining the prognosis, establishing a therapeutic action and evaluating the outcome. These are gross features of a very complicated process of problem solving. Two main elements dominate the medical process: diagnosis and prognosis. That is exactly what the patient is asking : "Doctor, what is it, and what is going to happen with me?"

When the physician is unaware of the underlying process his prediction is vacuum based. The prognosis as based upon the knowledge of the (natural) course of the disease, primarily demands the diagnosis of the disease.

Paragraph 1

DIAGNOSIS

There is no general agreement on the term 'diagnosis'. Originally the word stands for 'distinguish'. In this sense, diagnosis can be conceived as the process of differentiating between a number of alternatives. This includes a preconception of the notion that there are really "things to be chosen". This preconception consists of three components:

a.) a preexisting series of categories or classes to provide the frame for the diagnosis (this matter will be discussed in par. 3.);

b.) the particular entity which is to be diagnosed; and

c.) the deliberate judgment that the patient belongs to this category rather than that (1).

But the physician has not only to differentiate between disease alternatives, but between disease and health as well. Not every patient who enters the health care system is -medically speaking - ill, while not every diseased person is considered to be ill in social life. As we have previously described, health and disease are largely abstract concepts which are time, social and cultural based. The stress between these concepts makes diagnosing, especially in primary health care, a troublesome business. Judgment about the state of health and disease can raise conflicts between the patient's judgment about his well-being and that of the physician, which may largely differ.

Medical knowledge left physicians virtually empty-handed in making (objective) judgments about the question: "ill versus not-ill". Nevertheless, society urges physicians to decide upon these questions in matters of insurance, social institutions, labour, etc. Definitions about both states are left conveniently vague. They do not permit physicians to decide unambiguously upon ill versus not-ill.

We can agree with Gross (2) that only generally accepted definitions designing clearcut descriptions of the states of health or disease justify the

physicians' judgments on the insurance and social domains.However, these definitions and descriptions are wanting.

There exist tremendous differences between medical criteria used by the examining physicians and the relativistic sociocultural standards employed by lay people to identify the presence of illness. This does not mean that those diagnosed as not-ill according to the medical criteria were not suffering from some measure of discomfort, unhappiness, or inefficiency because of their symptoms (3). Brown found that non-sickness accounted for almost 25% of the diagnosis made on 12.835 patients: patients who labelled themselves as being ill. In England Balint estimated that 25 - 50% of all patients who go to doctors are not suffering from any pathological entity or nosological syndrome (quoted in 4). Approx. 60 - 70% of people who experience some discomfort do not go to an official medical institution (5).

The very notion of a disease is so nebulous, that to deal with it in terms of recognition algorithms may just give the discussion a superficial and spurious exactness. Nevertheless, the physician is summoned to differentiate between state of health and disease, and between diseases. This latter aspect mainly designates the term diagnosis.

The process of diagnosis is discussed in literature mainly in two ways: a deterministic way and a probabilistic one. The deterministic approximation can be examplified by two definitions:

1.) "The diagnostic process employed by a physician can be seen as an attempt to establish the similarity of the presenting symptoms and signs of the patient to a particular disease prototype" (6);

2.) "The process of diagnosis is a representation of the problem-solving process wherein the problem-solver (a physician) is required to examine,evaluate and select information in order to reach a goal (a disease specification)(7).

According to the first definition, the process of diagnosis consists of comparisons between the pattern of features which a patient presents with those of described diseases or memorized from personal experience by the physician. The degree of resemblance establishes the degree of correctness of the diagnosis. In this sense, the purpose of diagnosis is to identify a patient's disorder as a classifiable entity.

The second way refers to some kind of strategy in order to acquire and to process data relevant and in accordance with a constellation explanatory to a disease.

The term 'diagnosis' gradually shifted from the process of differentiating to the outcome of the process, the name of a disease. It can be seen as a process of converting observed evidence into names of diseases. The evidence consists of data obtained from examining a patient and substances derived from patients. When each symptom, each sign, each lab test etc. is regarded as a piece of evidence, also called indicant, then a disease can be considered as an ordered set of indicants. Diagnosis becomes the allocation of a set of indicants to a disease class (8).

This conception of more of less fixed indicants constituting more or less fixed disease configurations led to a decision theoretical approach, in which indicants and diseases are related in a calculable probabilistic way.

Card & Good (9) defines diagnosis as:

1.) the activity of deciding which of a number of diseases obtains; or which of a number of treatments is advisable; or of attempting by acquiring evidence to allot probabilities to the diseases in the hope of making one of the probabilities close to 1; or of attempting to determine the expected utilities* of the various possible treatments in the hope that, with probability close to 1, the treatment selected will be the one of maximum expected utility on all the evidence that could be collected";

2.) the disease whose probability is close to 1 after diagnosis has been made in the restricted sense".

* expected utility can conveniently be described as the value which can be attached to the expected quality of a state of health with regard to the respective disease and treatment.

In the probabilistic version of diagnosis one has to make use of numerical values in order to calculate probabilities and utilities. The variability and the lack of consistency in classification and nomenclature explains the great difficulty encountered by the statisticians in modelling and delineating in mathemathecal terms the diagnostic process for in practice they are attempting to model something which does not exist as a single entity (10).

Diagnosis as a procedure is often viewed as a stepwise, sequential process.

Dudley (63) views diagnosis as a two-step process:

1.) the physician collects data, and only when this process is completed,

2.) he reflects upon it for the purpose of inferring a diagnosis (or a set of diagnostic hypotheses = differential diagnosis).

In the discipline of medical decision making the diagnostic process is considered to be a sequential process in which the physician employs a test to obtain more information in order to test the new information, and so on until a 'final hypothesis' or 'diagnosis' is reached. Every step in the process is considered to be quantifiable by means of numerical values.

Actually, these numerical values are largely unknown, not only to medical knowledge but also, and especially to the physician.

A stepwise approach consisting of three phases can be depicted as:

Step 1. : select a diagnostic test (or question)

Step 2. : carry out the selected test and observe its outcome.

Step 3. : either (i) select a further test and so return to step 1 or (ii) make a diagnosis in the light of the outcomes so far obtained (12).

This conception presupposes that there is something to test upon. This 'something' is mainly the preexistent idea (hypothesis) that comes up in the physician's mind after the presentation of the complaint.

Because of variability and uncertainty of disease concepts, and people's -often- lacunar memory , we think that the theoretical stepwise-test approach does not always cover the reality of the diagnostic process in routine practice.

Paragraph 2

DATA ACQUISITION, PROCESSING, AND RECORDING

Patients entering the physician's consulting room present neither symptoms nor data. The former ones present the medical translation of patient's feelings. The latter ones are artificial elements preferably written in a (alpha)-numerical way based upon the symptoms of the patient. But before a certain feeling of a patient can be acknowledged as a symptom or a datum a number of conditions have to be fulfilled.

1) A certain process takes place before the individual is acknowledged as a patient (the 'sick role'). That means that his feelings must be avowed as belonging to a possible disease or disability. This depends on:

a) disruption of normal functioning
b) visibility to others
c) perceived seriousness
d) elicited embarassment
e) expectation of effective treatment
f) incapacitating symptoms.

2) The complaints must convey a medical significance. When the presented complaint cannot be related to a -known - disease-entity of some disturbed state of health, it is most of the time ignored by the physician. There are two possibilities for denying the symptom:

a) the complaint indeed cannot related to any known disease entity e.g. "Doctor, when I blink my right eye, I feel an itch in the little toe of my left foot".

b) the physician is unaware of the existence of the disease to which the symptom can be related. This possibility arises when the disease in question is -extremely- rare (global or to his environment), or there is really a gap in his knowledge memory.

In extension, there are some communications about particular features in physicians' behaviour like:

- he only considers the prominent symptoms as he knew them from past sicknesses (20)
- the physician is too prone to see things in patterns (21).
- he only looks for positive answers. The power of the positive respons is more than 100 times greater than a negative one for a single attribute (22).
- in physical examination, a sign is noted when certain sensory stimuli have passed from the observed entity to the doctor's cerebral recognition. Errors can occur before the stimuli reach the doctor's mind because of sensory defects. Alternatively the doctor's cerebration may still fail to recognize and identify the entities (13).
- most physicians take into account the unreliability of patient's answers, physical signs, laboratory tests, reports etc. (23).

When we consider the pace in which the physician performs his consulting hours, the overload of information and the extreme variability of presented problems (especially for family physicians), then we can become aware of the extreme difficulty of data collection and the construction of a reliable data base.

3) The symptoms and signs presented must be sufficiently deviant from the range before they are acknowledged as such by the physician (13). Problems may

arise because of:

a) a lack of reference values for the symptoms and signs in medicine. Of course, everybody has some intuitive perception of norms, largely related to one's own feelings, but it disregards the variability of the norms as they vary over sociocultural values, age and time. It does not take into account the personal variability of the physician. Different physicians place different emphasis and interpretation on deviations from the normal in similar recordings (14). Striking differences in appreciation between physicians are found when performing estimations of the attributes: "obesity", "thinness", or "normal weight" (15).

b) the difference of presentation of complaints and bodily signs by the patient. Zola (16) describes the different types of presentation of complaints by Italian versus Irish patients and their physicians' reactions. Mechanic (16a) described the social psychologic factors affecting the presentation of bodily complaints.

c) the doctor-patient communication. About this item an extensive literature has been published. E.g. doctor-patient communication in pediatric clinic (17); Talking with the doctor (18); Assessment of Nonverbal communication in the Patient-Physician interview (19). Physician's enquiry techniques, physician's non-listening behaviour, physician's attitude towards patient's emotions, physician's inabilities to meet the patient etc. are all subject of an extensive amount of - most psychological - studies. Because most of these studies focus only on one or two elements of the doctor-patient communication, which in turn is part of the data-collection procedure of the physician, their contributions are fragmentary.

Data Recording

Data processing in medicine is a method by which the data base is transformed into a problem list. It is the important step whereby information gathered about the patient is filtered and the clues are selected and grouped into meaningful problems (24). However, an ideal data base has never been formulated. The medical record is too often accepted without proper criticism (25). Often physicians are blamed for their inappropriate records, but one has to admit that there are few guidelines for physicians. A major attempt to streamline the medical record has been delivered by Weed (26). He stated that the Problem Oriented Medical Record (P.O.M.R.) can help in ameliorating a variety of difficulties now besetting medicine, among which are:

1) medical problems dealt with or without context;
2) inefficiency in practice;
3) lack of continuity in care;
4) the apparent inapplicability of "basic science" facts and principles;
5) inefficiency in education;
6) the absence of meaningful audit in the practice of medicine;
7) increasing complexity;
8) explosion of technologic tests;
9) withdrawal or apathy of the supervisory personal.

All these bottlenecks to a sound medical practice are fully recognisable, but to blame them almost exclusively to the medical record seems to go too far. We agree with Feinstein stating that "the medical record was a reflection rather than a cause of the difficulties; a symptom of the intellectual maladies but not the disease" (27).

Anyway, Weed's proposal was an attempt to set guidelines to a standardization of a medical data base. Unfortunately, this did not much alter the main problem: the reliability of the stored data. While proclaiming the P.O.M.R., Weed forgot to define what problems are. The only difference was the

remodelling of the source-structured recording into a problem-structured one. Doing so, he lifted the record into a kind of vacuum, because people cannot handle problems as singular data, nor is there any explicit or implicit context to process them.

Data collection is inefficient when the information collected does not lead to the formulation of useful hypotheses or to the testing of existing ones and may, therefore, be circumstantial or irrelevant. Besides, data have to be assessed to determine whether the information is true. Next the data have to be evaluated to determine whether they reflect an altered body state or are a normal variant (28).

Reliability of Data

Data can be described as symptoms, signs and tests (laboratory, Xray etc) for which some degree of reliability and / or relevance have been assessed. Although the reliability of a symptom, the precision of the questioning procedure and the accuracy of the truth of the patient's complaint, cannot be assessed in numerical terms (29), the carefulness of the medical procedure is often considered as a qualitative norm. The relevance of the symptom is often seen as its degree of contribution to a particular disease. The diagnostic value of a symptom is disease-conscious and its relevance is likely to be different for different diseases (30). This relevance is often expressed in a weight to each symptom for each disease. The disease which produces the largest ratio of the patient's weighted symptoms to the weighted sum of all characteristics for that disease is considered the correct diagnosis.

The scoring process (the measurement of the weight) uses numerical values which reflect the likelihood with which various clinical findings will occur in a given disorder (7). The values are found with the help of mathematical-statistical procedures (as in medical decision making, see chapter V) or with the help of fuzzy set theory (e.g.20,31,32,33). The mathematical theory starts from the principle that symptoms are the atoms in medicine. When we know the symptom configuration of the disease we only have to weight the symptoms, or rename them to pathognomic or corrollary symptoms, symptoms part of the rule, valued attributes, distinctive features etc., and everything can easily be fitted into the right place, using only a program to find the "goodness of fit".

Regrettably , it is not as simple as that. Every physician knows that the same symptoms can have different meanings not only to different diseases but to the same disease also. Pain in the breast with a younger man can be indicative for a pneumothorax, while in an elderly man it can point to a myocardial infarction. It can make a lot of difference if a symptom comes in an acute, an interrupted or an insidious way, even if it concerns only one and the same disease.

There is some evidence that symptoms are not the often assumed atomic pieces. They can vary according to a number of subsidiary aspects related to the symptom (we shall discuss this matter in greater detail in chapter VI). But to describe these minor and often very subtle details explicitly is hard labour and largely unrewarding. However, the variation can account for large differences and yield disagreement between different physician's data bases, and perhaps even within one data base over time. The suggestion that feeding back the information to the examining physicians which could help to reduce the differences between physicians and improve the results may only scarcely be sustained (34).

Information gathering is the accumulation of a profile of data concerning the patient. There are innumerable "facts" to be gathered and there are many reasons why physicians can go astray or misinterpret the results of their labour. This may especially be true for elements the physician is less acquainted with, e.g. laboratory and X-ray results. Physicians often feel

reassured with laboratory tests they neither need nor really use. Zieve (35) listed 10 items concerning misinterpretation and abuse of laboratory tests.

The present 'habit' to order multiple tests seems not to contribute to a clearer understanding of the problem or to help in diagnosing. They are only helpful when:

a) all test-results are normal thus tending to exclude a disease;
b) all test-results are abnormal thus tending to confirm a disease.

They are least helpful when some are positive and others negative (36).

The interpretation of the result is also influenced by the definition of the cut-off point, the point where the calibration change from normal to abnormal. When this point is directed towards diseased patients, specifity (defining non-patients) increases but sensitivity (defining patients) decreases, and as it is moved in the direction of patients without disease, the reverse is true.

This dilemma of the interpretation of results leads to another peculiar phenomenon in the medical process: the observer error. It is the physician's job to observe. However, as we have discussed before, observation is said to be 'theory-laden': we observe particular things in a situation because we have theories which impute relevance to some of them and not to other (37). There cannot be a purely neutral and indifferent collection of clinical facts.

Each observation is person, time and place based. What really worries in constructing a reliable data base is that these highly personal valued data give rise to several errors in the data processing storage and retrieval. These can be

1) errors of omission : neglect to observe

2) errors of detection : fail to observe

3) errors of interpretation : as discussed above

4) errors in recording : only a limited number of symptoms, perhaps only relevant to the doctor, is recorded. (see also 38).

5) errors of memory recall : inaccurate recall of assumed similarity of patterns

6) errors of retrieval : fail to find recorded symptoms, therapies etc.

7) errors of classification : incorrect classifying of patients into the data system.

The main problems to these errors can be summarized as:

a) what is the magnitude of observer error?
b) how can this error be minimized?
c) what is the significance of the residual variation? and
d) what are the effects of observer variation (39).

Two major drawbacks from observer variation can be detected:

I. the variability of observation and recording by one physician of repeated observations (intra-observer variation), and

II. the disagreement of observation and recording between two or more physicians (inter-observer variation) of the same phenomenon.

In literature these features are described in two ways, dependent on viewpoint: intra/inter-observer errors and intra/inter-observer agreement.

The review of Koran (40,41) discusses intra- and inter-observer agreement as observed in literature. Intra-observer agreement for a number of physical

signs values ranged between approx. 60 and 80 per cent. The inter-observer agreement scored lower, ranging from about 40% to around 80%. It was noticed that the more diagnostic categories there were to consider, and the less severe an abnormality, the lower the inter-observer agreement. From all sources of information history taking is the most important but also the least accurate. Agreement among physicians concerning patient's questionnaire was only 2 in 25 patients (42).

The overall diagnostic error rate of even the most senior clinician has been shown to around 21 per cent (39). According to Gill et al. (39) the error rate in diagnosis due to observer variation forms a small proportion of the total error rate in diagnosis. In fact, somewhat less than one quarter of all diagnostic error observed can be attributed to the acquisition of 'faulty' data by the doctor.

Spitzer & Fleiss (43) found striking unreliability of psychiatric diagnosis. They also report marked differences in diagnosing a particular disease. The percentage of psychiatrists diagnosing schizophrenia ranged from 2 per cent in the British Isles to 69 per cent in the United States. In a similar way, Baumann et al (44) found diagnostic disagreement between psychiatric clinics in Zurich and Berlin.

Strong disagreement of the diagnosis of chronic non specific lung disease (asthmatic bronchitis) between Great Britain and the U.S. was found by Reid at al (45).

Wagner et al (46) collected about 1000 papers on errors in medicine, of which 383 about errors in diagnosis. "Although we know it (making errors), nevertheless, we hold the most striking judgments, data and findings for pure gold, and build with these elements our judgments, albeit in general their range is unknown to us" (47).

Medical records are of the utmost importance to the medical practitioner and his patient, and, when legible, to his colleagues while attending to the practice. A data base is an important tool in building the foundation of medicine or of the specialty to which it concerns. We, physicians, are urged to search and research for guidelines in delineating reliable medical records and medical data bases. Let us hope the statement of Cochrane et al. (48) "Although medical disagreement has long been recognized as common and was published by Alexander Pope (1723), there appears to be some general reluctance to recognize it as due to the subjectivity of medical judgment and even more reluctance to investigate it quantitatively." has become obsolete from now on.

Information Load

The (over)load of information in diagnostic data processing is another hot item in this field. It is one of the main claims for computerizing or computer-assisting the diagnostic process. Better than man, the machine can easily deal with large amounts of data. It is hypothesized that a heavy information load may lead to simplification. Information load is generally conceived as the amount of data to be processed per unit of time. In one of his studies, de Dombal et al (49) showed, in an artificial setting, that accuracy of processing data declined after 8 elements of information load, substantiating Miller's theory on channel capacity (see chapter II). However, the study disregards Miller's suggestion, about the creation of 'chunks' with increasing experience and skill.

"Massing of data" was a phenomenon found by Lichtenstein et al. (50). It enables people to economize their information processing, but on the other hand influences people's estimation of occurrences. In the judges' opinion, occurrences tend to be massed rather than distributed over time. E.g.: People believe air crashes to occur in crops, physicians believe myocardial infarction to happen in outbreaks.

The "massing of data" may be an explanation for differences between laboratory and real life experiments, the former mainly derived of sensational or emotional charges of items. People heavily respond to emotionally charged events, especially when stressed to be catastrophic. They exhibit strong and often consistent biases. Some portion of these errors may be due to the unrepresentative coverage of these events in news media. Zebrowski (quoted in 50) notes that "fear sells": media dwell on potential catastrophes and not on succesful operations. The more is published about some kind of disease, the more the physician tends to diagnose it. Not only the physician but the patient also is influenced by this psychological effect. It effects the presentation of the complaint. Lichtenstein et al (50) found two explanations:

a.) encoding variability: spaced repetitions are more likely to receive differential coding than massed items;

b.) deficient processing of massed items. The studies of Tversky and Kahnemann (51) made clear that people tend to estimate weights of events and their frequency by a number of erroneous heuristics. As estimates play an essential role in medical decision making, this matter will be discussed in chapter V.

Uncertainty in data processing

Uncertainty is another element in data processing and diagnosing. Before a patient enters the consulting room, uncertainty as to the true state of the individual is maximal to the physician. But by means of a number of manoeuvres from the part of the physician and the patient, information is obtained which reduces or attenuates the uncertainty to the point, it is hoped, where the information is prescriptive of a course of action. Several studies have shown the influence of uncertainty on the diagnostic process. Slovic et al. (52) define decision-making as "reducing uncertainty in a problem situation (not to acknowledge and quantify it)". Members of professions such as law and medicine frequently are confronted with uncertainty in the course of their routine duty. In these circumstances informal norms have developed for handling uncertainty, so that paralyzing hesitation is avoided. These norms are based upon assumptions that some types of errors are more to be avoided than others; assumptions so basic that they are usually taken for granted, seldom discussed, and therefore slow to change (53). These errors can be described as:

Type I error, or error of the first kind: judging a sick person well (rejecting a hypothesis which is true), and

Type II error, or error of the second kind: judging a healthy person sick (accepting a hypothesis which is false).

Many medical judgments are based on avoiding type I error, which can be examplified by the proverbs: "when in doubt diagnose illness" and "Better safe than sorry" (53).

Another feature of uncertainty is described by Driscoll & Lanzetta (54). They describe information search as a monotonic function of uncertainty. The amount of information declines under high levels of respons uncertainty. Both stimuli and response uncertainty influence the physician's state of uncertainty. This observation has been underscored by the perceived differences of behaviour between family physicians and (general) internists. Family physicians, operating in more vague task environments, asked fewer items of information, fewer history questions, fewer questions about life situation and mental status, used fewer items of physical examination, and ordered fewer laboraty and related tests (55,56).

Eventually, the medical data processing leads to diagnosis and disease classification. We shall proceed to this latter item.

Paragraph 3

CLASSIFICATION .

"The process of classification, the recognition of similarity and the grouping of organisms and objects dates back to primitive man. He must be able to perceive similarities in stimuli for survival. The study of classification has always had two major interrelated components: "how do we classify?" and "how should we classify?". The first component belongs to the domain of the psychology and philosophy of sense perception: What is similarity, how do we know we recognize similarity, how do we perceive regularity and relationships, what are the criteria, how does classification affect everyday life etc. The second component is the subject matter of taxonomy, the science of classification". (The theoretical background of classification is borrowed from Sokal,57). From these components three characteristics have to be distinguished:

1) there must be something to perceive, to identify;
2) we must be able to place these identified objects into a class of similar objects; and
3) we have to know how these classes have to be interrelated.

Identification will be defined as the allocation or assignment of additional unidentified objects to the correct class, once such classes have been established by prior classification. The definition of classification is the ordering or arrangement of objects into groups or sets on the basis of their relationships. And finally, the definition of taxonomy will be defined as the theoretical study of classification including its basis, principles, procedures and rules.

We can examplify this with an example of botany:take e.g. a flower. One could easily identify it as being cultivated or not,but it needed a genius like Linnaeus to set up a taxonomy of plants in order to give you the possibility to find the exact class and subclasses and name. Finally, a more or less experienced person must be able to find the category and name of the plant.

All classifications aim to achieve economy of memory. The world is full of single cases. By grouping numerous individual objects into a class, a taxon, the description of the taxon subsumes the individual description of the objects contained within it. By grouping the plant into a taxon we know, although perhaps they might not be there yet, that eventually all flowers of this plant (these plants) will become yellow.

Another purpose of classification is manageability The objects are arranged in systems (that may or may not be hierarchic) in which the several taxa can be easily named and related to one another, e.g. the Mendeleiev system for chemical elements. The paramount purpose of classification is to describe the structure and relationships of the constituent objects to one another and to similar objects and to simplify these relationships in such a way that general statements can be made about classes of objects. It is easy to perceive structure when it is obvious and discontinous. But much of what we observe in nature changes continuously in one or another characteristic with equally steep gradients for each characteristic.

Classification that describes relationships among objects in nature should generate hypotheses. In fact, the principal scientific justification for establishing classifications is that they are heuristic and that they lead to the stating of a hypothesis that can be tested. (Heuristic is meant here as a force stimulating interest as a means of further investigation).

Two kinds of classification can be distinguished:

a) Monothetic: Classes differ by at least one property which is uniform among the members of the class;

b) Polythetic: The group or sets of individuals or objects that share a large proportion of their properties but do not necessarily agree in any one property.No single propperty is required for the definition of a given group nor

will any combination of characteristics necessary define it.

Medical classification obviously falls within the latter distinction. In the polythetic classification there is a strong need of weighting characters, especially the ones that highlight the class to which they belong. But how to weigh? Should certain characters be weighted more heavily than others? Who determines the importance of the characters to be weighted? It is important to note that most classificatory labour is not based on fundamental scientific principles but largely on considerations of practicality. This practicality often leads researchers to weight discordant characters less than others. The difficulty with such weightings is that one needs initial classification to provide weights for the characters. But once classifications are established there is a reluctance to change, especially when the classification is assumed to be correct (57).

In medicine a firm basis for classification was laid by pioneers like Morgagni and Virchow, using gross morbid anatomy and microscopy for explaining the clinical manifestation of a great number of diseases. From the second half of the 19th century on the diagnostic process was directed towards the name of a disease, explaining all/many of the phenomena observed. Regrettably, the history of medicine almost always deals with 'ideal types' which are approximated, by the majority of disease entities, and ultimately defined explicitly by their joint pathology and etiology. It may explain why medicine is more interested in the "presence of a disease" rather than the presence of a "diseased state". This presence of a disease encompasses the assumption of the presence of causal factors that are likely to produce illness (58).

In medicine, the classification must take into account both the history of disease in nature and the history of disease in persons, universal and singular phenomena. Therefore, the capacity to form a unifying theory is severely restricted (59).

The human body changed relatively little over millennia, but since medicine expanded its purview to include the concerns of mental health and illness, and since medicine in general must subserve, the deterministic ideology and ulterior goals of given societies, the actual conception of diseases cannot but reflect the state of the technology, the social expectations, the division of labour, and the environmental condition of those populations (58).

"Disease is a normative concept, indicating a state of affairs as undesirable and to overcome, dealing with lesions, organs, functions etc, while the patient speaks about illness, suffering, incapacitation etc". This is why physicians customarily talk about diseases as if they exist as real entities.

But the names of diseases, according to Scriven (60), must be accepted as no more than a verbal shorthand by which to refer to as the conclusions of the diagnostic process. And this shorthand is highly personal. Differences between two or more clinics, or between two or more countries, have been studied regularly (44,61).

Difficulties arise because of ill-defining. But it also includes an ill-defined classification, descending from the morbid anatomy classification of the pathologists of the former century. The physician observes, inferes, deduces, induces etc of what is presented to him as the patient's problem, and translates these observations into symptoms and signs, more or less causally related, to a proposed disease. The pathologist classifies what he sees. But what makes that what the pathologist observes classifiable as disease? Both the physician and the pathologist provide a basis for claiming that physicians observes symptoms and signs.

Criteria for the inclusion of symptomclusters and diseases into classification system can be formulated as:

1) stability, or reliability across time, for the same judge and the same data;

2) consensus, or reliability across judges, for the same data and the same occasion;

3) convergence, or reliability across the data sources, administered on the same occasion and interpreted by the same judges (62). The majority of

reliability studies have focused upon judgmental consensus and have come to widely disparate conclusions: from extremely high to virtually no consensus.

Besides these conditions there remains the demand for unequivocal terminology. Much can be improved by agreeing on the same terminology, techniques, rules, criteria, elements, classification etc. The assumption that accurate diagnosis will be reached despite disagreement regarding particular signs and symptoms remained to be proved (40).

The art of diagnosis consists of classifying patients into a group corresponding to some pathological entity. If we know everything about the patient, provided we can use a standardized terminology, then this classification could correspond to a logical decision. But, of course, this is an idealised situation and we must be able to classify a patient with only partial information about his medical status (40).

But first of all the definition of identification has to be fulfilled, i.e. the establishing classes by prior classification. The need for such a classification (of diseases) was first formulated in 1893 by the International Statistical Institute. It originated an International List of Causes of Death, I.C.D. This classification was revised every decade. The sixth revision of 1948 enlarged its usefulness for morbidity applications by increasing the specificity of rubrics and by emphasizing manifestations of disease rather than etiology. For family medicine the need of a standardized method of recording diagnoses was formulated by Abercrombie in the early fifties. But it was only in 1963 that a practical classification for primary care was edited, the so-called E-book after Eimerl & Laidlaw (63).

In 1974 the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (W.O.N.C.A.) initiated and adopted a new classification and coding system, the international classification of health problems in primary care (ICHPPC) which was updated in 1979 and related to the ICD classification, 9th edition. The difference between the ICD9 and the ICHPPC-2 classification is in the various levels of sophistication. In the ICHPPC system it is recommended to enter each problem at the highest level of diagnostic refinement the user can be confident of at the time. Thus the interpretation is left to the doctor. That means:

1) a subjective interpretation of the observed symptoms and signs;(when observed);

2) a subjective judgment about the aggregated symptoms and signs ("diagnosis").

When the diagnosis depends on many variables including training and experience, habits of thought, acquired skills and the availability of diagnostic technology then we must ask: what is diagnosis / diagnostic refinement in practice? Can diagnoses of physicians be compared to each other (condition of consensus)? Can this type of classification provide the reproducible data necessary for building a knowledge base of medicine?

Meanwhile, various types of these classifications are in use.

Next to the ICD classification a new kind of classification was originated, Reasons for Encounter, based on a taxonomy of problems. However, this poses a new difficulty caused by ambiguity and imprecision in the concept of a "problem". In replacing the diagnostic nomenclature of disease by a pragmatic nomenclature of problems, it has exchanged a standarized but inadequate taxonomy for a, perhaps, less inadequate but unstandardized taxonomy (27). Again, like Weed's book, the word 'problem' is not actually defined whatsoever. But when we do not know whether an individual phenomenon is a separate problem or collectively forms a single problem or a collectivity of problems we are unable to register it, or in a reverse way we shall ignore the certain phenomenon. Moreover, when we do not know which patient is diseased or not, how can we definitely define its classification? Stating that X is a patient, Tautu & Wagner (64), designed the scheme (Di being an abbreviation of disease).

Schema 1

	X = really a Di	X = really a non-Di
Assert X = Di	a correct classification	c misclassification
Assert X = nonDi	b misclassification	d correct classification

square b misclassification stands for type I error;
square c misclassification stands for type II error.

The proportion of correct diagnostic bids is known as the Non Error Rate.

This proportion is unknown, and varies among all kinds of diseases, physicians, cultures, societies etc. Because of natural variation in disease manifestation, the original disease entity can shift into another one. If we really want a classification system, and we have to have one because of its fundamental significance for medical knowledge, it has to transcend the variation and physician's disagreement. Unless we are able to (re) build a reliable system, we cannot really gain insight in what is the paramount feature of the medical process, the prognosis, the prediction.

Paragraph 4.

PREDICTION

"Doctor, what is going to happen with me?" is the essential question with which the patient addresses his physician. Most of the time he asks it differently but the essence remains. He is not really interested in the name of the disease otherwise than the distinction between catastrophic (e.g. cancer, heart-disease) and non-catastrophic (e.g. influenza, superficial infection) diseases. Besides, medicine has invented, and still invents lists of - especially to patients - meaningless names, or adjectives like rheumatic, essential, idiosyncratic etc., which really mean nothing. Neither is the patient interested in the names of the medical treatment apart from some global indications as drugs, operation or physiotherapy. The pharmaceutical industry provides physicians and drugstores with an innumerable variety of drugs under fancy names, differing from country to country, covering only a couple of thousand chemical compounds. The patient, and who is not, is interested in what lies immediately ahead of him. Unfortunately, predicting is very difficult, especially when it concerns the future. The physician, although it is often assumed otherwise, is not an exception to the rule. He has the advantage of the layman that he has some knowledge about the natural history of the disease just diagnosed in the patient. Supposing the diagnosis and the classification is correct, the physician has the knowledge or can find in textbooks some general statements about the course of the disease. Principally, three types of actions can be distinguished:

1) in case of a self-limiting disease, no action has to be undertaken to change the course.

2) the disease has a more serious character, and therapy is - easily - available: the physician will not hesitate to influence the course.

3) the disease is fatal, no causal therapeutic action is really known. The physician will try and prescribe all kinds of actions known to mitigate and

alleviate the symptoms and possible incapacitating outcome of the disease. It presumes a triad of conditions:

a) the diagnosis is correct

b) the natural course of the disease is known

c) the actions and effects of the proposed therapeutic management are known.

The qualities of the first condition have been discussed.

With regard to the second condition the answer can be very simple: about most diseases the natural history is unknown. What we know about the natural history of diseases is descended from medical knowledge of about fifty to a hundred years back. Nowadays practically every patient is 'treated' with all means medicine can provide. We are only able to observe a therapy-induced course of a disease, superimposed by varying reactions to the therapy.

This evokes the dilemma of the third condition: when we introduce some therapy e.g. a particular drug, into a patient, whose diseased state has an unknown course, how can we possibly know what the (re)actions of this drug are? Without the possibility to make accurate predictions how can we possibly know that we shall reach what we want to reach? Or in the famous words of Mager: "If you don't know where to go, you may very well end up somewhere else, and not even know it".

The most fundamental characteristic in the medical action is its selective nature; doing one thing precludes doing another. It is selective in two ways: selective to the nature of the therapy, and selective to the receiver of the therapy. One can only study the reactions of one drug in one patient, excluding other and/or similar drugs. And our actions are selective to the receiver in that while giving some kind of therapy it precludes the not-giving of it.

On the part of the receivers other problems are posed. What are the selection criteria for the particular drug? Is the physician especially interested in this disease? Which patients with the disease will be selected: the most ill ones or not, men or women, old or young etc. Or is the physician especially interested in the drug? And in the latter case, will criteria be different from the target population in the former domain of interest? When the physician chooses the most ill patients he almost inevitably will find a change in their condition. While improving, there may be a very likely alternative explanation to this finding: he observed simply a regression effect, because of his higher or lower judgments between the first and the second encounter.

To say anything about the validity of the judgments, it is necessary to disentangle the effects of the treatment. This can only be done by random distribution. But this is not likely to happen while the resources mostly are scarce and the randomized sample is not always randomly distributed. Action is always selective, meaning that as we select certain cases for treatment, we by that very act also select other cases who do not get treatment. The only cases we can observe, therefore, are the true positives and the false positives. Subjects tend to focus only on the number of true positives i.e. they follow the strategy of using only confirming evidence as we have observed earlier (65). The second and third condition about predictions can scarcely be fulfilled.

Nevertheless, prediction is medicine's almost exclusive and powerful instrument for measuring the effects of treatments and the performance of physicians. With regard to the former question, the necessity to establish criteria becomes more and more urgent. In the time of Alexander Fleming things were easy. The physicians and the whole medical world had a clearcut notion of the natural histories of the known infectious diseases. The introduction of penicillin could be easily observed. Nowadays we can only rely on global descriptions, maybe obsolete, or on minor details in the symptom patterns of diseases or their inanimate data from laboratories.

It becomes extremely difficult, both for medicine and pharmacy, to develop and, moreover, to test new drugs and therapies in order to judge them as a substantial addition and improvement to the pharmaceutical and medical arsenal.

The performance of the physician may be measured by the accuracy of his clinical judgment, aimed at the prediction of significant outcomes in the life

of another individual. When the same type of prediction is made repeatedly by the same judge, using the same type of information as a basis for his judgments, then the process becomes amenable to scientific study (62). But, alas, we are far removed as yet from this goal. Meanwhile, medical audits have the status of exchanges of views, not more, not less. Unless we realize this primitive status of the testing of performances, we shall remain like the medieval scholars, preferring discussion to scientific experimentation, which will eventually lead to a convenient standstill.

Paragraph 5

THERAPEUTIC MANAGEMENT

When one takes a glance at the contents of most medical journals, the number of papers about all kinds of treatments is impressive. Some journals are exclusively dedicated to -mostly- drug therapy, partly because they are sponsored by the pharmaceutical industry, partly because they focus on one of the physician's main interests. One can wonder why so much attention is paid to a subject that is, as I argued before, of no great importance to the medical process. Drugs and therapies come and go, the latest hardly any better than the previous (ones).

Of course, it is essential to keep in touch with developments in the medical and pharmaceutical world. But real innovations in pharmaceutics are rare. After the discovery of penicillin in the late thirties, the principle of antibiotics has not changed. The first benzodiazepines have been introduced in mental health care in the beginning of the sixties, and afterwards hardly anything has changed with respect to the basic compound, except for an overwhelming variety of derivations, all with similar qualities as the basic substance. Many drugs have existed for decades without any change. One can hardly observe any need for the abundance of articles, reviews, survey, books etc on treatment. Nevertheless, medical congresses are predominantly dedicated to therapy. Sessions of physicians are almost exclusively dominated by one subject: treatment. So there must be an explanation for this remarkable phenomenon in the medical world. We can think of two possible explanations:

1) Often, and especially in primary care, the physician is confronted with an unclassifiable complex of complaints. No diagnosis or only a rudimentary one can be made, while the patient is in trouble, or supposed to be. The patient looks expectingly at the doctor, who is left empty-handed by medical knowledge. The only thing which remains is to act like the ancient Greek doctors and prescribe drugs for the various symptoms, without possibly knowing what the (re)actions will be. Every communication with a colleague in a similar, or assumed similar, situation is to be welcomed. Every hint in literature, even the remotest, may be of help in forthcoming acute or chronic but unexplained cases. For diseases, for which cure is unavailable or unknown, various types of drugs are prescribed. Feinstein (66) states: "every treatment is an experiment", but we have to realize that it is not a scientific experiment.

The drug is seen as a tool in the hands of the doctor; an tool of extraordinarily powerful tool without which no health can be gained, no real help can be offered. "Are you quite sure, doctor, that there is no other drug for my complaints?" Instead of resolutely denying, the physician prescribes another drug which he has vaguely heard of. He is part of the belief in health and the health bringing medicaments as they are loudly advocated in the news media.

2) the physician makes a straightforward diagnosis. He can choose among many widely available and powerful drugs. Within a couple of days the patient will be cured or at least restored from the most incapacitating symptoms. But in the case the illness remains or even deterioriates, there may be a number of possibilities:

a) it is a variant of the natural history of the disease;

b) the disease, or cause of the disease, is resistant to the drug;

c) the patient is not compliant with the advised regimen: e.g. he does not takes the tablets as prescribed;

d) the physician made an inaccurate diagnosis.

Approximately in this order the physician considers these alternatives. Although the latter alternative may be the most probable one, it often does not enter the physician's consideration, because the physician's personal judgment and commitment about this case (and diagnosis) hamper an objective and renewed view. Every physician is acquainted with a great number of variants. Needless to say that these variants, if existing, are unmentioned in the textbooks. Every physician feels a need to compare his variant and its treatment with those of colleagues, verbally or from literature. Resistance to particular drugs is often mentioned in a somewhat vague and general way (it has to be distinguished from specific items like the resistance of a particular bacteria to a specific antibiotic). It is often blamed to a number of disturbances like: insufficient resorption, insensibility to the drug, pharmacological processes etc. In these cases a similar drug is prescribed, with varying results. To be in touch with all possibilities of 'resistance' the physician consults other experts and experienced colleagues, and of course various sources of information.

Literature about non-compliance is growing. The importance of this problem is unknown. Its discussion is beyond the scope of this paragraph.

Unmentioned, thusfar, is acting in emergency. Sometimes the physician has to initiate a treatment before a diagnosis is made and approved. Fortunately, these emergency cases are relatively rare in practice and do not require an extensive knowledge of pharmaceutical compounds. In relevant cases only a small number of drugs is used. Janis & Mann (67) listed some conditions to this emergency decision making:

1) awareness of serious risks if no protective action is taken;

2) awareness of serious risks if any of the salient protective actions is taken;

3) moderate high degree of hope that a search for information and advice will lead to a better solution;

4) belief there is sufficient time to search.

Each physician has to be aware of his actions. The powerful drugs, extensive surgery and other interventions make the consequences in contemporary medicine so far-reaching, that no one can afford to stay in the wings.

Paragraph 6

STRATEGIES

The original aim of this study has been the question, "Do recognisable strategies exist according to which a (family) physician reaches a decision?". The answer to this question cannot be disentangled from the contents of the medical process. We now face the question whether there are strategies which can be generalized. It is claimed that, because of the uniqueness of the patient-physician relation, every physician has not only his particular way of working, but he also has a manifold of strategies at his disposal, for each problem a different one. Leaper et al (10) denied a specific diagnostic process in physicians' problem solving. But the remarkable phenomenon exists that

physicians all over the world, providing they are using 'Western Medicine', recognize each others methods, schemes and problems. It reaches beyond the borders set by cultures, societies or health care systems. What is this peculiar thing that connects physicians all over the world?. What is the characteristic (are the characteristics) that makes physicians recognisable to each other?

Of course, one of the elements is the patient presenting his problem.But this cannot provide all explanations. We suggest that (a) joint strategy(ies) is a major element in connecting physicians all over the world. Literature on this subject is scarce. The first signs of interest can be traced back to the fifties, when investigators, mainly working for departments of education, tried to restructure medical examinations. Rimoldi (68) tried to simulate the medical process by means of written information and instruction. (an survey of this subject will be given in chapter VI).

In 1959 Ledley, a mathematician, and Lusted, a radiologist, wrote their fundamental paper on "Reasoning foundations of medical diagnosis" (69). They analyzed the medical process and structured it into a logical and probabilistic system, following the line of Meehl (70). They introduced Decision Theory into the medical discipline, which enabled them to calculate the "predictive value", the probability that the patient has a particular disease given the presented symptoms. Using a series of notched cards and a rod these cards could be sorted in order to provide calculable probability estimates. This device can be seen as a predecessor to the Hollerith machine and the computer.

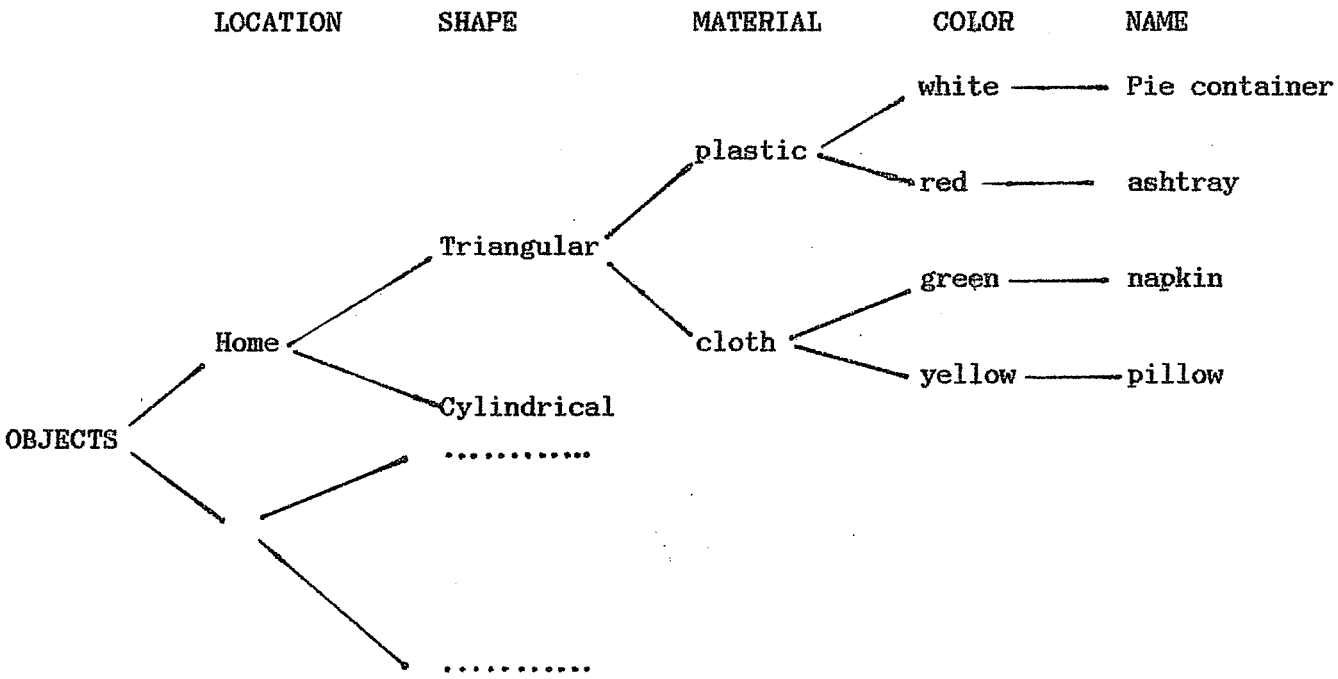
Soon after this intriguing study the first attempts to mathematize the diagnostic process were published (e.g. Warner et al, 1961 (71), Nash, 1963 (72),and in 1967 one of the first surveys about the mathematical - statistical (Meehl originally called it : actuarial) foundations in diagnostics was published (73). As a consequence, investigations in medical decision making strategies and schemes were largely modelled according to the probabilistic and hierarchical system. Several authors express their belief that physicians actually do work in this way. "It is assumed that expert clinicians may use logical methods in their everyday work" (21); "a linear sequential process of hierarchical steps" (74); the primary aim (..) is to express the activities of the clinician in mathematically recognisable terms" (75); "Ideally, a physician needs to calculate and compare the probabilities of various diseases that could have caused the patient's problems (76).

Next to this trail two other methods were developed, both originated in cognitive psychology. In a non-chronological order these trails can be described as 1) multiple branching method, and 2) artificial intelligence.

The first one goes back to the pre-war Gestalt psychologists, Duncker (on problem solving) and Wertheimer (Productive thinking), 1945. The theory of concept formation was further developed in the fifties by Bruner, Goodnow and Austin (77), Miller (78) and Hovland (79), leading to experiments with psychologists ((Mandler (80), Kleinmuntz (81)) and clinicians (Wortman (82,83). The latter author tested Mandler's theory about the organization of information in memory into hierarchical structures. Wortman found three stages of a recoding process: a) the perception of the hierarchy of categories. b) chunking within a subordinate category, and c) the establishment of the hierarchical relationship between superordinate and subordinate categories. This led to a multiple branching scheme of organizational hierarchy like the following abbreviated scheme of Wortman (83)

Figure 1

MULTIPLE BRANCHING SCHEME OF ORGANIZATIONAL HIERARCHY



Multiple branching can be seen as a logical progression down one of the many possible paths in which the response to each test or question automatically determines the next step. This scheme is more widely known as a decision-tree giving each step the weight of a probability estimation towards a previously determined outcome.

Artificial intelligence, although it seems a brand new branch of science, goes back to the beginning of this century. The experiments of Wundt gave a fresh impulse to epistemology. Especially the philosophical theories of Selz (Ueber die Gesetze des Geordneten Denkverlaufs : About the laws of regulated thinking processes, 1913, and "Die Gesetze der Produktiven und Reproduktiven Geistestaetigkeit: The laws of productive and reproductive mental activities, 1924), influenced investigators to more practical theories as Cybernetics (Norbert Wiener, 1948) . Cybernetics was a general name encompassing three systems:

1) information theory,
2) theory of feedback systems (servomechanism theory), and
3) the system of electronic computing. When we know what is going on in the human mind, we must be able to mimic this process. It gave rise to investigations with observation and thinking aloud protocols of human problem solving behaviour. The thesis was, that an explanation of an observed behaviour of the organism is provided by a program of primitive information processes that generates this behaviour. One of the first studies was performed by A.D. de Groot (84), studying the problem solving behaviour of chessplayers. He laid one of the foundations for what is nowadays known as artificial intelligence. Especially the thorough and highquality investigations by Simon and Newell, stretching over three decades, must be mentioned in this particular context. Several diagnostic computerprograms using theories and techniques of artificial intelligence science have been developed. There are three main reasons for this

development:

1) the rapidly increasing storage capacity;

2) the sheer quantity of medical research and knowledge;

3) the greater volume of data physicians have to handle (85).

The most ambitious program now is Caduceus containing diagnostic elements to diagnose more than 500 diseases (86). Whether these three trails mimic the actual daily work of the physician is an open question. Neither of them have been developed in the original setting of medicine. They all share a number of assumptions about the medical process among which are:

1) all diseases are unequivocally classifiable entities;

2) all hypotheses, as intermediaries towards a diagnosis, fit in a hierarchical structure;

3) all tests to hypotheses can be expressed into probabilities;

4) symptoms can be -invariantly- assigned to diseases in terms of weights (mathematical-statistical) or grades of membership (fuzzy set theory, Boolean algebra).

5) each outcome of a diagnostic process can be expressed in terms of value estimation.

6) to each value estimation a probability can be assigned.

7) because we know (or at least ought to know) the frequency distribution of a disease in a (or the physician's practice) population, we can calculate its prevalence.

These concepts are elements necessary towards the organization of a formalized decision-making system. Newell & Simon (87) claim that the benefits to be gained from the computer or information processing approach are:

a.) it provides a formal language for dealing with complex problems;

b.) requires a precise formulation of the theory;

c.) allows for a direct and unequivocal test of the theory by running a program.

The outcome of the program can be compared with the outcome of the physician's problem-solving activities in similar cases. The advantage of the information processing approach is its possibility to retrack the various steps in the process. As the problem solving activities are supposed to be directed by the problem, the problem solving space reflects the organizational structure of memory in information processing as well as in human problem-solving. In other words, an information processing approach can simulate human behaviour problem solving. It will lead to the formulation of a number of conditions for a problem solving program or simulation.

1.) It should predict the performance of a problem-solver handling specific tasks;

2.) it should explain how human problem solving takes place;

3.) what processes are used;

4.) what mechanisms perform these processes;

5.) it should predict the incidental phenomena that accompany the problem solving process (88).

These conditions are not always met in simulations and investigations in medical informatics. In our opinion the omission alienated the physicians from this formalized approach in medical decision making. The information processing programmes became less and less recognisable to the practising physician, which resulted in desillusion and lack of interest. On the other hand, information scientists became annoyed by the physician's behaviour. Their straightforward and clear (in their opinion) programs and algorithms were not accepted by the physicians, indeed, they were rejected although proven superior in some fields.

In our opinion, three obstacles towards amalgamation of the formalized and the 'informal' ways of decision making can be distinguished.

1.) The strategies used in both procedures differ too much.

2.) physicians know, or at least intuitively feel, the relativity and

unreliability of the medical data necessary in the information processing approach to calculate the various probabilities.

3.) the restriction of formal decision making to a number of sometimes less interesting (to the physician) fields. Indeed, the restriction itself hampers interest.

The strategies were supposed to happen according to a hierarchical structure: a strict relationship between superordinate and subordinate categories, as we have described in chapter III, paragraph 3. The physician moves in a top-to-bottom direction, starting at the lowest possible level (82). E.g. the physician begins to choose between the option somatic versus mental illness, and via a number of down-grading subordinate hypotheses comes to a specific diagnosis, respectively decision. The test to the former hypotheses automatically determines the following step. However, except for a small number of simulations or theories (22,89,90) a hierarchical structure has never been ascertained in routine medical practice. In our investigation we challenge this formal, deductive strategy.

The structure can be delineated as a sequence of hypotheses and tests, the first hypothesis reflecting the prevalence of the disease in a - circumscribed - population: the prior probability. Each test and test result can change this probability estimate towards a final estimation: the posterior probability. Each test to a hypothesis can have four outcomes: the predictions that the disease is present or absent given the test result is positive or negative. It can be schematized as follows:

Table 1

BIVARIATE PROBABILITIES OF DISEASE STATE AND TEST RESULT
(frequencies)

DISEASE / TEST	PRESENT	ABSENT
POSITIVE	a	b
NEGATIVE	c	d

The probability that the test or procedure result will be positive when the disease is present, is called Sensitivity and is calculated as:

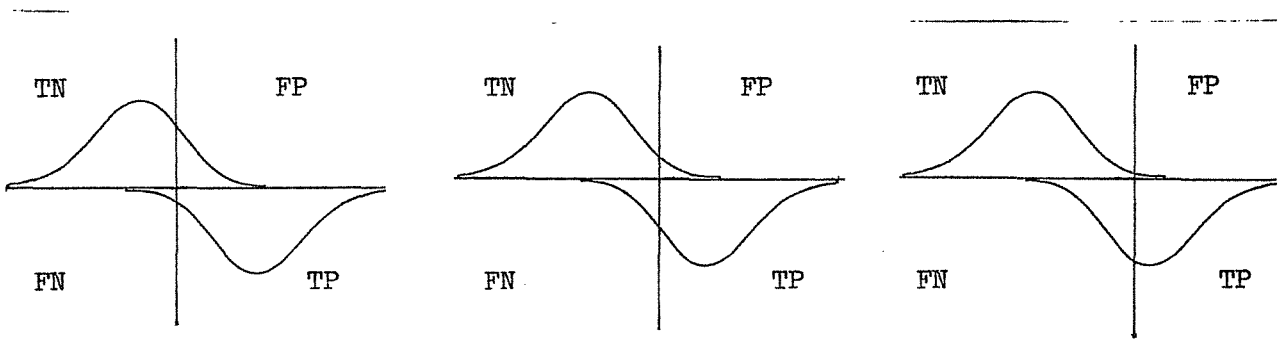
$$\frac{a}{a+c}$$

The probability that the test or procedure result will be negative when the disease is not present, is called Specificity, and is calculated as

$$\frac{d}{b+d}$$

Assuming that the test is meticulously and perfectly executed, the results are influenced by several variances of natural and non-natural origins: regular human variance, drug-induced variances etc. One of the major obstacles in judging test results is the establishment of the margin(s) between 'normal' and 'abnormal'. Several test results e.g. vary with age or sex, sometimes vary from culture to culture, race to race, hour to hour, day to night. Generally, the line between 'normal' and 'abnormal' is drawn on basis of the mean values for that particular test found within a sample of 'normal healthy' adults. This value, the cutoff-point, should mark the difference between a diseased and a nondiseased state. However, this cutoff-point is seldom based on firm, i.e. scientific, foundations. The cut-off point determines to a certain extent, the sensitivity and specificity. The more the point is shoven to the "normal" side, the more people are included in the diseased group and the more excluded from non-diseased group. The test influences not only the sensitivity and specificity values, they diverse in their directions: the higher the sensitivity the lower the specificity and vice versa.

The next schemes may elucidate these relationships (see Wulff (29)). It is assumed that the results of a particular test follow a so-called normal or Gaussian distribution ranging from absolute abnormal to absolute normal. The upper curves in the schemes present the negative test results (specificity) and the lower the positive test results (sensitivity). On the left side of the vertical line the negative or 'normal' values (true and false negative) are depicted, on the right hand side the positive or 'abnormal' values (true and false positives). Moving the vertical bar in either direction clearly influences the values in all four quadrants.



after Wulff (29)

The predictive value of the test in case the disease may be present can be calculated as:

$$\frac{a}{a+b}$$

The prevalence being the percentage of the numbers of diseased people proportional to the total population:

$$\frac{a+c}{a+b+c+d}$$

With the help of some formulae the probability that a disease is present or absent given the symptoms being present or absent can be calculated. This matter will be discussed in chapter V.

Although the theory is tempting, one can scarcely see how the conditions towards the calculations can be fulfilled. Even if we unequivocally and unanimously agree on the values for a, b, c, and d, the basic questions about the medical framework remain. It may may take a very, very long time to obtain that kind of figures which enable us to calculate the prediction. Meanwhile we can underline Feinstein's statement (91) that "iatromathematical enthousiasts could make substantial contributions to clinical medicine if the efforts now being expended on Bayesian and decision-analytic fantasies were directed to the major challenges of algorithmically dissecting clinical judgment, based on the way the judgments are actually performed".

We have to acknowledge that medical decision making actually contributed to the understanding of strategies used in medical problem solving.

Several investigators formulated one or more strategies. In a concise version we shall give some short descriptions.

1.) Elstein and co-workers (90) found in a simulated environment one strategy practiced by all participants, the hypothetico-deductive one: on the basis of the earliest clues a short list of potential diagnoses is formulated, successively tested until a final hypothesis, the diagnosis, remains.

2.) Sackett (quoted in 92) analyzing the literature, found 4 strategies.

a.) <u>exhaustive method</u> : noting every detail of the patient's history;

b.) <u>pattern recognition</u> : conformation to a previously learned model, triggered by sensory stimuli;

c.) <u>Multiple branching</u> : logical progression down one of many possible paths in which the response to each test or question automatically determines the next step;

d.) <u>hypothetico-deductive method</u>

3. Simon & Newell (88) mention two strategies:

a.) <u>Scan & Search mode</u> : a small number of nodes on a decision tree (multiple branching) are searched over a short distance: new nodes are generated when searching fails;

b.) <u>Progressive Deepening Strategy</u> : the top-to-bottom deductive way of reasoning.

4.) Kleinmuntz & Kleinmuntz (93) tested three strategies:

a.) <u>Generate-and-Test Strategy</u> , which picks treatment randomly, until an acceptable one is found: there is no attempt to diagnose the disease;

b.) <u>Heuristic Strategy</u> : this strategy views people as satisfiers who indulge in a limited amount of search, until a satisfactory rather than optimal solution is reached. Only positive evidence is sought to confirm a hypothesis. Their concept is supported by the study of Balla (21).

c.) <u>Bayesian Expected Utility Strategy</u> : this strategy is based on probabilistic reasoning using Bayes' theorem for calculation.

With the latter strategy, we are back with the formalized decision-making strategies. Its achievements and obstacles shall be discussed in the next chapter.

REFERENCES.

1. King L.S.,: What is a diagnosis? JAMA, 1967, 202, 154 -157.
2. Gross R., : Was ist eine Krankheit? Gibt es stabile Krankheitsbilder? Metamed 1, 1977, 115 - 122.
3. Kellert, S.R.,: A sociocultural concept of Health and Illness. J. Med. Philos.,: 1976, 1, 222 - 228.
4. Drucker E.,: Hidden values and health care. Med. Care, 1974, 12, 266 - 273.
5. Mechanic D.:Medical Sociology. New York,Free Press,1978.
6. Gorry G.A., Modeling the diagnostic process. J. Med. Educ. 1974, 45, 293 - 302.
7. Wortman, P.M.: Medical Diagnosis: an information processing approach. Comp.Biomed. Research., 1972, 5, 315 - 328.
8. Card W.I.,: The Diagnostic Process. J. Roy, Coll. Physcns., London, 1970, 4, 183 - 187.
9. Card W.I., Good I.J.,: A logical analysis in medicine. A companion to medical studies. Blackwell Sc. Pub., 1976.
10. Leaper D.J., Gill P.W., Staniland J.R., Horrocks J.C., Dombal F.T. de, Clinical Diagnostic process; an analysis. Br. Med. J. 1973, 3, 569 - 574.
11. Dudley H.A.F., Pay-off, heuristics and pattern recognition in the diagnostic process. Lancet, 1968, 723 - 726.
12. Taylor T.R., Aitchison J., McGirr E.M.,: Doctors as decision makers: a computer assisted study of diagnosis as a cognitive skill. Br. Med. J. 1971, 3, 35 - 40.
13. Feinstein A.R.: An analysis of diagnostic reasoning. II. The strategy of intermediate decisions. Yale J. Bio. Med. 1973, 46, 264 - 283.
14. Anderson G., Lerena C., Davidson D., Taylor T.R.,: Practical application of computer assisted decision making in an antenatal clinic - a feasibility study, Meth. Inf. Med., 1976, 15, 224 - 229.
15. Lahaye D., Roosels D., Viaene J.,: The value of Subjective Appreciation in the Medical Record. II. The Evaluation of the general condition of the Patient. Meth. Inf. Med., 1978, 17, 103 - 105.
16. Zola I.K.: Problems of communication, diagnosis, and patient care: the interplay of patient, physician and clinic organization. J. Med. Educ., 1963, 38, 829 - 38.

16a Mechanic, D: Social psychologic factors affecting presentation of bodily complaints. New Engl.J.Medicine,1972,286,1132 - 39.

17. Korsch B.M., Negrete V.F.,: Doctor - Patient Communication. Scientific American, 1974, 7, 66 - 74.
18. Coulthard M., Ashby M., :Talking with the doctor J. of Communication, 1975, 25, 140 - 147.
19. Larsen K.M., Smith C.K.,: Assessment of nonverbal communication in the patient-physician interview. J. Fam. Practice, 1981, 12, 481 - 8.
20. Adlassnig, K.P.,: A fuzzy logical model of computer assisted medical diagnosis. Meth. Inf. Med., 1980, 19, 141 - 148.
21. Balla J.I.,: Logical thinking and the diagnostic process. Meth. Inf. Med., 1980, 19, 88 - 92.
22. Blois M.S.,: Conceptual issues in computer aided diagnosis and the hierarchical nature of medical knowledge. J. Med. Phil., 1983, 8, 29 -50.
23. Dornfest F.D.,: The problem solving model of diagnosis in general practice. South. Frica Med. J., 1981, 59, 298 - 301.
24. Cutler. P.,: Problem solving in clinical medicine, Baltimore, Williams & Wilkins, 1979.
25. Goldfinger S.E.,: The problem oriented record: A critique from a believer. N. Eng. J. Med. 1973, 288, 606 - 608.
26. Weed L.L.,: Medical records, medical education and patient care. The problem oriented record as a basic tool. Chicago, Year Book Medic. Pub. Inc. 1970.

27. Feinstein A.R.,: The problem of the 'problem-oriented medical record'. Ann. Intern. Med., 1973, 78, 751 - 762.
28. Young D. W.,: A survey of decision aids for clinicians. Br. Med. J., 1982, 285, 1332 - 1336.
29. Wulff H.R.,: Rational Diagnosis and treatment. Oxford, Blackwell Scientific Publ., 1976.
30. Cumberbatch J., Heaps H.S.,: A disease conscious method for sequential diagnosis by use of disease probabilities without assumption of symptom independence. Int. J. Biomed. Comp. 1976, 7, 61 - 78.
31. Esogbue A.O., Elder C., Fuzzy sets and the modeling of physician decision processes, part I. Fuzzy sets and systems,1979,2,280 - 291.
32. Esogbue A.O., Elder C.,: Fuzzy sets and the modeling of physician decision processes, part II: Fuzzy diagnosis decision models. Fuzzy sets and systems, 1980, 3, 1 - 9.
33. Sanchez E.,: Inverse of fuzzy relations. Application to possibility distributions and medical diagnosis. Fuzzy sets and systems, 1979, 2, 75 - 86.
34. Lahaye D., Roosels D., Viaene J.,: The value of subjective appreciation in the medical record. I. The evaluation of dyspnoea. Meth. Inf. Med., 1978, 17, 100 - 103.
35. Zieve L.,: Misinterpretations and abuse of laboratory tests by clinicians. Ann. N.Y. Academy of Sciences. 563 - 572.
36. Griner P.F., Mayewski R.J., Mushlin A.I., Greenland P.,: Selection and interpretation of diagnostic tests and procedures. Principles and applications. Ann. Int. Med., 1981, 94, 553 - 600.
37. Popper K.R.: Objective knowledge. Oxford, Clarendon, 1972.
38. Nobrega F.T., Morrow G.W., Smoldt R.K., Offords K.P.,: Quality assessment in the hypertension: analysis of process and outcome methods. New Engl. J.Med., 1977, 296, 145 - 148.
39. Gill P.W., Leaper D.J., Guillou P.J., Staniland J.R., Horrocks J.C., Dombal F.T. de,: Observer variation in clinical diagnnosis. A computer aided assessment of its magnitude and importance in 552 patients with abdominal pain. Meth.Inf.Med., 1973, 12, 108 - 113.
40. Koran L.M.,: The reliability of clinical methods, data and judgments, N. Eng. J. Med., 1975, 293, 642 - 646.
41. Koran L.M.,: The reliability of clinical methods, data and judgments (second of two parts). N. Eng. J. Med., 1975, 695 -701.
42. Stern R.B., Knill-Jones R.P., Williams R., : Use of computer program for diagnosing jaundice in district hospitals and specialized liver unit. Br. Med. J. 1975, 2, 659 - 662.
43. Spitzer R.L., Fleiss J.L.,: A re-analysis of the reliability of psychiatric diagnosis. Br. J. Psychiat., 1974, 125, 341 -347.
44. Baumann U., Schneidewind G., Angst J., Helmchen H., Hippius H., : Zur Entscheidungslogik bei Symptom-Beurteilungen. Arch. Psychiatr. Nervenkrankheiten, 1975, 220, 225 - 235.
45. Reid D.D., Anderson D.O., Ferries B.G., Fletcher C.M. : An Anglo-American Comparison of the Prevalence of Bronchitis. Brit. Med. J. 1964, II, 1487 - 1491.
46. Wagner G., Tautu P., Wolber U., : Problems of medical diagnosis. A bibliography. Meth. Inf. Med. 1978, 17, 55 - 74.
47. Wagner G.,: Fehlerforschung als Aufgabe der Medizinischen Dokumentation. Meth. Inf. Med. 1964, 3, 93 - 94.
48. Cochrane A.L., Chapman P.J., Oldham P.D.,: Observer's errors in taking medical histories. Lancet, 1951, 1007 - 1009.
49. Dombal de F.T., Horrocks J.C., Staniland J.R., Guillou P.J., :Pattern recognition: a comparison of the performance of clinicians and non-clinicians - with a note on the performance of a computer-based system. Meth. Inf. Med. 1972, 11, 32 - 37.
50. Lichtenstein S., Slovic P., Fischhoff B., Layman M., Combs B.,: Judged frequency of lethal events. J. Exp. Psych., 1978, 4, 551 - 578.

51. Tversky, A., Kahnemann D.,: Judgment under uncertainty: Heuristics and biases. Biases in jedugment reveal some heuristics of thinking under uncertainty. Science, 1974, 185, 1124 - 1131.
52. Slovic P. Fischhoff B, Lichtenstein S.,: Behavioral decision theory. Ann. Rev. Psych., 1977, 28, 1 - 19.
53. Scheff T.J.,: Decision rules, types of error and their consequences in medical diagnosis. 97 - 107.
54. Driscoll J.M., Lanzetta J.T., : Effects of two sources of uncertainty in decision making. Psych. reports, 1965, 17, 635 - 648.
55. Gerritsma J.G.M., Smal J.A.,: De werkwijze van huisarts en internist. Een vergelijkend onderzoek m.b.v. een interactieve patienten simulatie. Utrecht, W.U. Bunge,1982.
56. McWhinney,I.R.: Family medicine as a science. J.Fam.Pract. 1978,7,53-58.
57. Sokal R.R.,: Classification, purposes, principles, progress, prospects. Science, 1974, 185, 1115 - 1123.
58. Margolis J.,: The concept of disease. J. Med. Philosophy, 1976, 1, 238 - 255.
59. Gutman Rosenkranz B.,: Causal thinking in erewhon and elsewhere. J. Med. Phil., 1976, 1, 372 - 384.
60. Scriven M.,: Clinical judgment. In: Engelhardt H.T., Spicker S.F., Towers B.,: Clinical Judgment, a critical appraisal. D. Reidel Publ. Co. Dordr. 1979.
61. Dombal de F.T., Staniland J.R., Clamp S.E.,: Geographical variation in disease presentation. Med. Dec. Mak. 1981, 1, 59 - 69.
62. Card W.I.,: Clinical decision making (1): an analysis. Health Bull., 1977, 207 - 212.
63. Einhorn H.J.,: Learning from experience and suboptimal rules in decision making, in : Wallsten T. (ed.): Cognitive processes in choice and decision behaviour. Hillsdale N.J., Lawrence Erlbaum, 1979.
64. Tautu P., Wagner G.,: The process of medical diagnosis: routes of mathematical investigation. Meth. Inf. Med. 1978, 17, 1 - 9.
65. Brehmer B.,: In one word, not from experience. Acta Psychologica, 1980, 45, 223 - 241.
66. Feinstein A.R.: Clinical judgment. Baltimore, Williams & Wilkins, 1967.
67. Janis I.L., Mann L.,: Emergency decision making: a theoretical analysis of responses to disaster warnings. J. Human Stress, 1977, 35 - 48.
68. Rimoldi, H.J.A.,: A technique for the study of problem solving. Educ. Psycholog. Measurement, 1955, 13, 450 - 461.
69. Ledley R.S., Lusted L.B.,: Reasoning foundations of medical diagnosis. Symbolic logic, probability and value theory aid our understanding of how physicians reason. Science 1979, 130, 9 - 21.
70. Meehl, P.E., Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis, Univ. of Minnesota Press. 1954.
71. Warner H.R., Toronto A.F., Veassy L.G., Stephenson R.A.,: Methematical approach to medical diagnosis. JAMA, 1961, 177, 177 - 183.
72. Nash F.A.,: The application of computers to medical and veterinary science. Proc. Roy. Soc. Med. 1963, 56, 555 - 560.
73. Koller S.,: Uebersichten. Mathematische-statistische Grundlagen der Diagnostik. Klin. Wochenschrift, 1967, 21, 1065 - 1072.
74. Reichertz P.L.,: Towards systematization. Meth. Inf. Med. 1977, 16, 125 - 130.
75. Crean G.P., Clinical decision making: an introduction. Health Bull., 1977, 205 - 206.
76. Eddy D.M., Clanton C.H.,: The art of diagnosis: solving the clinico-pathological exercise. N. Eng. J. Med., 1982, 306, 1263 - 1268.
77. Bruner J.S., Goodnow J.J., Austin G.A.,: A study of thinking. New York,Wiley,1956.

78. Miller G.A.,: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psych. Review, 1956, 63, 81 - 97.
79. Hovland C.I.: Communication analysis of concept learning Psycholog. Review, 1952, 59, 461 - 472.
80. Mandler G.: in Spence, K. W., Taylor Spence J.T., : The psychology of learning and motivation. New York,London,Academ.Press, 1967.
81. Kleinmuntz B.: The processing of clinical information by man and machine, in: Formal representation of human judgment. Pg 149 - 186. New York, John Wiley & Sons, 1968.
82. Wortman P.M.,: Medical diagnosis: an information processing approach. Comp. Biomed. Research, 1972, 5, 315 - 328.
83. Wortman P.M., Greenberg L.D.,: Coding, recoding and decoding of hierarchical information in long-term memory. J. Verbal Learning and Verbal Behaviour (currently J.Memory & Language), 1971, 10,, 234 - 243.
84. Groot A.D. de,: Het denken van de schaker. Amsterdam, North Holl. Publ. Cy., 1946.
85. McMullin E.: Introduction. J.Med.Phil. 1983, 8, 3-4.
86. Szolovits P., Pauker S.G.,: Computers and clinical decision making: Whether, How and for Whom?. Proceedings IEEE, 1979, 67, 1124 - 1226.
87. Newell A., Simon H.A.,: Human problem solving, Englewood Cliffs, New Yersey, Prentice Hall, 1972.
88. Simon H.A., Newell A.,: Human problem solving. The state of the theory in 1970. Am. Psych., 1971, 26, 145 - 159.
89. Doroszewski J., : Unity and Diversity of the medical action: A review of its components and their interconnections. Metamedicine, 1981, 2, 155 - 68.
90. Elstein A.S., Shulman L.S., Sprafka S.: Medical Problem Solving, Harvard University Press. Cambridge (Mass.) 1978.
91. Feinstein A.R.,: The haze of Bayes, the aerial palaces of decision analysis and the computerized Ouija Board. Clin. Pharm. Ther. 1977, 21, 482 - 496.
92. Sacket D.L.: Four strategies, in: Weltman R., : unsolicited data - to ignore or not? JAMA, 1977, 238, 2466 - 2469.
93. Kleinmuntz D.N., Kleinmuntz B.,: Decision strategies in simulated environments. Beh. Science, 1981, 26, 294 - 305.

CHAPTER V

MEDICAL DECISION MAKING

Humans make decisions, a tremendous number everyday, consciously or unconsciously, ranging from small ones, 'how do I like my egg in the morning?' till large ones, 'how many million dollars to invest in an uncertain business prospect?"; from the move of a chess player till the general's decision to sacrifice the lives of several hundreds of soldiers. Most of these decisions are based on intuition or "common sense", or what is assumed to be a more firm basis, experience. The various unsatisfying outcomes, not to speak of disasters, resulting from this kind of decision making, urged people to look for strategies yielding more successful outcomes, and ways were sought to originate and define strategies which enable us to predict more reliably the outcomes of our decisions.

It can hardly be coincidental that a large part of investigation in this field has been focused on that particular situation where personal gains and losses are so obviously demonstrated: the gambling table. When decision making is seen as the optimalization of the reliability of predictions, the study of decision making started in the 17th and 18th centuries with philosophers and mathematicians like e.g. Descartes, Poisson, and de Moivre (doctrine de Chances) in France, Gauss and Cramer in Germany, Bernouilli in Switzerland, and John Stuart Mill and Thomas Bayes in England. Cramer and Bernouilli were the first who, in the 18th century, began with the rational economic man concepts (1). Whether the formulated mathematical solution really brought what their principals (high-ranking noblemen) expected, is highly questionable.

People still continue to lose their money in the various gambling houses in the world, to the intense joy and pleasure of the owners of these houses. Besides, the very complicated mathematical statistical procedures prevent routine application of the proposed statistical formulae in these types of gambling / decision making situations. The various theories and procedures remained dormant till the introduction of equipment that made the calculation of the proposed formula much easier and quicker: the computer. The second world war accelerated the process, and in the fourties and fifties of this century, a 'revival' of the gambling theories, mostly renamed 'Decision Theory', could be greeted.

The first applications of these theories to practical situations were in the areas of warfare and economics . In the next decades these theories and their various modifications were applied and tested in other fields of science, e.g. medicine. It is against this background that medical decision making developed as an independent discipline not actually related to the actual process of medical practice, but by using 'medicine' as a means to test the theories. In this way it had to be 'normative' or 'prescriptive' to the actual decision making.

Patients more and more question the -sometimes fallible- predictions of their physicians, extreme variance in therapeutic outcomes, variabilities of morbidity and mortality figures, rising cost in health care without a substantial increase in health status, yes sometimes even decreasing, and several other elements which worked together to question the 'va-et-vient' of the medical practice. Several attempts to gain insight in the medical process, and to optimalize it, have been made and are made. Among these attempts are e.g. Problem-Oriented-Medical Record (2), Medical Audit, Protocols (how to diagnose and act in certain cases prescribed by experts), Flow Charts (3), and the several variants of medical decision making (sometimes also called medical or clinical decision analysis), Thusfar, the results of all these attempts are disencouraging.

Decision theory has much to learn from medicine without trumping up their

methods, and medicine in its turn can gain much advantage from the views and procedures of decision theory. Outstanding people like Simon & Newell, Kleinmuntz, Feinstein etc. are more and more sought after to try and bridge the gap between medical problem-solving and medical decision.

Paragraph 2

DECISION THEORY

Decision theory is a group of related constructs that seek to describe or prescribe how individuals or groups of people choose a course of action when faced with several alternatives and a variable amount of knowledge about the determinants or the outcomes of those alternatives (4). The theory can be distinguished into two types, a theory concerning descriptive (how people do behave) or prescriptive (how people should behave) decisions, both for either individual or group decisions. Decision making can be defined as the process of thought and action involving an irrevocable allocation of resources that culminates in choice behaviour. In making a decision, a decisionmaker is dealing with environments, characterized by risks, uncertainty, complexity, changes over time and conflict (1).

The decisionmaker invariably has to choose among a number of alternatives either diagnoses or therapeutic actions. The quality of a decision depends upon how well the decisionmaker is able to acquire information, to analyse the information, and to evaluate and interpret information such as to discriminate between relevant and irrelevant bits of data; it also depends upon how well the decisionmaker is able to cope with the stress which is invariably encountered in important decision circumstances (1). It is essentially for human decisionmakers to bring order into their information acquiring and processing activities, when confronted with an excess of information, unreliable information or a lack of sufficient information.

People are flooded by information which must somehow be reduced and simplified to allow efficient processing and to avoid an otherwise overwhelming overload (something like trying to read all medical journals at once with no system for ordering, reducing, or altogether avoiding data overload). However, this reduction and simplification process, called after Mischel (5) "cognitive economics", involves various dangers. Mischel is especially concerned about the routes of simplification and the growth of self-knowledge and rules for self-regulation with maturition.

Decision theories give rules for optimal processing of data. One of the best-known (and on which most of the decision procedures are based) is the Theory of Games and Economic Behaviour (Von Neumann & Morgenstern, 1947), which proposed several decision rules to handle decisions of complete ignorance; "to find mathematically complete principles which define 'rational behaviour' (...) as a set of rules for each participant which tell him how to behave in every situation which may conceivably arise". Decision theory rests on five main conceptual components:states of nature, actions, outcomes, probability and utility functions.

a) "the set of states of nature" is assumed to form a mutually exclusive and exhaustive listing of those aspects of nature which are relevant to this particular choice problem and about which the decision maker is uncertain. In decision theory two possibly very different states of the world which are generated or are related to the same outcome for a given action, will be labeled the same (4);

b) The set of possible actions and outcomes is finite, complete, and invariant for a given problem;

c) The optimal solution depends directly on the probability assignments (weightings to the component elements of the state of nature)

d) The sum of the probability of all outcomes for a given action must be unity;

e) The probability assignment reflect the confidence of the decisionmaker in the likelihood of outcomes (4).

Outcome directed distinctions are:

1) Decisions under certainty : where each alternative course of action has a single well-specified outcome;

2) Decisions under risk or uncertainty , where each alternative course of action has a well-defined set of possible outcomes each with a probability of occurrence;

3) Decisions under ignorance : when each action results in a range of possible outcomes but the probability of occurence of each outcome is unknown (Luce & Raiffa) (4,6).

Distinctions 2) and 3) especially refer to the medical situation. The physician mainly has to make a choice among a number of alternatives of which the outcomes with their probabilities concerning this particular patient are either uncertain or completely unknown. His choice will depend, to a large extent, upon the decision situational, structural model as reflected in the contingency task. He chooses the most preferred alternative generally from partial or total ordening or prioritization of the alternatives (1).

Implicit in this description is the assumption that there is an underlying logical structure (7). Ledley & Lusted (8) proposed three mathematical-logical, disciplines to the medical process: symbolic logic, probability and value theory; each of them had led to separate, although interrelated, areas of study, based on the Von Neumann & Morgenstern theory.

Most decision theorists base their constructs on a cybernetical disease concept regardless of medical substantiation. "Since doctors themselves don't take active part in it, the main thrust is carried by people coming from informatics as such. What the physician needs, if even, is a decision support for the steps in his clinical decision, not the ultimate classification of the problems" (9).

Although comparisons with physicians' work are regularly made, and are spoken of as 'computer-aids' to doctors, physicians are often blamed for not behaving in a 'probabilistic fashion', or 'unable to estimate probabilities and utilities', suggesting that doctors should think and act in these ways. "They violate the principles of rational decision making when judging probabilities, making predictions, or otherwise attempting to cope with probabilistic tasks" (10). But, the cybernetical concept of medicine as it stands may be questioned, as was discussed in the previous chapters.

The stress between theory and reality colours the discussion about medical decision making. It also differentiates decision making from problem solving. Medical problem solving can be viewed as an analysis of the steps physicians take to reach a solution. It is mainly descriptive in its nature. The theory of clinical decision making is a prescriptive approach to the process of clinical judgment and is based on a theory, which has not been originated within the medical world. It can be viewed as a multistep process which culminates in the selection of one alternative in preference to another.It delineates the steps a physician ought to make in order to arrive at an optimal decision.

These step can be delineated as:

1. Ascertain the need for a decision;
2. Establish decision criteria;
3. Allocate weights to criteria;
4. Develop alternatives;
5. Evaluate alternatives;
6. Select the best alternative (11).

It would be accurate to state that the theory of decision making is a normative theory, presenting how decisions should be made. Yet, there is a large body of descriptive literature on decision making which gives a different picture. It suggests that actual decision making is not nearly as objective and straightforward as the normative picture presents. It largely reflects the difference between the optimizing and the satisficing decision making theories,

as discussed in chapter II. The optimizing man is completely rational, making "optimal" choices in a highly specified and clearly defined environment. When confronted with a problem the "optimizing man" should

a.) clearly define the decision criteria;
b.) be knowledgeable of all relevant alternatives;
c.) be aware of all possible consequences against the decision criteria;
d.) evaluate the consequences against the decision criteria;
e.) rank alternatives in a preferred order of rational impact;
f.) select the alternative which rates highest in terms of health state (11).

Simon rejected this economic man model in favour of one that he believed more accurately reflected reality. His "satisficing" theory, or "administrative man model", proposes that the decision maker will

1.) recognize only a limited number of decision criteria;
2.) propose only a limited number of alternatives;
3.) be aware of only a few of the consequences of each alternative;
4.) formulate a simplified and limited model of the real situation;
5.) select the alternative which presents a satisfactory solution. (11).

(compare with heuristic strategy of Kleinmuntz & Kleinmuntz (12)).

The decision criteria in medicine are largely defined by symptoms, diagnostic hypotheses and the prevalence of the diseases in the diagnostic phase and diagnoses, therapeutic actions and outcomes in the therapeutic phase. In this paragraph the main elements of the decision-theoretical approach of the medical process will be discussed.

This approach is based on the assumption of a linear sequential (logical) process of hierarchical steps. Its primary aim is to express the physician's activities in mathematically recognisable terms (13). As we have sketched in Chapter IV the structure of the cybernetical concept can be delineated as a sequence of hypotheses and tests. The hypothesis is generated by the physician on basis of the acquired evidence, his medical knowledge, and his past knowledge about the presumed diseases, mirroring their prevalences. Each test (which can be any question to the patient, any physical examination or laboratory test) and test result is weighed against the possible alternatives (hypotheses), as they are depicted in a multibranching decision tree. The weighing is calculated as the proportion that a disease is present or absent given the result as positive or negative. These proportion are called after Yerushalmy & Palmer (14) sensitivity respectively specificity.

Sensitivity is the ability of a test to give a positive finding when the person tested truly suffers from the disease in question. It can be calculated as follows:

$$\text{Sensitivity} = \frac{\text{diseased persons with a positive test}}{\text{all diseased subjects tested}} \times 100$$

Specificity is the ability of a test to give a negative finding when the person tested is free of the disease under study. It is calculated as:

$$\text{Specificity} = \frac{\text{nondiseased persons negative to the test}}{\text{all nondiseased subjects tested}} \times 100$$

These elements are customarily denoted as the probability of having a symptom (or positive test) given the disease: P [S|D], or the reverse, the probability of not having a symptom (or negative test) given the absence of the disease, P [s|d]. The chance relation of a symptom (or test) to a - specified - disease entity is called conditional probability, the probability of (not) having the symptom given the (non) disease. In the following we shall use capital letters for the positive results and lower case letters for the negative side.

As we have sketched in Chapter IV the bivariable distribution can be pictured as:

Schema 1

Bivariate Probability Distribution

DISEASE / SYMPTOM	PRESENT	ABSENT
POSITIVE	p.P[S:D]	(1-p)P[S:d]
NEGATIVE	p.P[s:D]	(1-p)P[s:d]

The total population consists of a proportion (p) of subjects who actually have the disease, and the remainder of nondiseased subjects (1 - p).The proportion of these subjects yielding a positive test will equals p.P[S:D] (true positives), plus the false positives [1 -p][1 - P(s:d)]. It follows that the proportion of subjects in the total population yielding a negative test will consist of [1 - p]P[s:d] (true negatives), plus the false negatives p.P[s:d]. these relations will be expressed in the next table

Table 1

Proportions of subjects in the total population who have or do not have actual disease and the proportions of positive and negative tests (symptoms present or absent) in these groups.*

Classification	Total Population	Yielding Positive Test	Yielding Negative Test
Diseased	p*	p.P[S:D]	p.P[s:D]
Nondiseased	[1-p]	[1-p]P[S:d]	[1-p]P[s:d]
Total	p + [1-p]	p.P[S:D] + [1-p]P[S:d]	p.P[s:D]+[1-p]P[s:d]

* symbols are explained in the text

If we know the prevalence p of a disease, and the sensitivity and the specificity of a test or symptom given the disease, we are able to calculate the various aspects of this type of disease-symptom relationship.

For tests with a continuous scale of values (e.g. blood pressure, pulse rate, blood sugar levels) various cutoff-points can be selected to adjust the sensitivity and the specificity of the test in conformity with the decision maker's goals. These cutoff-points can be arranged in curves which are called after Lusted the Receiver Operating Curves (ROC)(15).

When test results are expressed dichotomously, the predictive value of a positive test result may be defined as the frequency of times that the presence of a symptom will detect a diseased individual. This can be calculated as

$$\text{Predictive Value} = \frac{\text{Number (or proportion) of diseased persons with a positive test/symptom}}{\text{number (or proportion) of persons with a positive test /symptom}}$$

$$\text{PV positive} = P[D|S] = \frac{p.P[S|D]}{p.P[S|D] + [1-p][1-P(s|d)]}$$

The negative predictive value is given by

$$\text{PV negative} = P[d|s] = \frac{[1-p]P[s|d]}{[1-p]P[s|d] + p[1-P(S|D)]}$$

It may be clear that positive predictive value or discriminative force is related to the frequency with which a disease is present in the population. When the presence of the disease in the population is practically zero the probability that a certain symptom or symptom configuration predicts its presence, or can discriminate among other diseases, is rather low. This can be examplified. When the sensitivity is fixed at 80% and the specificity at 90%, then the PVpositive and PV negative change with the disease prevalence according to the following table.

Table 2

The relation between predictive values and disease prevalence.

Actual Disease Prevalence	Pv pos.	Pv neg.
0.1%	0.8%	99.98%
1%	7.45%	99.78%
10%	47.1%	97.59%
20%	66.7%	94.74%
50%	88.9%	81.82%

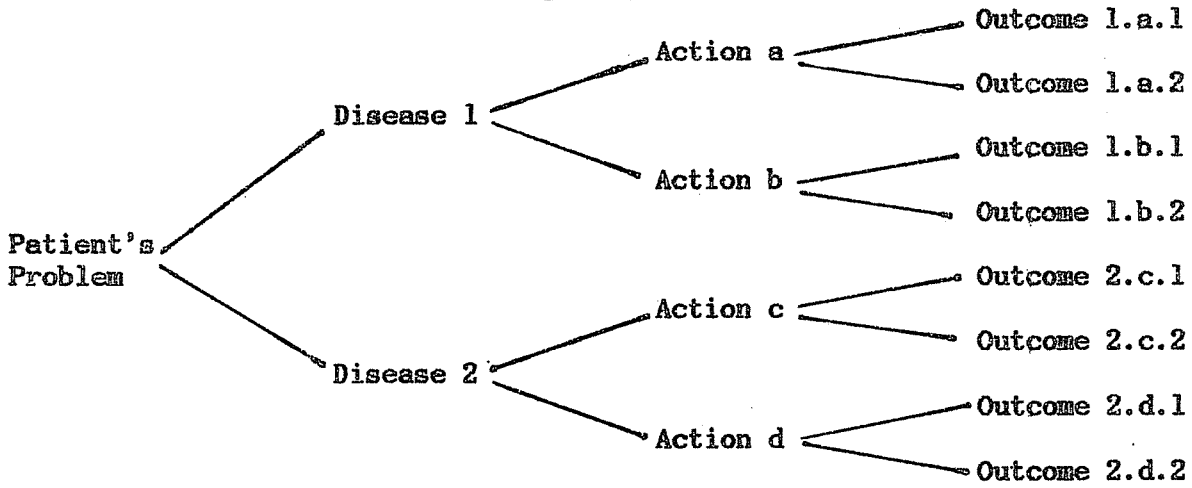
With the sensitivity and the specificity prefixed the disease prevalence particularly effects the predictive values for positive test results. This is especially evident for the lower disease prevalences, the specific domain of primary care.

(The relation between PV positive and PV negative with a fixed disease prevalence is exemplified in Vecchio's paper)

The essence of the decision making is choice. The decision maker has to choose among alternatives, e.g. between diseases or therapeutic actions. The decision process is customarily depicted as a multiple branching tree, a decision tree.

Schema 2

An example of a decision tree.



Where branches come together two kinds of nodes can be distinguished: Decision (or choice) nodes, symbolized with O, and Chance nodes, symbolized as O.

A decision node denotes a point in time at which the decision maker can select one of the several alternative courses of action; a chance node denotes a point in time at which one of the several possible events beyond the control of the decision maker may take place.

A branch or path (sometimes called scenario) in a decision tree is a particular sequence of actions or events beginning with a particular choice at the initial choice node and following a particular event or choice at each subsequent chance or choice node from left to right.

(For a comprehensive survey on this subject see e.g. Weinstein & Fineberg (22)).

The decision tree can be viewed as an aid in identifying alternative clinical strategies. The optimal strategy can be found by assigning 'weights' or 'values' to the various alternatives. The choice between two diagnostic hypotheses can be calculated by means of the values of two ratios. The ratio of two related probabilities can be expressed in odds. When we assume e.g. p[D] to be 0.1, then the corresponding p[d] must be [1 - p], and the odds are

$$\frac{p\,[D]}{p\,[d]} = \frac{0.1}{0.9} = 1 : 9$$

The ratio of two conditional probabilities is called the likelihood ratio. A likelihood of a hypothesis given an event is by definition proportional to or equal to the probability of that event given the hypothesis. The likelihood ratio for a symptom given a disease to the absence of the symptom if the disease is absent can be expressed as

$$\frac{p\,[S|D]}{p\,[s|d]}$$

As is described in Chapter III, section 5, it is a way of judging

hypotheses concerning unknown events. A future event is to be regarded as the more probable the greater the relative frequency of similar events observed sofar under similar circumstances. The predictive value for a disease, symbolized as P [D|S], as calculated with the help of the aforementioned elements can, therefore, only express an inductive probability (when we have seen five white swans, the next one is assumed to be white also).

Having decided upon a certain disease the decision maker is then assumed to make a choice between the various therapeutic actions. Only in rare instances the therapeutic action follows unequivocally from the knowledge about the cause of the established disease entity. Generally, there are several ways leading to Rome, i.e. reaching the optimal state of helath for this particular patient with this specific disease.

According to the conceptual components of Decision theory the likelihood of the outcomes must be expressed in the probability assignments the decision maker attaches to these outcomes. The quantified values as attached to the maximal attainable status of the patient is called 'Utility', a term borrowed from the economics. Since we have no guarantee to actually achieve this goal, we speak of "Expected Utility". The physician's objective becomes the maximization of the expected utility (18).

"Utility theory" is an axiomatic theory of decision making under risk. It specifies the course of action a decision maker should choose to be consistent with his preferences and judgments. It states that the decision maker should choose that course of action which maximizes his expected utility (Luce & Raiffa, quoted in 19). It is, as mentioned, measured in terms of probability and it is this uniformity of measurement that enables him to combine it with that other feature in the decision making process, the probability of the diagnosis, in order to produce a decision in favour of a specified therapeutic action (21).

Utility has been studied by the odds that a subject would demand when offered a "fair" bet (29). Utility or Value theory requires

1) the necessity to assess utilities in all the consequences to be considered;

2) if the likelihood of the expected outcome can be expressed as probabilities, it is possible, by working backwards through the various alternative actions to determine the expected utility of each course of action at each decision point (19).

In Utility theory a number of elements is assumed to exist in the model:

1) a set of policy alternatives, 'a';

2) the set of possible consequences of choice or future states of nature or decision outcomes called 'S';

3) utility function 'U(s)' that is defined for all elements 's' of set 'S';

4) information as to which outcomes will occur if a particular policy alternative 'a' in 'A' is chosen; and

5) information as to the probability of occurrence of any particular outcome if an alternative a A is chosen.

This leads to the simplest form of expected utility of alternative A and can be computed as

$$E\{U(a_i)\} = \sum_{j=1}^{n} P\{s_j(a_i)\}U\{s_j(a_i)\} \qquad (1)$$

All utility theories (discussion of the various theories is beyond the scope of this study) assume that:

a) preferences are governed only by the utilities and outcomes;

b) complete ordering of the expected outcomes is possible;

c) the preference order of the outcomes is transitive, which axiom enables the decisionmaker to calculate the utility of one state if the utilities of the other states are known;

d) independence exists between the various outcomes;

e) enhanced expected outcomes are preferred if and only if a basic expected outcome is preferred (1).

In practice single utilities are not assessed, but groups of them, comparing the various outcomes in different ways to understand aspects of the utility structure (21).

For a medical example we consider 10.000 patients for all of whom two tentative diagnoses, D1 and D2, have been made, with probability 5/8 and 3/8 respectively. Suppose also that there exists a treatment T that is 80% effective against disease complex D1, and 40% effective against disease complex D2. The expected value is the proportion E, which is the sum of the products of the value of the treatment for curing the disease and the probability that a patient suffers from the disease.

$$E = \frac{80}{100} * \frac{5}{8} + \frac{40}{100} * \frac{3}{8} * 10.000 = 0.65$$

which is the probability that an at random selected patient recovers when he is treated (8).

However, between theory and reality there yawns a gap. The theory, thusfar, counts at least two obstacles.

1.) it assumes diseases, symptoms, signs and tests as independent elements, which can hardly be the case in medicine;

2.) we bear no knowledge about the exact values of the various attributes.Some of the problems to these obstacles shall be discussed.

The starting point of the decision making process is considered to be the prior probability. The prior probability is regarded as a summary of what one knows about the presented patient's problem in general, before there is any specific information available (23). But how can we possibly know what physicians know in general about this particular situation? As an accomodation the prevalence or the incidence of the assumed disease is accepted as a value for the prior probability. However, the conception of Balla (23a): "the strength of a physician's initial disposition to regard his patient as having a certain illness is directly proportional to the incidence* of that disease in the population he serves", is contrasted by e.g. Weinstein & Fineberg (22) and many others, assuming the prevalence reflecting the prior probability.

* note.(Incidence can be described as the rate with which a certain disease in a circumscribed population is represented to a physician within a given time span.)

Although incidence is a more realistic term to a physician than prevalence, - he can estimate what he observes -, customarily the prevalence rate is taken as value for the prior probability. However, reliable figures for the prevalence rate are not generally available. Figures available pertain almost always to (too) limited populations and medical fields.

Gustafson et al (24) advocated subjective estimation of the various values by the expert physician. They define subjective probability as any estimate of the probability of an event, which is given by a subject or inferred from his behaviour. Subjective estimates can be derived from experts, frequency data from literature or medical records or from interviewing a random sampled group of physicians. (Opposite to subjective probability is the objective, or statistical, or actuarial probability. Figures, mostly concerning disease prevalence rates, are derived from specific investigations or from epidemiological studies). The Gustafson et al (24) study concludes that the subjective model's developmental cost and time requirement is much lower (than objectively acquired figures), while it functions as well as either actuarial model. In most studies subjectively acquired probability estimates are used nowadays.

Several authors, however, question the validity and reliability of these estimates. They point to the highly unreliable estimation of chances of people. Physicians, they state, like other scientists and indeed like the ordinary lay person, easily tend to infer, generalize, and predict too much while observing too little. Problems inherent to prior probabilities are e.g.:

1,) the estimates discriminate against rare diseases;

2.) prior probabilities as figures for prevalence are (almost) not available, and when they are, they are hardly reliable. Incidence figures are not available at all;

3.) sampling errors;

4.) estimates may only be applicable to a given population in a particular geographical location;

5.) estimates can easily be distorted by seasonal and investigational influences;

6.) estimates may have a personal and situational emphasis. Physicians do not rely heavily on general accessible knowledge like textbooks, journals, continuing education etc. Many estimates rely on direct experience with their own patients. Kochen (25) found that physicians primarily rely on:

a.) own experience : 100%
b.) literature : 65%
c.) medical school : 24%

About the reliability of the physician's memory, Kochen found that

- physicians remembered cases and exactly what was done: 5%
- only recalled names of past patients : 65%

These findings question the reliability of subjective judgement. Moreover, the judgements are subject to certain systematic biases that can produce serious distortions and oversimplifications in inferences and predictions (5). Because the very factors that enhance the judge's subjective confidence are often negatively correlated with the accuracy of their estimation, the "illusion of validity" may be observed as a general trend. In their interesting studies Tversky & Kahneman (26,27,28,29) sketched significant biases in people's judgment under uncertainty. People, like physicians, use very peculiar rules in estimating figures from past events. Not only do these figures differ from background to background, population, and sample size but also from person to person, and in persons over time. The rules are commonly named heuristic rules. Tversky & Kahneman investigated and described four heuristic rules:

Representativeness : When making inference from data too much weight is given to results of small samples. As sample size is increased, the results of small samples are taken to be representative of the larger population. The judgements based on this rule are:

a.) insensitive to prior probabilities of outcomes;

b.) insensitive to sample size;

c.) liable to misconception of the randomness of a sequence of events: (gambler's fallacy);

d.) insensitive to predictability: favourable description favour prediction more than the reliability of that description;

e.) illusory validity: confidence in the representativeness or similarity of an event. Unwarranted confidence in a good fit;

f.) liable to misconception of regression (towards the mean), concerning randomness of scores around a reference point.

Availability : The decisionmaker uses only easily available information and ignores difficult available sources of significant information. An event is believed to occur frequently, that is with high probability, if it is easy to recall similar events. These biases are due:

a.) to retrievability of instances. Easy retrievability will appear more numerous than the reverse. In addition, familiarity and salience play their role in this heuristic;

b.) to the effectiveness of a search set;

c.) imaginability: the ease with which a certain instance can be constructed from memory;

d.) illusory correlation: presumed co-occurrences of two or more elements e.g. peculiar eyes and suspiciousness.

Adjustment and Anchoring : the decisionmaker is often flooded by an excess of data. He may reduce his mental efforts by fixing a reference point as an initial value, a starting point suggested by the formulation of the problem. The fixed value can be every datum chosen at random in the amount of data e.g. the mean, and then adjusts that value improperly in order to incorporate the rest of the data such as to result in flawed information analysis. Biases may arise from:

a.) insufficient adjustment;

b.) the evaluation of conjunctive and disjunctive events;

overestimation of conjunctions and underestimation of disjunctive events;

c.) anchoring (the fixed value or anchor) in the assessment of subjective probabilities distribution.

We know for sure that every reader recognizes several of these heuristics in his personal sphere. Otherwise he can hear and see them everyday on radio and television from politicians, news and advertising agents and the like.

But, unfortunately, these heuristics are not the only ones. In his inspiring review on information systems and processes and decision report, Sage (239) listed a total of 27 biases, to which Eraker et al (30) added several more. Nevertheless, people are quite confident about their fallible judgments, as was described by Fischoff et al in their entertaining article: 'Knowing with certainty' (31).

If there is a possibility for calculating a diagnosis, and it seems to be, medicine as a discipline is, up till now, unable to provide really reliable and valid figures for such procedures on a broad scale.

Nevertheless, a large number of exercises in this field show remarkable results. Several systems perform at least as well as physicians, often better. But we must realize that because of several built-in restrictions results are difficult to compare. Besides, differing classifications and taxonomies can distort conclusions. As a normative concept decision theory has to rely on a generally accepted medical taxonomy. As we argued before this conception may be different from the facts and actual medical practice.

The most difficult and ambiguous element of medical decision making is the expected utility theory. Lindley (21) believes that "it is one of the great triumphs of modern thinking to show that only by assigning such a single numerical measure sensible decisions can be made. Only by measuring the quality of life can rational decisions be made". Still, it is an open question how to quantify the quality of life and who is to quantify it. 'People often exhibit patterns of preference which appear incompatible with the expected utility theory. Because of the impossibility to establish objective criteria for measuring someone's quality of life, one has to rely on subjective estimation.

For expected utility theory it is that believed that people react in a risk averse way. Kahneman & Tversky (32) showed that the concept of utility as a concave function of money is violated in several ways according to the framing of the presented problem.

People are risk averse when one has something to gain but risk seeking when one has something to lose. Risk aversion in the positive domain is accompanied by risk seeking in the negative domain (the reflection effect). Besides, people overweight outcomes which are considered certain, relatively to outcomes which are merely probable - a phenomenon which is labelled the certainty effect. It stands next to the pseudocertainty effect when descriptions of decision problems favour conditional evaluation with equal outcomes. (33). People often disregard components that the alternatives share, and focus on the components that distinguish them. This approach to choice problems may produce inconsistent preferences, as a pair of prospects which can be decomposed into common and distinctive decompositions lead to different preferences. Many carriers of value or utility are changes of circumstances rather than final positions. Kahneman & Tversky (32) propose the Prospect Theory to accomodate the violations of

expected utility theory.

As long as the diagnosis is clearcut, the number of possible (therapeutic) actions - very - limited and the outcomes as states of health can be defined in two states: dead or alive, the problem of utility can easily be solved. But when do we, physicians, meet such a situation? And if we meet it, do we need these uncertain calculation to govern our steps? Doctors' problems always are concerned with nonsimple questions, with several possible actions, and a variety of different outcomes, largely beyond determination. Regarding medical options in terms of life and death is inadequate. People are more than a mere biological phenomenon.

Every individual estimates his own "utility", as well as society which may place a numerical worth upon its subject. But also the physician implicitly places a utility upon the patient. Whenever physicians make clinical decisions they integrate their own value system to generate preferences for alternative diagnoses, therapies and outcomes. The greater the cultural gap between patient and physician, the more difficult this appreciation of values (4). This seems especially true for decisions under risk.

The estimation of a utility value is submitted to the appreciation of a specific situation, in which the preferences of the physician and the patient can be conflicting. Apart from rough distinctions as life and death, and a number of years to live (assuming it can be forecast), we cannot see how normative, or objective, utility estimates can be determined.

It means that for every specific process and for every specific state-of-health a generally accepted numerical value has to be established. Of course, every physician makes this type of estimation every day to every patient. And indeed, most processes and most qualities of patients' health states show many resemblances. But apart from the question of an overall agreement among physicians, neither physician nor patient regard themselves as part of statistics. Besides, the computation of subjective estimates of physicians does not make the value more objective. Theories and standards applied in financial and economic disciplines cannot be transplanted to medicine as such. Analyses of preferences will often lead to ill-considered, often accidental incompleteness. Lindblom (34,35) indicates a number of limitations to analysis:

a.) it is fallible, never rises to infallibility, and can be poorly informed, superficial, biased, or mendacious;

b.) it cannot wholly resolve conflicts of value and interests;

c.) sustained analysis may be too slow and too costly compared with realistic needs;

4.) issue formulation questions call for acts of choice or will and suggest that analysis must allow room for policies.

The variability of disease characteristics, predictions, therapeutic and treatment possibilities and appreciations of health states makes fundamental establishment of utility theory and a normative classification of utilities illusory. According to Lindblom means and ends, treatment and outcomes, are often confounded. The identification of values and goals is not distinct from the analysis of alternative actions. Agreement on a good policy does not necessarily include that it is the most appropriate means to an end. Besides, analysis may be limited, important options neglected, and outcomes not considered. It often seems that there is a greater preoccupation with ills to be remedied rather than positive goals to be sought (35).

The changing views on health, disease and cost make normative utility valuation fallacious. What may be wisdom today, may be foolishness to-morrow.

In daily practice the estimation of the utility is most of the times conveniently left to the physician: "Doctor knows what is good for me". When the expected utility is expressed on a scale between 0 and 1, the physician invariably aims at the upper part of the scale. People expect him to react in this way. Every result below this expected outcome leads to severe

disappointments and the patient urges him to diligent looking for more powerful treatments. Moreover, the physician decides upon the whole of the medical process, in which each element depends on and influences the other. A physician is really amazed when he is asked to valuate a single feature from the medical process. Besides, the personality of the physician influences for a great deal the estimation, as was studied and shown by several psychologists (e.g. Tversky & Kahneman, 33,28, Lichtenstein et al. 36, Slovic et al.10).

But when objective values cannot be provided and subjective utilities are unreliable, what about decision theory? De Dombal (90) spoke about utility as "the chief unsolved problem in medical decision making".

Paragraph 3

MEDICAL DECISION MODELS

We shall now consider the process along which solutions can be found for presented problems. In contrast to what might be thought on the basis of a decision theory, there is no uniform model for this process. Instead, an unlimited number of publications on various models have been published, and possible will be still more in future.Most models are based on a couple of similar principles, but authors claim variation. The number of reviews on this subject is still increasing (e.g. Kleiber, Gachowitz & Huber found in the period 1971-1975 more than 1000 relevant references about the application of decision models using Bayes' theorem; Wagner, Tautu und Wolter published a not-exhaustive bibliography of over 800 titles in the twenty year period from 1957 (37); Koehler, Wagner & Wolber compiled Interactive Data Processing in Medicine in the period 1970 till 1976: they found 617 papers.It is beyond the scope of this chapter,and my ability to present a review on this matter. This paragraph presents nothing more than a personal opinion about the subject as it was "sensed" from literature.

Because in medical literature the issue has been left vague, one has to decide whether physicians process the information in a parallel or a sequential way. Ledley & Lusted's paper (8) assumes a parallel fashion. The first application of mathematical-statistical decision making has been performed in this way. Warner et al (38) compiled data on 83 patients and generated a symptom-disease matrix consisting of 53 attributes (symptoms & signs) and 35 disease entities. The system required that all 53 attributes had to be observed for every patient. The processing used a lot of computer time, and patience from the part of the physician. It became clear that a parallelled processing system was hardly feasible, especially with regard to the state of equipment in those days. Using a kind of slide-rule Nask (1963) constructed a Mark IIIA "logoscope" to find data on the rule in a sequential way, but processing had to be performed in a parallel fashion. The logoscope had a repertoire of over 700 million combinations (39); as far as I know this model never found practical aplication. In 1968 Gorry & Barnett (40) introduced a sequential model in which observations were to be considered stepwise. In the same year Edwards et al 41) introdced a Probabilistic Information Processing (P.I.P.) system also based on the principle of sequential processing. Sequential strategy seems to be the preferable choice in medical decision making, realizing the total amount of facts within medical science. Pauker et al (42) estimated the number of facts within internal medicine at about two million.

The sequential statistical inference is based on a definition of probability as a particular measure of the opinions of ideally consistent people. Statistical inference is modification of these opinions in the light of evidence (42). Starting at the level of prior probability every new piece of evidence (symptom, sign, lab result etc) can stepwise change this measure towards the posterior probability: the probability of suffering from the disease in the light of the acquired evidence. Each piece of evidence can give rise to choice among alternative paths. In order to select an alternative plan or course of action for ultimate implementation, the decision maker applies one or more decision rules which enable comparison prioritization, and ultimately, selection of a single policy alternative from among a set of choice alternatives.

The purpose of a decision rule is to specify the most preferred alternative generally from a partial or total ordering of alternatives (1). The choice of a decision rule depends largely on the contingency task structure. Sage (1) recognizes three types of decision rules: wholistic judgment (reasoning by analogy and intuitive affect), heuristic elimination (simplified approximations to holistic decision rules are used), and holistic evaluation: (an attempt to consider all aspects of a decision situation in evaluating choices by means of disaggregation of various choice components). The latter type has been

subdivided in the various utility theories, and will be largely considered here. The holistic decision rule comprises the following requirements:

1) outline the structure of the problem i.e. make a decision-flow-diagram or decision tree showing each possible course of action and the possible outcomes of each action;

2) at each branch point where the outcome is determined not by choice (decision node) but is left to chance (chance node) estimate numerically the probability of occurrence of each outcome;

3) for each possible outcome, assign a relative value called a utility;

4) multiply each utility value by the associated probability of occurrence to obtain a score for that outcome.Sum the scores of the possible outcomes at a chance node to obtain an expected utility for that node. Then sum the scores at the chance node associated with an action to obtain the expected utility of that action;

5) choose the initial action that has the highest calculated expected utility (44).

Distortions can arise at each point of this program. At first the constructing of a flow chart or outlining a decision tree is not only an exhaustive task but is also liable to incorrectness, omissions, personal preferences or opinions etc. A flow chart consists of a hierarchy or sequential system of progressively more specific questions leading ultimately to the most specific diagnostic output justified by the evidence at hand. This encompasses the relationships of each set of alternative acts and outcomes to each stage of decision. It had to contain provision for every contingency that is pertinent. Most of the time these charts or trees are crude, incomplete, and sometimes indeed misleading (45). It assumes that all pathophysiologic, predictive and management facts are available, known or at least traceable in medical knowledge. In our opinion, this is an illusion. And as far as the facts can be elicited, reliable data for most of the necessary probability values are not available. These values are not available from literature because the reported cases can greatly differ from the presented problem.

However, as Kassirer (46) points out, when literature provides unsatisfactory answers, the situation requires the "judgment of an experienced clinician" or applying "common sense". It implies relying on personal judgment and personal intuition, because there is no evidence that probability estimation of experts is more reliable than of laymen. But this is putting the cart before the horse: if you want to improve the decision, don't use inadequate figures or dates. We have to acknowledge that by screening the medical literature we will come across large gaps in existing knowledge. It may only be suggestive for future research.

When constructing and analyzing a complete decision tree seems too elaborate, it is suggested to cut a number of branches, limiting the breadth and the length of the decision tree to the depth of 5 or 6 steps and 3 or 4 different branches. But here we meet another obstacle: who and what determines the cutting and controls the heuristic (limiting search) process? Criteria and rules to this 'cutting procedure' are generally not provided. Moreover, heuristics may not only vary from problem to problem, but also nobody tells us what these heuristics are.

Most models do not allow the patient to suffer from more than one disease at a time (19). In their study Alperovitch et al eliminated 22% of the subjects from the total population under study. It seems to us that the effectiveness of models becomes questionable.

Decision tables may be convenient to portray certain diagnostic decisions that depend on a particular array of information, rather than on a specific sequence in which each component of the array is noted (47). These tables are often illustrated with flow charts, which contain diagrams and graphic symbols for each act of reasoning in the strategy.

Sequential strategy is sometimes referred to as algorithms and protocols. The word algorithm is commonly used in computer activities to refer to the plan

or strategy to solve a problem. Algorithms may be statistical and logical: statistical refers to algorithms which calculate the most likely diagnosis from explicit statistical analysis of disease-symptom frequencies and disease probabilities; logical refers to algorithms which usually proceed in a sequential branching fashion: a decision is made at each step based on a logical 'if A then B' or similar type of reasoning (48). Algorithms have been discussed in detail in Williams (49), and published in medical journals like JAMA, Patient Care, Practitioner. Protocols have been produced to ascertain prescriptive courses of actions. They have especially been developed for the use by paramedical personnel. Their scope is limited because they cannot meet the requirements of the data assessing stage (50).

Flow charts for single medical problems have been designed by Edwards (on dysphagia,51), and Essex (covering rural care in developing countries, (52), and published in books (Patient Care, (3). More complex systems are e.g. ONCOCIN, which used some of the formalisms of the Mycin program based on Artificial Intelligence (53). These latter devices are mainly based on a personal or group interpretation of the specific medical knowledge pertaining to the diseases under study.

The formal decision theoretical models are customarily classified into two groups:

a) mathematical statistical;

b) artificial intelligence.

The former group has been widely investigated and applied, and has an extensive list of literature. Within this group another classification can be made:

1) statistical models using Bayes' theorem;

2) models not using Bayes' theorem.

Because several techniques as linear discriminant function, cluster analysis, and the like can easily be incorporated into this theorem, the "Bayes models" far outrange the latter group. The 'Bayes' models' are generally classified according to a proposal of Card (54), (also used by Wardle & Wardle (55)), into three types:

1) Mark I models which allocate each complete set of symptoms and signs to a disease class, regardless whether a parallel or a sequential structure was used;

2) Mark II models which are diagnostic tree searches in which measures of greatest expected informativeness are used to direct the selection of the next step in the medical process.

3) Mark III models which introduce positive and negative utilities ('costs') in a maximizing expected utility decision tree.

Who was Bayes and what is his theorem?

Reverend Thomas Bayes, an English minister, who lived from 1702 to 1761, left "An essay towards solving a problem in the doctrine of chances" (56), which was communicated to the Royal Academy of Science by his friend Mr. Price. In an introduction Bayes wrote "that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times". "This problem is by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to (provide) a sure foundation for all our reasonings concerning past facts, and what is likely to be hereafter".

Given the number of times in which an unknown event has happened and failed, the problem can be defined as:

THE CHANCE THAT THE PROBABILITY OF ITS HAPPENING IN A SINGLE TRIAL LIES SOMEWHERE BETWEEN ANY TWO DEGREES OF PROBABILITY THAT CAN BE NAMED.

Bayes stated the following definitions:

1: Several events are inconsistent, when, if one of them happens, none of the rest can.

2: Two events are contrary when one, or other of them must, and both together cannot happen.

3: An event is said to fail, when it cannot happen,
or (which comes to the same thing), when its contrary has happened.

4: An event is said to be determined when it either happened or failed.

5: The probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening.

6: By chance I mean the same as probability.

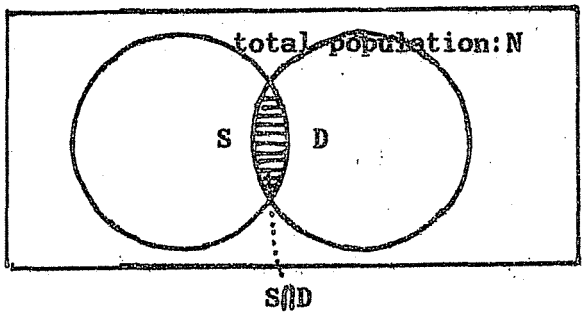
7: Events are independent when the happening of any one of them does neither increase nor abate the probability of the rest.

In the first four definitions Bayes defined the events and their relationships, in definition seven the requirement of independence to the essential formula as it was described in definition five.

This formula can be approached in the classical doctrine of chances. The probability of a die showing a 1 is the ratio of this single face to the total number of all possible faces, or the computed number of ones and non-ones:

$$p\,[1] \quad \frac{1}{1+[\text{non-}1]} = \frac{1}{1+1}$$

(the negation is denoted in lower display). Essential to Bayes' Theorem is the notion of the conditional probability: the probability of an event given another event, not contradicting the former one. It involves two events, with the occurrence of the second event depending on the previous occurrence of the first. In our jargon we can denote this as the probability of a symptom or sign or test given a disease-entity, (P[S¦D]). This means that in an arrangement of two sets of elements, S and D, in a finite population, we are interested in the intersection of the sets S and D, diagrammatically shown in a Venn diagram (see e.g.45,17).



This intersection means nothing else than the true positive symptoms given the disease. In this way we can denote four possibilities in conditional probability:

P[S¦D] = sensitivity
P[s¦D] = false negative rate
P[s¦d] = specificity
P[S¦d] = false positive rate.

If N is the total number of people under study, the constituent groups can be enumerated as n(S), n(s), n(D),
n(S D) etc. By definition of conditional probability:

$$P(D|S) = \frac{n[S|D]}{n[S]} \qquad P \text{ being } \frac{n}{N}$$

which can be transformated to

$$P[S|D] = P[S|D] * P[D]$$

Since we now have two things equal to P[D|S], we set them equal to one another so that

$$P[D|S] * P[S] = P[S|D] * P[D]$$

Solving this equation for the left hand term, we get:

$$P[D|S] = \frac{P[S|D] \times P[D]}{P[S]}$$

In this equation we did not consider the other possible outcomes. Remembering the example with the die, the single event has to be compared with all other events. Having the latter in the denominator we can write the classifical formula as

$$P[D|S] = \frac{P[S|D] * P[D]}{P[S|D] * P[D] + P[S|d] * P[d]}$$

or in words

$$\text{Predictive value} = \frac{\text{Sensitivity} * \text{Prevalence}}{\text{Sensitivity} * \text{Prevalence} + (1\text{-Specificity}) * (1\text{-Prevalence})}$$

The particular feature of Bayes' theorem is that it strives to combine the two types of statistical probabilities: the probability of relative frequencies and the sequence of relative frequencies. As we may remember from chapter III according to Carnap, Keynes and Popper numerical values cannot be attached to a singular hypothesis, which P(D|S) is. The statement "the probability of the next cast with this die being 1 equals 1/6" is not really an assertion about the next cast. Nevertheless, Bayes' Theorem strives to predict future events from past ones, provided that a number of restrictions is fulfilled. Bayesian statistics are based on a definition of probability as a particular measure of the opinions. Bayes' theorem specifies how such a modification of opinions can be made.

The tools include the theory of specific distributions and the principle of stable estimation, which specifies when actual prior opinions may be satisfactorily approximated by a uniform distribution (43). This means that quite a number of requirements have to be fulfilled before reliable calculations can be made. Some of them will be mentioned:

a) a set of reasonable beliefs can be represented by a probability function defined over propositions S and D; reasonable changes of belief can be represented by a process called "conditionalization", i.e. construction of new statements in terms of conditional probabilities (57). It is believed that these beliefs and their changes can be expressed in numerical terms.

b) the measures reflect the opinions of ideally consistent people.

c) the particular set of diseases should be well-defined, so that the frequency of characters in the individual diseases and the frequency of these diseases in the particular population under study are known (58). It requires

that this population, or sample size, is actually known, and of sufficient size to permit satisfactory estimation.The actual performance of a Bayesian calculation is difficult, because it depends on quantitative data that are seldom available.

To apply Bayes' theorem one has to make use of personal judgments of more or less experienced physicians about these probabilities. The drawbacks of this procedure have been sketched elsewhere.

d) the diagnoses used must constitute a complete diagnostic system, i.e. be mutually exclusive and together cover all possible diagnoses (59). Especially mutual exclusivity is only rarely to be found in the setting of the context formation task which occurs when the physician is presented with the patient's chief complaint (60).

e) for being useful the Bayesian approach encompasses also the assumption that the model captures the essential determinants of the judgmental process. Research suggests that it does not (27).

f) posterior binomial estimates have to be determined by sample difference rather than sample proportion. In subjective probability estimation people tend to react in a reverse way. They also do not defend on the population proportion.In his evaluation of evidence, man is apparently not a conservative Bayesian (as is often believed): he is not Bayesian at all (27).

g) Bayesian inference is aimed at specific diagnoses and actions, not at the uncertain decisions of clinicians (45). In other words, it lacks reality.

h) each new piece of evidence acquired by research will change nomenclature, frequency distributions of disease and symptoms and their co-occurrence. It will hamper the application of Bayes' theorem.

j) the theory requires that symptoms and diseases be independent. One of the most serious statistical problems outstanding in diagnosis is the one of symptom interactions. (61). Norusis & Jacquez (62) presented a simple example of the importance of the joint distribution of two variables.

Schema 3

Joint distribution of two symptoms S1, S2.

	Disease 1 S1			Disease 2 S1	
	present	absent		present	absent
present S 2	0.5	0	present S 2	0	0.5
absent	0	0.5	absent	0.5	0

Although both symptoms have equal marginal probabilities in each of the two disease states, considering them jointly permits perfect discrimination. The usual methods of analysis, based only on marginals, would have excluded these variables as unimportant. Norusis & Jacques found that the independence assumption can substantially decrease the effectiveness of a Bayesian classification system. Usually this is not recognized, for true misclassification (of diagnosis, J.R.) probabilities are not known and a basis for comparison is unavailable.

Even moderate perturbations of probability estimates may result in changes of the classification rule, which leads to increases in the misclassification rate. When diseases are easily differentiable, and correlations not extremely

large, the independence assumption may be acceptable (63). It may be therefore that results reported in literature usually focus on diseases for which good classifications have been obtained, and thus it cannot be recognized how deleterious the independence assumption can be for the more challenging situations.

A number of variations of the Bayesian model have been propagated, among which are the Bahadur expansion and Discrimination models. Bahadur's distribution specifies the order of dependence to be accounted for. The model can be set to take account of lesser order dependencies and ignore all higher order dependencies (64,65).The model always chooses that diagnosis which appears with the greatest frequency in the original sample for the profile in question (64). Examination of models which include only pairwise dependencies (linear discriminant function, second-order Bahadur, and optimum dependence trees) suggested that higher-order interactions should be incorporated into estimation procedures for all three second-order models and led to unpredictable results when compared to independence (62). Bahadur's complex mathematical model effectively asks for the relationships which existed , not realizing that patients with the same profile could have different diseases.(64)

Fisher's Linear Discriminant Function also incorporates second order correlations in a multivariate normal framework. The method is simply to weigh the symptom values in such a way that the variation between the disease classes is maximized, relative to the total variation in the sample. Interesting is that the independence and linear discriminant function models are affected very little by sample size. An increase of over four times in sample size gave for LDF procedure an almost negligible reduction in misclassification (63). For very small sample sizes, the discriminant function may lead to slight gains over independence (63). Discriminant functions in their most widely available form are not really appropriate to the discrete data which arise from a study of the presence or absence of the clinical signs and symptoms e.g. because they cannot cope with missing items of data which often occur in practical clinical medicine (66).

Croft (67) tested 10 of the most commonly used mathematical diagnostic models to a same large set of data. He found that their diagnostic accuracy is more sensitive to variations between the diseases than variations between the mathematical models. He recommends future researchers not to continue with increasingly sophisticated mathematical techniques, but, instead, tackle the real obstacles to practical computer-aided medical diagnosis. These obstacles are:

1) lack of standard medical definitions;
2) lack of large, reliable medical data bases, and
3) lack of acceptance of computer-aided diagnosis by the medical profession.

Croft & Machol (68) consider medical diagnosis as a pattern-recognition problem. Confirming the suggestions of Baron & Fraser (58), they recommend to study and reexamine the entire taxonomy of diseases first.

We have to consider that if little is known about the distinguishing features of a disease then there is no magic formula whereby the mathematical statistical procedure can discover such knowledge. In most of the studies little attention has been paid to the dilemma of observer variation. It can largely deteriorate the accuracy of the procedure. One way of increasing accuracy is to eliminate patients because their disease (or disease structures) do not fit in the model. Fraser, who discarded 30 out of 100 patients, points out the fallacy of approaching computer-aided diagnosis from a set of diseases rather than a set of symptoms (cited in 55). Sample size is another matter which influences accuracy. As we pointed out probability, and thus Bayesian statistics, depends on the frequency distribution of symptoms and signs. If these probabilities are derived from a small sample of patients, they will be subject to a large degree of error (55). The most striking difference between Mark I and Mark II models is the amount of information. Especially in Mark I the amount is redundant for

diagnostic purposes. Knill-Jones (69) and Taylor (70) found a marked increase in efficiency with a decrease of information items of approximately 20 items per case. The Mark III models are hampered by the previously mentioned troubles about the estimation of cost and utilities. These models are still in statu nascendi.

Statistical programs normally require a block of data to be entered initially. Those that elicit data sequentially may use either an information or utility theory selection procedure, which may influence the outcomes.

Apart from these statistical models a Fuzzy Set Theory to medical decision making has been developed. This theory is, as far as we know, hardly applied in practice. For some literature on this part see 71,72,73,74.

Furthermore, we have to realize that decision theory does not provide the kind of focusing mechanisms that physicians tend to use when they assume an initial hypothesis in dealing with a patient. As mentioned before, it may be alien to the physician.

Artificial Intelligence approaches started in the early 1970's. The term "artificial intelligence" is generally accepted to include those computer applications that involve symbolic inference rather than strictly numerical calculation. It is based on the hierarchical structured cognitive models as have been originated by investigators like Newell & Simon (75), Kleinmuntz (76), Wortman (77,78) (see also chapter II).

Their efforts have been aimed at understanding and, afterwards, to a simulating and delineating of the human -i.e. the physicians'- reasoning processes. In the renal failure program Gorry analyzed the processes. His conclusions include four points:

1) clinical judgment is mainly based on gross chunks of empirical and/or pathophysiologic knowledge. From a good deal of experience rules of thumb (heuristics) are derived.

2) physicians' knowledge is rather judgmental. The rules they learned allow them to focus attention and to generate hypotheses quickly. These rules permit them to avoid detailed search through the entire problem space.

3) physicians recognize levels of belief or certainty associated with many of the strategies or rules they use. They do not quantify these concepts of belief or certainty, neither do they use them in a mathematical-statistical manner.

4) experts can state their heuristics in response to perceived misconceptions in others easier than to generate them a priori (79).

Crucial to the solution of simulating and delineating physicians' reasoning processes seems to be the knowledge of the heuristics they use. But, "the evidence on rational decision making is largely negative evidence, evidence on what people do NOT do"(7).

Nevertheless, several researchers believe in the 'primacy' of technology, which asserts, in its boldest form, that we can only hope to understand what we ourselves have made, and, as a consequence, that which we have NOT made only insofar as we can handle it 'as if' we had made it. The history of human thought can provide many examples of the use of contemporary technology as a source of explanatory concepts for acquiring an understanding of natural processes of various kinds, and in retrospect we often comment to the effect that it was not possible to understand a given phenomenon until such a source of explanatory concepts became available (80).

We are at the core of a great scientific dilemma: to develop technology before or after the knowledge and understanding of the speciality to which it refers. Should Medical Artificial Intelligence (AI) be developed as a high-qualified decision support for the physician or as an independent tool in medical decision making? Thusfar, the former approach have been chosen, although real understandig of methods in medicine has not been attained.

Principal factors in order to enhance rapid high-qualified decision support involve a number of fundamental limitations:

1) the need to insure substantive or input-support rationality, such that evaluation of plans and decisions are veridical;

2) the need to insure process rationality, such that the information system accomodates the capabilities of, and the constraints placed upon the user;

3) the need to understand and cope with human cognitive limitation as they affect the formulation, the analysis, and the interpretation of decision situations and alternatives, and

4) the need to understand and to integrate the normative or prescriptive components with the descriptive components of the decision situation in order to evolve realistic adjuvants for the formulation, analysis, and interpretation of decision options (1).

AI systems represent knowledge in the logical structure of If (premise), Then (action) syllogism, which is known as the Production rule model. It is claimed that this model allows the coding of general and specific medical knowledge, module structures, explanation and checking.Some authors assume it to be too rigid (53).

The inexact medical reasoning is viewed as one of the main stumbling blocks to the strict algorithmic structure of AI programs. Shortliffe and Buchanan (81) have turned to "Confirmation Theory" which is descended from the distinction of Carnap between two types of probability:"degree of confirmation" (53,82,83) and "relative frequency", which was incorporated and implemented in Mycin, one of the first AI programs. As inference rules to these probability estimates do not exist, one has again to rely on subjective estimations.

In one of the most sophisticated AI programs, CADUCEUS, the inference rules are largely based upon the introspection of one or more internists (84).The contents part was derived and updated from medical literature (pathophysiological conferences etc.). Unfortunately, because of its inductive inference rules there is no automatic learning function built into the system.The crucial notion the system employs is that of "evoking strength" (ES), relating manifestations (M) to disorder/disease (D). The various possible diseases are organised into a hierarchy generated as information about each disease. Each disease corresponds to a node in the tree and is described by a set of atomic findings, each having two associated weights: an evoking strength ES and a frequency number (FN). ES depends on the likelihood that D and M in the pertinent population are causally related, estimated by the physician on an integer scale running from 5 (high) to 0 (low). ES, therefore, depends on the availability of alternatives to D as causal explanation of M. The strength with which M evokes D will depend on whether M brought about by other causes D1,D2... Dn, that is the likelihood of M:D, (84). Or, in other words, when one encounters manifestation M how seriously would one entertain the hypothesis of D? Such a question, however, is contents dependent (85).

When some manifestations in M accompanying D are not found in the presented illness pattern CADUCEUS uses a second measure to weigh the causal strength of M to D. This measure is called FN (frequency number), ranging from 5 (all) to 0 (none) which measures the frequency in the target population with which D is accompanied by M. If FN is 5, then the anomaly is serious. FN is based on frequency counts in medical records, not on estimated likelihood. An "import number" (IN) weighs the clinical importance of each M to D, based on pathophysiological theories and knowledge. However, the physician does not consider each M separately and implicitly assign measures to it. He looks for large scale patterns of a familiar sort where the M's are interrelated in a causal way (84). McMullin mentions two inadequacies to the system:

a.) the absence of patterning ability that would allow the recognition of common manifestation-clusters (M);

b.) the inadequacy to stimulate the developmental aspect of diseases, the way they present over time (84).

In my opinion Shortliffe hits the Mark about AI systems when stating:

"They try to model their system on human cognition. The problem, of course, is that we must first learn how human cognition works". (81)

Paragraph 4

APPLICATIONS

Medical decision making would have been unthinkable without the existence of the computer. What can computers do? E.g.

a) read and translate a symbol;

b) move a symbol from one storage location to another;

c) compare two symbols and execute one program command when they are identical, but execute another program command when they are not;

d) associate two symbols, allowing access to one symbol when the other is given (86).

Some more advantages of computers are the following. They can:

1) store large quantities of data without distortion over long periods of time;

2) recall data, on receipt of the appropriate message, exactly as stored;

3) perform complex logical and mathematical operations at high speed;

4) display many diagnostic possibilities in an orderly fashion (87).

As the limitations of man as an effective problem solver has been repeatedly demonstrated (26,31,75,88), and overloading by information is assumed to be one of the important stumbling blocks to rational decision making, the computer may be a major tool for aiding physicians in their routine medical problem solving.

Kleinmuntz & Kleinmuntz (12) have raised some questions towards the application of computers in medical practice:

1) what cognitive processes underlie judgment and choice?

2) can computers be used to predict human inference?

3) what strategies are optimal for humans and machines?

4) do humans benefit from the feedback about their optimal and suboptimal rule application?

5) how are discrepancies between optimal models and human judgment to be evaluated?

Most of these questions still wait for an answer, which marks the computer-assisted medical decision programs as they stand. The results of the application of computer-aided systems in real medical practice (apart from administrative purposes) thusfar have not been very successful (89). It seems to me that after the over-enthousiastic period of the seventies, a period of reflection, but also of hesitation and lingering has entered the scene. Maybe researchers do not always reflect questions like the above-mentioned, switch the goals and the means, zealously overrun practical applications, sanctify the program more than the physician, do not reflect the specific conditions? We do not know, but when computer-aided decision making promises insight into the physician's thought processes (42) and a resulting better understanding of the human diagnostic process (87), it thusfar failed its mission.

What are the problems met in originating computer programs for medical decision making? (90)

1) <u>relevant information.</u> The problem is that we do not know the relevancy prior to the solution. A study of chest pain (91) showed that out of 498 information items under study, 55 was the maximum required for effective differential diagnosis and nearly 90% of the total information available was either redundant or irrelevant both for the description of the disease entities and the differential diagnosis. But Pipberger et al could only determine this figures afterwards, not during the diagnostic process.

2) <u>reproducibility of information, clear terminology.</u> Indeed, this is another stumbling block in medical practice, as we have discussed earlier. Not only the names of diseases but also the names of symptoms are sometimes confusing. The symptom 'dyspnea' can mean 'tightness in the chest' as well as

'afraid', 'nausea' or 'dizzy', varying from situation, time, geographical area, culture etc. Besides symptoms can be accompanied by a number of adjectives e.g. whether the 'oppression' is attackwise or continuous, day or night, with exertion or in rest etc. Computer programs require the use of vocabularies that are free of ambiguity (92). Unless medical science solves this problem, computer directed diagnostic aids will play only a small part in medical decision making.

3) comparability of terminology. Especially, classifications and taxonomies of diseases must be revised in order to develop a uniform system for unambiguous use by all physicians in the world.

4) test optimization. The paper of Zieve (93) about 'misinterpretations and abuse of laboratory tests by clinicians' outlines of the problem involved in this condition.

5) analysis of data and allocating the patients to a problem/disease class. The current taxonomic system of 'disease' is grossly unsatisfactory for both science and care in classifying the difficulties that patients present to doctors (94). By abandoning this nomenclature and allowing each problem to be expressed in its own observational terms, Weed (2) provided an intellectual liberation from the nosographic shackles imposed by the restricted scope of entities in the conventional taxonomy of "disease". But in replacing the diagnostic nomenclature of disease by a pragmatic nomenclature of problems, Weed has exchanged a standardized but inadequate taxonomy for an adequate but unstandardized taxonomy (95).

6) limitations of problem classes. Unfortunately, present medical taxonomy is more likely to expand its classes (e.g. with psychosocial disturbances) than to restrict them. In a study by Lipkin 21 haematological diseases were (conveniently) grouped into 9 classes, apparently chosen because the grouping reduced the time for the computer to process the data. This resulted in some remarkabe appositions - e.g. of uremic anemia and acute post-hemorrhagic anemia- in the same class, while megaloblastic anemia could only be found under the headings 'pernicious anemia' and 'tropical sprue' (96). This can lead to highly confusable situations.

7) prior and conditional probabilities. This matter has been discussed in former paragraphs.

8) stability of the probabilities. It is far from plausible that these probabilities remain stable. Changing views on health and disease, health care, medical research as well as progress in treatment possibilities make stable probabilities illusory.

9) quantification of utilities , and

10) clarification of outcome states .

The problems these items pose are assumed to be unsurmountable, as was discussed before.

Apart from de Dombal's list of conditions, computer programs meet some other hindrances (68):

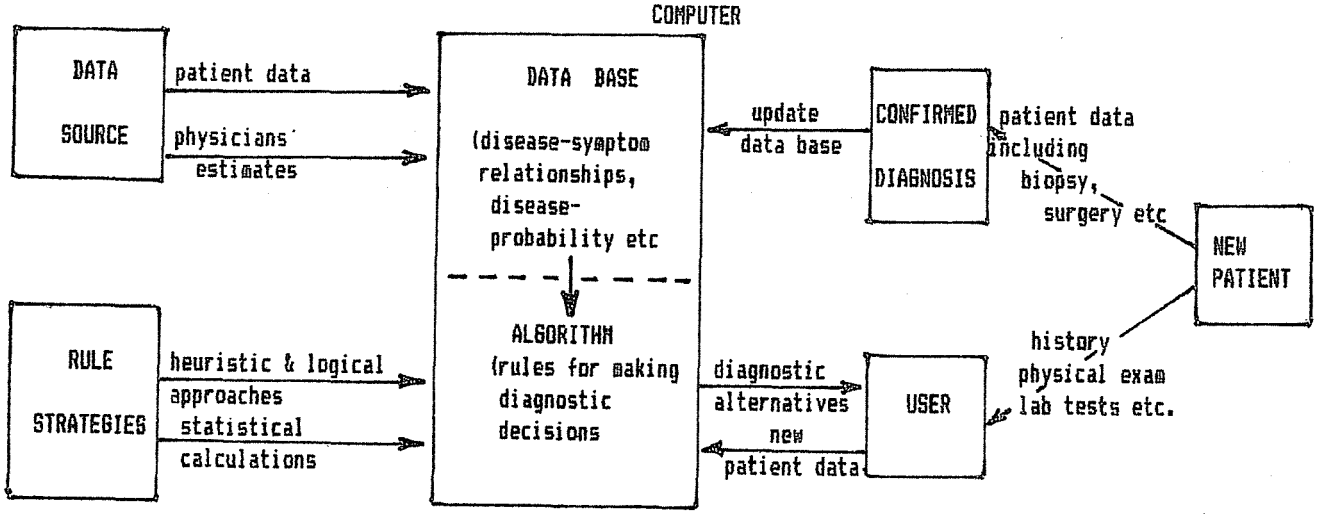
a) change of symptoms during the course of the disease. The diagnostic pattern of a disease in medical circumstances when the patient is at home, can largely differ from that of the same disease in hospital because of time lag. In chronic diseases various phases of the disease are known to exist, each phase with its own specific pattern, and, unfortunately, not always in strict order.

b) changes in prior probabilities with seasonal or epidemic conditions.

c) assumption of statistical independence on each datum.

A special limitation comes from the part of computer science. Each computer program has to make use of an algorithm. A computer algorithm is composed of the logical or statistical processes used to derive a solution to a diagnostic or therapeutic problem, from the information included in the data base and the information obtained from the new patient. However, algorithm nomenclature is inconsistent from author to author (87). The computer program, in essence, consists of a large data base and a set of rules to which the specific care of the individual patient is compared. It has been schematized by Rogers, Ryack, and Moeller (87)

SCHEMA 4



Rogers,W.,Ryack,B.,Moeller,G.(87)

The computer data base consists of

a) disease - symptom relationships;

b) disease probabilities;

c) other medical information pertinent to diagnosis and treatment of the particular disease involved (i.e. drug interactions, further diagnostic tests etc) (87).

The diseases included in the data base are summarized in three ways:

1) general disease class to which they belong (according to a classification system e.g. ICD9);

2) specific disease category or major symptoms which best describes the disease problem explored;

3) the actual number of diseases included in the system (87).

A set of rules assigns the patient's data set to the disease classes. In most computer programs all diagnoses are pursued equally and ranked strictly according to their probability. We must infer that they are given the same value or utility. However, this equal ranking of diagnoses is neither necessary nor realistic (4). Computer assisted decision aids can only ensure a certain amount of completeness, perhaps diminishing variation of interpretation, possible

diminishing emphasis differences, and maybe enhance comparability (97).

Sterling et al (cited in 53) identified three possible areas of application:

1) initial structuring of a clinical problem with little prior information; this is especially applicable for AI but which doctor wants a (very) lengthy dialogue with such a system? Lengthy because the computer must ask questions humans can see at once: in case of peritonitis a computer has to ask: "Is there an arrow piercing your abdominal wall?"

2) differential diagnosis within a restricted clinical problem area;

3) automatic interpretation of test results such as EEGs and ECGs.

This humble approach contrasts with wider scopes as "supporting the physician improving the quality of his decision making", "improving medical education", and "act as an intelligent interface to medical databases". But perhaps it is more realistic. There are presently too many uncertain elements and too many unresolved problems to proclaim a more or less comprehensive support system for physicians. No computer-aided system presently has the capability of diagnosing a large number of diverse diseases accurately (87). The systems almost always work in highly circumscribed areas. The fewer diseases and the more diseases are differentiated from each other, the higher the resulting diagnostic accuracy, whether diagnosed by physician or computer. The generally higher accuracy of the thyreoid studies is assumed to be due to the smaller number of diseases and greater distinguishability among the diseases involved (87). This accuracy is also affected by the exclusion of atypical patients in the data base, and by the way the assessment of the system is executed.

Frequently in constructing a decision aid system the more typical patients are included. This will result in an underestimation of the number of difficult cases to be expected, an overestimation of the performance of the decision aid and also in an atypicality detection mechanism which will too readily label new patients 'atypical' (physicians recall often more the 'atypical' patients over the 'typical':estimations of 'atypicality' are therefore distorted) (98). Although the use of a new test sample (evaluation sample) is fundamental to realistic assessment of a system's accuracy (Fisher et al, 1975) many studies ignore this requirement and report diagnostic accuracies testing the system on the developmental sample (87). Fleiss reported that the statistical algorithms (Bayes, Linear Discriminant Function) produces higher accuracy when tested on the developmental sample than when tested on a new sample from the same population (99). The explanation may be that since any new sample will be somewhat different, the algorithm cannot be expected to perform as well on the new sample as they do on the developmental sample.

In his review Reggia (100) distinguishes five types of computer-assisted medical decision systems:

I Conventional Programming Methods, including simple branching systems, flow charts etc, inferences by simply executing the statements in the program.

II Statistical Pattern Recognition, including

a) Bayesian approach

b) Linear Discriminant Analysis

c) Matching procedures

Alternative statistical approaches give only marginal improvement, if any.

III Production Rule System, including various Artificial Intelligence Systems.

IV Cognitive models: such programs are an explicit attempt to model the abductive reasoning of the expert human physician as it is understood today.

V Data Base comparisons.

We shall discuss some of these types of systems, being aware that only a limited survey can be provided in this chapter. Koehler et al (101) already listed a bibliography on Interactive Data Processing in Medicine: man-machine Dialogue of 854 titles. Since then the number of publications increased tremendously. Krischer (102) collected 110 studies covering 15 years of study.

It is estimated that about 25% of the studies has not been published.

The studies were executed in various categories as:

a) Health care policy evaluation.
b) Diagnosis
c) Treatment
d) Diagnostic treatment
e) Other
f) Conceptual issues.

The methods and emphasis of the studies were mainly::

1) Probability assessments or analyses: they are mostly pertained to populations rather than individuals or from estimates based on clinical data or experience. The studies revealed systematic biases in subjective estimates.

2) Single Attribute Utility applications: only rudimentary analysis of utility structure was performed. No consensus has yet been formed as to whose utilities should be included in the analysis. In most studies simplifications were introduced: the utility function is arbritrarily assumed to have a specified form and is assumed linear in terms of the outcomes considered.

3) Multiattribute Utility Applications: they focus on formulation and assessment of utility structures for situations involving potentially fatal outcomes and long term degradations in the quality of life. Restrictions being the same or worse as in single utility applications. Because of its extreme difficulty sample size is often restricted e.g., the first study on M.A.U. model based on a sample size of 15 (103).

4) Application of group utility analyses. These studies, related to group utility functions in health care settings, are limited to only analytical and conceptual issues in group preferences.

In various reviews on this subject (e.g.4,55,100,102,104,105) it becomes clear that the statistical approach and more special the Bayesian model outnumber the other types. 60% of all studies included in Rogers et al review used Bayes' theorem; up to 90% combined Bayes and L.D.F. (87). Differences between the various approaches were hardly found. Scheinok & Rinaldo (106) reported differences on diagnostic accuracy between Bayes and LDF of 1%. Birk et al (cited in 87) found 4% difference between Bayes versus matching procedure. Croft found up to 13% differences in diagnostic accuracy for 10 statistical algorithms on a same data base (67). Inconsistency between accuracy rates was more due to number and variance of the diseases than to different algorithms. Modifications of Bayes' Theorem have shown to score only slightly different from the original. Sometimes it is stated in a more general form, and hence useful in situations one would hesitate to apply the theorem because of the uncertainty as to what the true state is. Applied to a well-circumscribed set of symptoms and diseases in a stable population, Bayes formula seems to be a robust approach, and regardless of the independence assumption, will give satisfactory results. But Scheinok & Rinaldo (106) found for the application of Bayes under the assumption of independence a total correct ratio of 0.5700, while with the Bahadur extension the ratio should be reconsidered to be 0.7668.

Salamon et al (59) question the Scheinok & Rinaldo figures because this ratio, which is the number of correctly diagnosed patients divided by the number of patients, is not really the measure of efficacy of a specific model, because patients with the same profile could have different diseases. They reported that 223 out of 300 could be diagnosed correctly by any static non-randomized mathematical model.De Dombal et al.(107) showed that a computer program can even outperform senior physicians. The proportion of correct decisions was 91% for the computer and varied between 42% and 81% for the different physician groups. On the other hand Dannenberg et al (108) found in a group of 20 physicians 12 performing better than the computer.

Politser (7) raised the question whether these studies that test physician

ability to make probability judgments, may be testing the ability of physicians to represent their beliefs with numbers rather than their ability to predict events. Knill-Jones (145) found a computer program diagnostic accuracy of 71% for the 65 cases in the original test group. In 55% of the cases in which a "certain" diagnosis was reached, the accuracy was 89%. But there are suggestions that an accuracy of 85-87% is usually reached by clinicians faced with a jaundiced patient. The overall accuracy as having one of the 11 diseases for the computer program was 69% (69).

Several investigations were founded on only a small amount of symptoms.

For example, Cumberbatch & Heaps (124) using the data of Scheinok based their calculations on 11 binary valued symptoms for six upper abdominal diseases. Gustafson et al (24) used a total of 38 symptoms and test results of which 20 symptoms were mentioned more than once. Taylor et al (70) used 30 tests in three diseases.

Warner et al (38) used 50 symptoms and signs, Horrocks et al (110) 19 symptoms of which the symptom pain was subdivided into 10 aspects of this symptom for diagnosing dyspepsia, and de Dombal used 25 elements, among which 10 aspects of the symptom pain for acute abdominal diseases (111). Noticing that Pauker et al (42) estimated the total number of symptoms and signs in the field of internal medicine at about 2 million, the figures used in the computer programs seem relatively small.

To build a program based on reliable estimates, the sample size- the number of patients enclosed in the data base -, is important. Regrettably, this information is not always provided in the papers. The figures range from very small, 55 (Boyle et al. (112), 175, Horrocks (113), 300 Scheinok, (64), 1033 referred patients (Warner (38); 1991 Croft, (67) and 3431 patients, Sebag & Hall (114) .

The evaluation of the program on a test population poses new problems to comparability of studies and their validity to patient care. A wide range of population sizes are mentioned, and many more papers do not mention any size at all. A few examples are: Horrocks (115) tested the case of dyspepsia in 50 patients; for lower gastrointestinal disorders in 82 patients, Warner (38) tested in 36, Stein et al (116) in 20, Gustafson et al (24) in 200, and Croft (67) in 437 patients.

Another phenomenon is often vague: the judgmental procedure for the estimation (subjective) probabilities and utilities. Its importance was stressed by e.g. Oskamp (117) : "Their certainty about their decisions become entirely out of proportion to the actual correctness of those decisions") and Goldberg (118): ("the amount of professional training and experience of the judge does not relate to his judgmental accuracy"). Oskamp consulted 32 judges, of which 8 experienced psychologists, and Goldberg 29 psychologists, Sebag & Hall 1 physician, de Dombal et al (119) the available clinicians in the hospital and Ginsberg and Offensend (120) two physicians.

Some investigations base their system and program on the description of one or two patients / case histories: Ginsberg & Offensend (120), and Card (20) each on one case history, and Pauker & Kassirer (121) on two case histories. Sometimes the researchers base their model on former studies. Pauker & Kassirer (122) based their study on the data described by Rifkin & Hood (123) and Diamond & Forrester; Cumberbatch & Heaps (124) on material of Scheinok.

In this latter example 7 case histories were used to examplify the model; no testing was performed. Testing of the program to the performance of physicians have been reported in only a few papers, and sometimes without defining the numbers: e.g. Miller et al (125) - hospital clinicians and case discussants.-Lahay et al (126) - five (groups of) physicians. The numbers of physicians are relatively small e.g. Spicer & Lennard Jones (127): one physician, Ginsberg & Offensend (120): two physicians, Taylor et al (70): six internists, de Dombal et al (119): nine, and Gustafson & Holloway (103) twelve physicians.

There also is an unequal distribution for the type of diseases as object for the investigation. Rogers et al. (87) reviewed 54 articles: 60% of them

involved only 3 of the 17 I.C.D. (International Classification of Diseases) classes:

22% for class III:endocrine, nutritional and metabolic diseases;
17% for class V : mental disorders;
21% for class IX : diseases of the digestive system.

General availability of computer-aided medical decision programs in routine practice is also hampered by their poor transferability. Spiegelhalter & Knill-Jones (53) listed some of the problems in this respect:

1) the different clinical data may be routinely collected;

2) the observer variation in eliciting indicants (symptoms highly characteristic for a certain disease) may vary from situation to situation, which may bias the probabilistic prediction;

3) the prevalence of disease may vary from place to place, either due to genuine geographical variation (128) or to different reasons of referral to clinical centres;

4) the presentation of the disease may vary. These variations have shown to affect considerably a system based on a conditional independence.

In our opinion, computer-aided medical decision-making is still in its infancy, and perhaps not even that. It is far too early to assume a broad application of computers in a wide field of practical health care.

Nevertheless, the more constraint economics place upon health care, the more the efficiency and effectiveness of the health care system is questioned. Some authors translate this question in a need for computer assistance in medical practice. Wardle & Wardle (55) listed five items in favour of computer-aided diagnosis:

a) to improve the accuracy of diagnosis. The necessity for computer assistance is given in a number of examples of physicians' inaccuracy.

Rosenblatt found an accuracy of 80-100% for cancer, but only 44% for cancer of the lung.

Gwynne found 33.4% false negative diagnoses in a sample of 1627 patients.

Prutting found an average inaccurate diagnosis of ca 50% in various series.

In France, in a report of 100 autopsies, in 55.4% the diagnosis was missed.

33% of the vascular lesions of the C.N.S. were not detected in a series of 1478 cases studied by Kagan. Garland reported an accuracy of 44% for myocardial infarction. This figure is consistent with the investigation by Van der Does & Lubsen (129)

Moore et al (130) report that the pattern of decisions made by a nurse did not differ significantly, either statistically or clinically from that of three doctors. They recommend to substitute the general practioner for a nurse.

b) to improve the reliability of diagnosis. Every doctor has his off-days: boredom, fatigue, illness, situational and interpersonal distractions all plague him with the result that his repeated judgments of the exact same stimulus configuration are not identical (131).

c) to increase speed, to reduce costs. Taylor et al (70) report that some physicians required many more tests than either other physicians or the computer. Zieve's paper (93) lists a large number of misinterpretations and abuse of laboratory tests by physicians. Redundancy of information has been discussed world-wide.

d) to improve training in diagnostic techniques. The educational aspects of medical decision making (and additional computer programs) is still in discussion without a substantial outcome up till now.

e) research. The study of a particular set of diseases and symptoms enhances collecting and analysing data in detail required for diagnostic programs. The reasoning processes were studied for close connection of computer programs with the routine practice of the physicians.

When the statistical or actuarial model cannot provide the necessary support, perhaps different systems can supply the wanted assistance. Various

aids have been developed, some of them (flowcharts etc) have already been discussed.

Patient-filled questionnaires, with or without a computer, must enable the physician to save time and at the same time to acquire a more complete history of the patient. Examples are the Cornell Medical Index-Health Questionnaire, the Multiphasic Health Checkup, Computer-based Medical-history system (132), Patient-filled questionnaires (133), Gladys (134). Although enthousiastically reported on, as far as I know, none of these devices is now in routine practice.

Consultant-type aids: mainly there are two types:

a) Data-base systems. The computerized system collects various kinds of information, categorizes, analyses according to certain rules, and store this information in order to provide the busy physician updated knowledge and figures about a host of various diseases.

b) Expert Systems, to be distinguished into two types:

1) capturing the special knowledge of experts in the given subject matter and make this joint knowledge base accessible to physicians.

2) Using Artificial Intelligence models for the same purpose as 1) These models can make use of 'expert knowledge' but not as a necessary condition. Miller et al. (125) e.g. made use of reports of clinico-pathologic conferences as they were published in journals to model Internist-I. It differs from the former mode in the embodiment of a certain type of logic in the information supplied to the user. In this sense, it is a prescriptive device.

Data Base Systems.

Basic features are:

a) patient data are entered into the system, only some elements of these data will result in any action by the system;

b) the data base holds relevant information on patient care;

c) the data base is wholly or partially organised into decision-making models;

d) the system will suggest specific lines of action depending on the patient data;

e) the system will show the basis of its suggestions;

f) the user decides what action to take (50).

Sage (1) formulated recommendations to build a reliable data basis:

1) sample information from a broad data base and be especially careful to include data bases which might contain disconfirming information;

2) include sample size, confidence intervals, and other measures of information validity in addition to mean values;

3) encourage use of models and quantitative aids to improve upon information analysis through proper aggregation of acquired information;

4) avoid the hindsight bias by providing access to information at critical past times;

5) encourage decision making to distinguish good and bad decisions from good and bad outcomes, in order to avoid the ' illusion of validity';

6) encourage effective learning from experience. Encourage understanding of the decision situation and methods and rules used in practice to process information and make decisions such as to avoid outcome irrelevant learning systems;

7) use structured frameworks based on logical reasoning, in order to avoid confusing facts and values and wishful thinking and to assist in processing information updates;

8) both qualitative and quantitative data should be collected, and all with appropriate 'emphasis'; none should be over-or underweighted in accordance to personal beliefs, values and the like;

9) remind the type or size of the sample so as to avoid the representativeness bias;

10) present the information in several orderings so as to avoid recency and primacy effects.

And we should like to add one more item

11) define as strictly as possible the medical data to be included into the base, avoid ambiguous and inconsistent classification, and employ an unequivocal taxonomy of all medical entities.

Several types of these consultation bases, or medical information systems, are now in use.

Systems like PROMIS,COSTAR, TMIS, CARE, ARAMIS, CASNET etc are more or less successfully in use. Some of these systems provide the possibility to the physician to create and maintain a computerized medical record for every patient. It will be clear that these systems can provide very precious sources for clinical and epidemiological research. On the other hand, the discussion about the privacy and ethics around these data is in its very first phase.

The 'Expert systems'are mainly Artificial Intelligence systems, using experts' knowledge or data-base information (an A.I. system called MYCIN uses information from CASNET) and a couple of rules which lead the user through the information towards the conclusion: a diagnosis or plan of action. Yu (135) views two types of obstacles: minor and major ones, although, in our opinion, the minor are major and vice versa. Yu considers as minor problems:

a) inprecise medical terminology;

b) non-independence of clinical parameters;

c) incorrect information supplied to the computer;

d) static representation of a patient's medical history (and physical examinations and tests, J.R.).

The major problems include:

1) requirement of massive data base;

2) representation of medical knowledge which tends to be general rather then specific;

3) physician error or misinterpretation.

One of the major problems of computer-directed medical data collecting is the outsize number of questions to ask. E.g. a patient with meningitis need not be asked: "Is there a spear sticking in the middle of your head?" because even physicians do observe such a striking phenomenon, but the computer has to ask this (135). Simple, basic information gathering processes then become a tedious task. Since so many concepts and so many facts are necessary in order to come to even the simplest diagnosis representation of knowledge for such purposes tends to be global in nature. But physicians need narrower subsets of knowledge, subsets which are hardly provided in general medical textbooks. The computer programs need very specific knowledge though; but doing so it goes beyond the possibilities and limits of medical knowledge. Therefore, most existing systems are limited to a -very- restricted province of medicine and/or concerned with a limited amount of alternatives or domains. Puff e.g.is designed to interpret pulmonary function tests results; Mycin allows optimal choice among antibacterial drugs. More substantial programs are Caduceus (originally Internist II but for legal reasons renamed) and its predecessor Internist I. These concepts of Myers and Pople now covers the diagnostic potency to elucidate several hunderds of diseases in the field of internal medicine.

Advantages of these types of programs are distinct: it can store enormous amounts of factual data, easy retrievability, accessible to an almost unlimited number of users at each time and in each place, it is physician nor patient bound. But disadvantages are also obvious. There is a risk of over-reliance on computer read-outs, mechanization can lead to dehumanization, but - important for the user - it lengthens, rather than shortens, the time spent on a typical case (136). McMullin (84) in a critical review describes the pro's and con's of the Caduceus system. Recently Clancey & Shortliffe (137) gave an survey about the first decade of medical artificial intelligence.

We disagree with those people, especially in the field of informatics and computer scientists, complaining about the lack of interest physicians and the medical world show with regard to these developments. Physicians are really interested in the attempts and progress made in this field. But Medical Informatics has to realize that unless it can offer devices of real aid and assistance to the practicing doctors, physicians will remain standing in the wings.

Paragraph 5

STATE--OF-THE-ART

For more than two decades attempts have been made to use computers in clinical diagnostics. Interest in these techniques persists despite little practical success in their application. The explanation of this lack of success is complex, and, according to Taylor (138), can mainly be focused to two headings:

1) the lack of understanding of the decision making process itself, and

2) the need for a theoretical framework within which such systems can be developed.

The existence of a computer program that solves a medical problem does not prove that its method is the one physicians use to solve the same problem (139). With regard to this dilemma two explanations may be mentioned:

-the "diagnostic process" does not exist (140) and hence, every attempt to elucidate something that does not exist is unlikely to be rewarding (141). The mental processes are assumed to be so complex that attemps to make computers mimic this procedure are doomed to fail.

-the "diagnostic process" exists but physicians are unwilling to define the contents in such explicit ways to accomodate the needs of computer programs. If the physician is willing to define the medical problem in terms of actions, outcomes, probabilities and utilities, he can delegate to the program the computational task of playing out the consequences of his judgment (142).

This state of affairs hardly offers a promising view for the future. Both explanations start from unproven statements: the former conveniently negates the existence of some kind - or kinds - of a method in the physician's problem-solving behaviour, the latter blames the doctor. We agree with Taylor (138) stating that the emphasis must be laid upon decisions rather than data and on people rather than statistics. The choice of appropriate statistical or computing techniques will become apparent during the analysis. In the preceding decades emphasis was mainly laid upon these techniques.

The question became: "to what extent has Bayesian diagnosis begun to replace conventional diagnosis?" (45) Feinstein (45) notices that almost all the publications have been isolated proposals. The authors have not supplied any follow-up publications. "I know", Feinstein says, "of no published work, or clinical setting, or specific world situation, in which Bayesian methods have made a prominent contribution that could not have been achieved just as easily without the Bayes formula. Bayesian inference rests on the idea that the posterior probability of an event is proportional to the product of its prior probability multiplied by the likelihood ratio. But the Bayesian statistician does not know the prior probability. He must make a subjective, personal guess called an estimate. When Bayesian methods depend on intuitions and hunches, we only replace clinical by statistical hunches which does not bring us any step farther".

Physician's guesses are assumed to be based on their knowledge of the prevalence of diseases and symptoms and their interrelationship in a finite population. The latter being essential, because if we do not know N, P(robability) is undefined and can picture all sorts of numerical values ranging from 0-1. Quantitative estimates about events with which we are not directly familiar are notoriously unreliable because they are frequently based on a superficial image (143). Another matter is that the ailment nor the symptoms are clearly defined and cannot be assumed as a clearcut entity. It is not the probability the physician has in mind but the spectrum of people with the presumed specific disease under observation. While Bayesian logic is aimed at specific diseases and actions, it deviates from normal routine medical work which deals with unspecified ailments and uncertain decisions.

Studies have also been, for the most part, confined to the well-structured areas of medicine. However, the bulk of decisions in medical practice are much

less clearly defined (138). The general opinion appears to be that the statistical approach is often too simplistic for realistic clinical problems, inapplicable because there are insufficient data, and incomprehensible to the user (53). Leaper et al (140) warn that studies based on artificial situations should be treated with extreme caution. Statistical modelling of conditional distributions of variables given diseases has been criticized for its over-simplicity, particularly for the frequent assumptions of conditional independence of variables and the restriction to mutually exclusive and exhaustive diseases (53).

Croft (67) argues that no substantial improvement in computer diagnosis is possible until clinical profiles of major diseases are more accurately defined. But modern medical research emphasises identification of mechanisms and specific therapies rather than determination of clinical profiles based on patient history and physical examination. Gorry (79) stressed the need for formal representation of medical concepts and for a dialogue which would include explanations in terms that a doctor could understand. The first question, however, should be whether the decision theoretical approach reflects the medical problem-solving process as it is practiced in routine health care delivery. If it proves to be an artefact the claim for being an aid or assistance to the physician must be dropped.

This does not mean that decision theoretical approaches not have an impact on and value to health care delivery, but as a separate discipline next to clinical medicine and not as a part of it. Computer-based processing systems may be valuable, but several investigators still believe that these technologies are unadapted to the medical decision-making process, that they are not applicable to complete and dynamic clinical problems (7). Recent reviews have shown numerous medical applications of decision analysis, but most of them involved hypothetical situations (4,7,55,102).

Friedman & Gustafson (89) discuss the failure of such computer applications in clinical medicine, and characterize the systems as having been inflexible, with a poor interface with the user and of no readily perceivable benefit to doctors or patients. Thusfar, the running programs lengthen rather than shortens the time spent on a typical case. Analysis leads to the conclusion that the more extensive the utility assessments, the more extensive the use of drugs(102), opposing the expectations of optimizing choice behaviour. There is little to show for the enormous amount of work invested in this area.

It must be clear that physicians will reject systems that dogmatically offer advice without proven benefit in allround situations even if it displays impressive diagnostic accuracy (50). When Schwartz (144) complains about the negative reactions of many students and physicians in trying to bring decision analysis into clinical use, this aforementioned dogmatism can alienate physicians and students from this approach. The physician has to realize that even a highly educated adult relies upon intuitive considerations rather than the appropriate abstract principles. In the absence of the appropriate code, individuals rely upon their own experience, derived from familiar content, in making an inference or judgment (143).

Impediments to a successful computer-physician and/or patient-computer communication have been listed in the publication of Friedman & Gustafson (89):

1) many computer(s) (terminals) have been poorly engineered, resulting in mechanical breakdowns. Not only the construction but the 'bugs' in the program can drive the user crazy. Meanwhile the former impediment seems obsolete.

2) the computer(s) (terminals) have often been placed in out-of-the-way places making them inconvenient to operate and useless for rapid data retrieval.

3) the computer response-time often has been quite slow because of low-speed teletype output or excessive delay between responses. About extending response-time Warner reports another impediment: the overload of the system by users' requests lead the computer to spend almost all its cycles controlling traffic and swapping programs and buffers back and forth to disc (145). Although

since 1977 equipment has been more and more sophisticated and memory-storage multiplied, the innumerable number of data used in medical practice still makes the use of huge main-frames necessary.

4) in order to obtain information from the computer the physician has often been required to take part in a long and complicated technical dialogue.

As programs almost always differ to a large extent from each other, and, therefore, the sign-on codes vary from program to program, and programs are not transferable, we are far from an easy handling of the machinery.

5) the use of computers often requires knowledge of special passwords, codes, or computer languages.

6) computer terminals have been expensive and this made it difficult to develop accessible yet cost-effective applications. This impediment has by now been solved (almost).

The most important impediment is perhaps, that although we have been able to demonstrate the reliability and validity of information collected by computer-based patient interfaces, we have not demonstrated that providers can make better or less costly decisions because of it. Remarkably, where in several other fields the computer performs tasks previously incomprehensible to mankind, in the field of health care one is quite satisfied when the machine can duplicate the physician's task. Setting the goal at such a low level, it is hardly astonishing to find a cool reception.

Successful innovations are characterized by user identification of the problems. Computer research is characterized by computer scientists' working on a solution to health problems without drawing sufficiently on the advice of actual providers or potential users. Computer scientists are busy searching for a problem to fit their solution. In other words, the existing system must adapt to the program rather than the other way round (89).

Friedman & Gustafson signalize six obstacles to computer assistance in medical care:

1) patient-computer and/or physician-computer interaction;
2) provision with equipment that exceeds the physician's capability;
3) inability to prove a significant positive impact on patient care;
4) difficult transferable system from one institution to another;
5) research has thusfar not been change-oriented;
6) man has not learned from previous mistakes.

In our opinion, all questions, problems and pitfalls cannot be solved in the near future. Indeed, applications are decreasing:

the Index Medicus cited 30 papers under Diagnosis: computer-assisted in 1982 compared with 83 in 1977 (53). It is essential to both disciplines, medical and computer science, to strive for one line of research, one common view on the medical process, one language. Both disciplines have to analyze their methods and content in a rational and objective way. From the medical part we may agree with Taylor, (138), when he states: "It is, therefore, more promising to begin projects of this kind with an analysis of the decisions made by physicians in the appropriate area of the health care system so that from the beginning the proposed system will fit as closely as possible to the needs of the existing system and to the physicians who will use it".

REFERENCES CHAPTER V

1. Sage A.P.,: Behavioral and organizational considerations in the design of information systems and processes for planning and decision support. IEEE transactions and systems, man and cybernetics, 1981, 11, 640 - 678.
2. Weed L.L.,: Medical records, medical education and patient care. The problem oriented record as a basis tool. Chicago, Year Book Medic. Pub. Inc. 1970.
3. Patient Care. Patient Care Flow Charts, Miller & Fink Corp., Darien, 1976.
4. Albert D.A., :Decision theory in medicine. Milbank Memorial Fund Quarterly / Health and Soc. 1978, 56, 362 - 401.
5. Mischel W.,: On the interface of cognition and personality : beyond the person-situation debate. Psychologist, 1979, 34, 740 - 754.
6. Mason R.O., Mitroff I.I.,: A program for research on management information systems. Management Science, 1973, 19, 475 - 485.
7. Politzer P.: Decision analysis and clinical judgment. Med. Dec. Mak., 1981, 361 - 389.
8. Ledley R.S., Lusted L.B.,: Reasoning foundations of medical diagnosis. Symbolic logic, probability and value theory aid our understanding of how physicians reason. Science 1979, 130, 9 - 21
9. Dombal de F.T., Gremy F.,: Decision making and medical care. Can information science help? Amsterdam, North Holland, 1976.
10. Slovic P. Fischhoff B., Lichtenstein S.,: Behavioral decision theory. Ann. Rev. Psych., 1977, 28, 1 - 29.
11. Robbins,S.P.:The Administrative Process. Englewood Cliffs N.J.,Prentice Hall, 1980.
11a. Simon, H.A.: Administrative Behavior. A study of decision making processes in administrative organization. New York, Free Press, 1976.
12. Kleinmuntz D.N., Kleinmuntz B.,: System simulation, Decision strategies in simulated environments. Beh. Science, 1981, 26, 294 - 305.
13. Crean G.P.,: Clinical decision making: an introduction. Health Bull., 1977, 205 - 206.
14. Yerushalmy J., Palmer C.E.,: On the methodology of the investigation of etiologic factors in chronic diseases. J. Chron. Disease, 1959, 10, 27 - 40.
15. Lusted L.B.,: Decision making studies in patient management. N. Eng. J. Med., 1971, 8, 416 - 423.
16. Vecchio T.J.,: Predictive value of a single diagnosis test in unselected populations. N. Eng. J. Eng. Med., 1966, 274, 1171 - 1173.
17. Wulff H.R.,: Rational Diagnosis and Treatment. Blackwell Scientific Publ., Oxford, Edinburgh, Melbourne, 1976.18.
18. Card W.I.,: Mathematical method in diagnosis. J. Roy, Coll. Phycns., London, 1975, 9, 193 - 196.
19. Betaque D.E., Gorry G.A., : Autonating judgmental decision making for a serious medical problem. Management Science, 1971, 17, 421 - 434.
20. Card W.I., Ruscinkiewicz M., Philips C.I., : Utility estimation of a set of states of health. Meth. Inf. Med. 1977, 16, 168 - 175.
21. Lindley D.V.,: The role of utility in decision making. J. Roy. Coll. Physcns. London, 1975, 9, 225 - 230.
22. Weinstein M.C., Fineberg H.V.,: Clinical Decision Analysis. Phyladelphia, W.B. Saunders, 1980.
23. Dowling A.F.,: Medically oriented computer based information systems. Med. Care, 1982, 20, 253 - 254.
23a. Balla,J.I.: The use of critical cues and prior probability in decision making. Meth.Inf.Med.,1982, 21, 9 - 14.
24. Gustafson D.H., Kestly J.J., Greist J.H., Jensen N.M.,: Initial evaluation of a subjective Bayesian diagnostic system. Fall, 1971, 204 - 213.

25. Kochen M.: How clinicians recall experiences. Meth. Inf. Med. 1983, 22, 83 - 86.
26. Kahnemann D., Tversky A.,: On the psychology of prediction, Psych. Review, 1973, 80, 237 - 251.
27. Kahnemann D., Tversky.: Subjective probability: a judgment of representativeness. Cogn. Psych., 1972, 3, 430 - 454.
28. Tversky A., Kahnemann D.,: Judgment under uncertainty: Heuristics and biases. Biases in judgment reveal some heuristics of thinking under uncertainty. Science, 1974, 185, 1124 - 1131.
29. Tversky A., Kahnemann D.,: Availability: a heuristic for judging frequency and probability. Cign. Psych., 1973, 5, 207 - 232.
30. Eraker S.A., Politser P., Arbor A.,: How decisions are reached: Physician and patient. Ann. Int. Med. 1982, 97, 262 - 268.
31. Fischhoff B., Slovic P., Lichtenstein S.,: Knowing with certainty: the appropriateness od extreme confidence. J. Exp. Psych., Hum. Perc. Perform, 1977, 3, 552 - 564.
32. Kahnemann D., Tversky A.,: Prospect Theory: An analysis of discussion under risk. Econometrica, 1979, 47, 263 - 291.
33. Tversky A., Kahnemann D.,: The framing of decisions and the psychology of choice. Science, 1981, 211, 453 - 458.
34. Lindblom C.E.,: The Policy-making Process, Englewood Cliffs, N.J., Prentice Hall, 1980.
35. Lindblom C.E., The science of "muddling through". Public Administr. Review, 1959, 19, 155 - 169.
36. Lichtenstein S., Slovic P., Fischhoff B., Layman M., Combs B.,: Judged frequency of lethal events. J. Exp. Psych., 1978, 4, 551 - 578.
37. Wagner G., Tautu P., Wolber U.,: Problems of medical diagnosis. A. Bibliography. Meth. Inf. Med. 1978, 17, 55 - 74.
38. Warner H.R., Toronto A.F., Veassy L.G., Stephenson R.A.,: Mathematical approach to medical diagnosis. JAMA, 1961, 177, 177 - 183.
39. Nash F.A.,: The application of computers to medical and veterinary science. Proc. Roy. Soc. Med. 1963, 56, 555 - 560.
40. Gorry G.A., Barnett G.O.,: Experience with a model of sequential diagnosis. Comput. Biomed. Res., 1968, 1, 490 - 507.
41. Edwards W., Phillips L.D., Hays W.L., Goodman P.C., : Probabilistic information processing systems: design and evaluation. Tr. systems Sc. and Cyber, 1968, 4, 248 - 265..
42. Pauker S.G., Gorry G.A., Kassirer J.P., Schwartz W.B.,: Towards the stimulation of clinical cognition. Taking a present illness by computer. Am. J. Med., 1976, 60, 981 - 996.
43. Edwards W., Lindman H., Savage L.J., Bayesian statistical inference for psychological research, Psych. Review, 1963, 70, 193 - 242.
44. Ranschoff D.F., Feinstein A.R.,: Is decision analysis useful in clinical medicine? Yale J. Biomed., 1976, 49, 165 - 168.
45. Feinstein A.R.,: The haze of Bayes, the aerial palaces of decisionanalysis and the computerized Ouija Board. Clin. Pharm. Ther., 1977, 21, 482 - 496.
46. Kassirer J.P.,: The principles of clinical decision making: an introduction to decision analysis. Yale J. Biol. Med. 1976, 49, 149 - 164.
47. Emerson P.A.,: Decision theory in the prevention of thrombo-embolism after myocardial infarction. J. Roy. Coll. Phcns., Lond. 1975, 9, 238 - 251.
48. Chalmers T.C.,: A shortage of reliable data. N. Engl. J. Med. 1976, 294, 721 - 722.
49. Williams B.T.,: Computer aids to clinical decisions. Vol I & II, Florida, CRC Press, 1982.
50. Young D.W.,: A survey of decision aids for clinicians. Br. Med. J. 1982, 285, 1332 - 1336.
51. Edwards D.A.W.,: Diagnostic procedures. History and sypmtoms of

esophageal disease, in: Vantrappen G., Hellemans J.,: Diseases of the Esophagus. Berlin, Springer, 1974.
52. Essex B.J., Approach to rapid problem solving in clinical medicine. Br. Med. J., 1975, 3, 34 - 36.
53. Spiegelhalter D.J., Knill-Jones R.P.,: Statistical knowledge based approaches to clinical decision support system with an application in Gastroenterology. J. Roy, Statist. Soc. 1984, 147, 35 - 77.
54. Card W.I.,: The Diagnostic Process. J. Roy, Coll. Physcns. London, 1970, 4, 183 - 187.
55. Wardle A., Wardle L.,: Computer aided diagnosis - a review of research. Meth. Inf. Med., 1978, 17, 15 - 27.
56. Bayes T.,: An essay towards solving a problem in the doctrine of chances. Studies in the history of probability and statistics. Phil.Trans. Roy. Soc., 1763, 53, 370 - 418. Reprinted in Biometrika, 1958, 45, 293 - 315.
57. Tautu P., Wagner G.,: The process of medical diagnosis: routes of mathemetical investigation. Meth. Inf. Med. 1978, 17, 1 - 9.
58. Baron D.N., Fraser P.M.,: The digital computer in the classification and diagnosis of diseases. Lancet, 1965, ii 1066 - 1069
59. Salamon R., Bernadet M., Samson M., Derouesne C., Gremy F., :Bayesian method applied to decision making in neurology-methodological considerations. Meth. Inf. Med. 1976, 15, 174 - 179.
60. Pauker S.G., Szolovits P.,: Analyzing and simulating. Taking the history of the present illness: Context information. Comp. Ling. Med., 1977, 109 - 117.
61. Anderson J.A., Boyle J.A.,: Computer Diagnosis, Statistical aspects. Brit. Med. Bull., 1968, 24, 230.
62. Norusis M.J., Jacquez J.A.,: Diagnosis. I Symptom Nonindependence in Mathematical Models for Diagnosis. Comp. Biomed. Res. 1975, 8 156 - 172.
63. Norusis M.J., Jacquez J.A.,: Diagnosis II Diagnostic Models based on attribute clusters: a Proposal and Comparisons, Comp. Biomed, Res. 1975, 8, 173 - 188.
64. Scheinok P., : Symptoms diagnosis: Bayes's theorem and Bahadur's distribution. Int. J. Biomed. Comp., 1972, 3, 17 - 28.
65. Jacquez J.A., Norusis M.J.,: The importance of symptom non-indepence in diagnosis, in: Decision making and medical care. Can information science help? Amsterdam, N.Holland Pub Co. 1976.
66. Knill-Jones R.P.,: Clinical decision making (2) diagnostic and prognostic inference. Health Bull., 1977, 213 - 222.
67. Croft D.J., Is computerized diagnosis possible? Comp. and Biomed. Research, 1972, 5, 351 - 367.
68. Croft D.J., Machol R.E., : Mathematical methods in medical diagnosis. Ann. Biomed. Eng., 1974, 2, 69 - 89.
69. Knill-Jones R.P.,: Stern R.B., Girmes D.H., Maxwell J.D., Thompson R.P.H., Williams R., Use of sequential Bayesian model in diagnosis of jaundice by computer. Br. Med. J. 1973, 1, 530 - 533.
70. Taylor T.R., Aitchison J., McGirr E.M.: Doctors as decision makers: a Computer assisted study of diagnosis as a cognitive skill. Br. Med. J. 1971, 3, 35 - 40.
71. Adlassnig, K.P., A fuzzy logical model of computer assisted medical diagnosis. Meth. Inf. Med., 1980, 19, 141 - 148.
72. Esogbue A.D., Elder C., : Fuzzy sets and the modelling of physician decision processes, part I. Fuzzy sets and systems, 1979, 2, 280 - 291.
73. Esogbue A.O., Elder C.,: Fuzzy sets and the modelling of physician decision processes, part.II: Fuzzy diagnosis decision models. Fuzzy sets and systems, 1980, 3, 1 - 9.
74. Sanchez E.,: Inverse of fuzzy relations. Application to possibility distributions and medical diagnosis. Fuzzy sets and systems, 1979, 2,

75 - 86.
75. Newell A., Simon H.A.,: Human problem solving. Englewood Cliffs, New Yersey, Prentice Hall, 1972,
76. Kleinmuntz B.: The processing of clinical information by man and machine, in: Formal representation of human judgment:pages 149 - 186. New York, John Wiley & Sons, 1968.
77. Wortman P.M., Medical diagnosis: an information processing approach. Comp. Biomed. Research, 1972, 5, 315 - 328.
78. Wortman P.M.,: Cognitive utilization of probabilistic cues, Behav. Science, 1970, 15, 329 - 336.
79. Gorry G.A.,: Computer assisted clinical decision making. Meth. Inf. Med. 1973, 12, 45 - 51.
80. Gedye J.L.,: Simulating clinical judgment: an essay in technological psychology, in Engelhardt H.T., Spicker S.F., Towers B.: Clinical judgment: a critical appraisal. Dordrecht, D. Reidel Publ. Co., 1979.
81. Shortliffe E.H., Buchanan B.G., Feigenbaum E.A.,: Knowledge engineering for medical decision making: a review of computer-based clinical decision aids. Proceedings IEEE, 1979, 67, 1207 - 1224.
82. Sadegh-Zadeh K.,: Subjektive Wahrscheinlichkeit und Diagnose. Meth. Inf. Med., 1974, 13, 97 - 102.
83. Braithwaite R.B., Scientific explanation. A study of the function of theory, probability and law in science. Cambridge, Cambridge University Press, 1968.
84. McMullin E.,: Diagnosis by computer, J. Med. Phil., 1983, 8, 5 - 27.
85. Blois M.S., Conceptional issues in computer aided diagnosis and the hierarchical nature of medical knowledge. J. Med. Phil., 1983, 8, 29 - 50.
86. Barr D.M.,: Primary and specialized care: implications for learning situations. J. Med.. Educ., 1978, 53, 176 - 185.
87. Rogers W., Ryack B., Moeller G.,: Computer aided medical diagnosis: literature review. Int. J. Biomed. Comp., 1979, 10, 267 - 289.
88. Janis I.L., Mannn L.,: Emergency decision making: theoretical analysis of responses to desaster warnings. J. Human Stress, 1977 35 - 48.
89. Friedman R.B., Gustafson D.H.,: Computers in clinical medicine, a critical review. Comp. Biomed. Res., 1977, 10, 199 - 204.
90. Dombal de F.T., Medical diagnosis from a clinician's point of view. Meth. Inf. Med. 1978, 17, 28 - 35.
91. Pipberger H.V., Klingeman J.D., Cosma J.,: Computer evaluation of statistical properties of clinical information in the differential diagnosis of chest pain. Meth. Inf. Med. 1968, 7, 79 - 92.
92. Dudley H.A.F., Pay-off, heuristics and pattern recognition in the diagnostic process. Lancet, 1968, 723 - 726.
93. Zieve L.,: Misinterpretations and abuse of laboratory tests by clinicians. Ann. N.Y. Academy of Sciences. 563 - 572.
94. Feinstein A.R.: Clinical judgment. Baltimore, Williams & Wilkins, 1967.
95. Feinstein A.R.,: The problem of the 'problem-oriented medical record'. Ann. Intern. Med., 1973, 78, 751 - 762.
96. Scadding J.G.: Diagnosis: the clinician and the computer. Lancet, 1967, 877 - 882.
97. Anderson G., Lerena C., Davidson D., Taylor T.R.,: Practical application of computer assisted decision making in an antenatal clinic feasibility study, Meth. Inf. Med., 1976, 15, 224 - 229.
98. Habbema, J.D.F.,Hilden, J.: Some pitfalls and difficulties in evaluating aids in clinical practice.
99. Fleiss J.L., Spitzer R.L., Cohen J., Endicott J.,: Three computer diagnosis methods compared Arch. Gen. Psychiatry, 1972, 27, 643. - ...
100. Reggia J.A.,: Computer assisted medical decision making: a critical review. Ann. Biomed. Eng., 1981, 9, 605 - 619.
101. Koehler C.O., Wagner G., Wolber U., : Interactive Data Processing in Medicine, Man-Medicine Dialogue. (Bibliography) Meth. Inf. Med. 1976,

15, 102 -120.
102. Krischer J.P., An annotated bibliography of decision analysis applicationns to health care. Operations Research, 1980, 28, 97 - 113.
103. Gustafson D.H., Holloway D.C.,: A decision theory approach to measuring severity in illness. Health Serv. Res. 1975, 97 - 106.
104. Mai N., Hachmann E.,: Anwendung des Bayes-Theorems in der medizinischen Diagnostik - eine Literaturuebersicht, Metamed, 1977, 1, 161 - 205.
105. see 87.
106. Scheinok P. Rinaldo J.A.,: Symptom diagnosis: The use of two different mathematical models. Int. J. Biomed. Comp., 1971, 2, 239 - 248.
107. Dombal de F.T., Leaper D.J., Staniland J.R., McCann A.P., Horrocks J.C.,: Computer aided diagnosis of acute abdominal pain. Br. Br. Med. J., 1972, 2, 9 - 13.
108. Dannenberg A.L., Shapiro A.R., Fries J.Rr.: Enhancement of clinical predictive ability by computer consultation. Meth. Inf. Med. 1979, 18, 10 - 14.
109. Knill-Jones R.P.,: The diagnosis of jaundice by the computation of probabilities. J. Roy, Coll. Physcns. Lond., 1975, 205 - 210.
110. Horrocks J.C., Dombal de F.T.,: Diagnosis of Dyspepsia from data collected by a physician's assistent. Br. Med. J., 1975, 3, 421 - 423.
111. Dombal de F.T., Horrocks J.C., Walmsley G., Wilson P.D.,: Computer aided diagnosis and decision making in the acute abdomen, J. Roy. Coll. Physcns. Lond., 1975, 9, 211 - 218.
112. Boyle J.A., Greig W.R., Franklin D., Harden R. McG., Buchanan W.W., McGir E.M.,: Construction of a model for computer assisted diagnosis: application to the problem of non-toxic goitre. Quart. J. Med., 1966, 35, 565 - 568.
113. Horrocks J.C., Dombal de F.T.,: Computer aided diagnosis of Dyspepsia. Am. J. Dig. Dis., 1975, 20, 297 - 406.
114. Sebag J. Hall P.,: Decision mmaking in clinical practice: Application of predictors, indicators and indices to the medical history obtained by a self-administered questionnaire. Int. J. Biomed. Comp. 1975, 6, 167 - 179.
115. Horrocks J.C., McCann A.P., Staniland J.R., Leaper D.J., Dombal F.T. de,: Computer aided diagnosis: description of an adaptable system and operational experience with 2.034 cases. Br. Med. J. 1972, 2, 5 - 9.
116. Stern R.B., Knill-Jones R.P., Williams R.,: Use of computer program for diagnosing jaundice in district hospitals and specialized liver unit. Br. Med. J. 1975, 2, 659 - 662.
117. Oskamp S.,: Over-confidence in case-study judgments. J. Consult. Psychol., 1965, 29, 261 - 265.
118. Goldberg L.R., : Simple Models or simple processes? American Psychologist, 1968, 23, 483 - 496.
119. Dombal de F.T., Horrocks J.C., Staniland J.R., Gill P.W., Simulation of clinical diagnosis: a comparitive study. Br. Med. J., 1971, 2, 575 - 577.
120. Ginsberg A.S., Offensend F.L.,: An application of decision theory to a medical diagnosis - treatment problem. Trans. sys. science cybern. 1968, 4, 355 - 362.
121. Pauker S.G., Kassirer J.P.,: The threshold approach to clinical decision making. N. Eng. J.Med., 1980, 302, 1109 - 1113.
122. Pauker S.G. Kassirer J.P.,: Clinical decision analysis by personal computer. Arch. Intern. Med. 1981, 141, 1831 - 1837.
123. Rifkin R.D., Hood W.B.,: Bayesian analysis of electrocardiographic exercise stress-testing. N. Eng. J. Med., 1977, 297, 681 - 686.
124. Cumberbatch J., Heaps H.S., : A disease conscious for sequential diagnosis by use of disease probabilities without assumption of symptom independence. Int. J. Biomed. Comp. 1976, 7, 61 - 78.
125. Miller R.A., Pople H.E., Myers J.D.,: Internist - 1, an experimental

computer based diagnostic consultant for general internal medicine. N.Eng. J. Med., 1982, 307, 468 - 476.
126. Lahaye D., Roosels D., Viaene J.,: The value of subjective appreciation in the medical record. I. The evaluation of dyspnoea. Meth. Inf. Med., 1978, 17, 100 - 103.
127. Spicer,C.C.,Lennard-Jones,J.E.: The prediction of success or failure of medical treatment for acute colitis. J.Roy. Coll. Physcns. London, 1975, 9, 252 - 256.
128. Dombal de F.T., Staniland J.R., Clamp S.E.,: Geographical variation in disease presentation. Med. Dec. Mak. 1981, 1, 59 - 69.
129. Does, v.d.E., Lubsen J.: Imminent Myocardial Infarction Rotterdam. Dissertation, Rotterdam, 1978.
130. Moore M.F., Aitchison J., Parker L.S., Taylor T.R.,: Use of information in thyrotoxicosis treatment allocation. Meth. Inf. Med. 1974, 13, 88 - 92.
131. Goldberg L.R., Man versus model of man: A rational, plus some evidence, for a method of improving on clinical inferences. Psych. Bull, 1970, 73, 422 - 432.
132. Slack W.V., Hicks G.P., Reed C.E., Van Cura L.J.,: A computer based medical history system. N. Eng. J. Med., 1966, 274, 194 - 198.
133. Sheldon M.G.,: Patient filled questionnaires. Update, 1976, 15, 651 - 658.
134. Knill-Jones R., Dunwoodie W.,: A computer assisted diagnostic system for dyspepsia. in: Decision making in the general practice. London, McMillan, 1985.
135. Yu V.L.,: Conceptual obstacles in computerized medical diagnosis. J. Med. Phil., 1983, 8, 67 - 75.
136. McMullin E.,: Introduction. J. Med. Phil. 1983, 8, 3 - 4.
137. Clancey W.J., Shortliffe .E.H.,: Reading in Medical Artificil Intelligence: first decade. Amsterdam, Addison-Wesley, 1984.
138. Taylor T.R.,: Clinical decision analysis. Meth. Inf. Med. 1976, 15, 216 - 224.
139. Palva J.P.,: Measuring clinical problem solving. Br. J. Med. Educ. 1974, 8, 52 - 56.
140. Leaper D.J., Gill P. W., Staniland J.R., Horrocks J.C., Dombal F.T. de: Clinical diagnostic process: an analysis. Br. Med. J. 1973, 3, 569 - 574.
141. Scriven M.,: Clinical judgment. In: Engelhardt H.T., Spicker S.F., Towers B.,: Clinical judgment, a critical appraisal. D. Reidel Publ. Co. Dordr. 1979.
142. Kozielecki J.,: A model for diagnostic problem solving. Acta Psych., 1972, 370 - 380.
143. Connella J.S., Goran M.J., Williamson J.W., Cotsonas N.J.,: Evaluation of patient care. An approach, JAMA, 1970, 214, 2040 - 2043.
144. Schwartz W.B.,: Sounding board. Decision analysis. A look at the chief complaints. N. Eng. J. Med., 1979, 300, 557 - 559.
145. Warner H.R., Haug P.,: Medical Data acquisition using an intelligent machine. In: Bemmel van J.H., Ball M.J., Wigertz O.,: Medinfo 83, Amsterdam, N. Holland 1983.

CHAPTER VI

MODELS AND METHODS

Paragraph 1

Models in Medical Problem Solving

When one asks a physician "How do you make a medical diagnosis?" his explanation of the process might be as follows. "First, I obtain the case facts from the patient's history, then I perform physical examination, and laboratory tests. Second, I evaluate the relative importance of the different signs and symptoms. Some of the data may be of first-order importance and other data of less importance. Third, to make a differential diagnosis I list all the diseases which the specific case can reasonably resemble. Then I exclude one disease after another from the list until it becomes apparent what the case can be" (1). We do not know which physician Ledley & Lusted asked the question, but it is hardly representative for the physicians as a group. We think Sober (2) is right in stating that "the clinician's description of his clinical diagnostics is no more than rationalization. It is false to the facts of his own psychological processes." It also contradicts the observations of Leaper et al. (3) who conclude:

I : The diagnostic process - viewed as a monolithic structure - does not in fact exist.

II : Each clinician has his own pathway to diagnosis.

III: We are in no position to draw up algorithms for different individuals.

It is the impression that we operate from rough guidelines and these cannot be adequately formalized either as statistical or as exact generalizations (4). A survey studying diagnostic pathways in family practice was described by Hull (5). He found a.o.:

1) A great variation between methods of examination between doctors;

2) The fact that previous knowledge of the patient by the doctor influences the doctor's decision how to examine his patients;

3) and that presenting symptom influences the doctor's decision to examine. Also the energy with which a doctor pursues a diagnosis in family practice depends chiefly on the patient's presenting symptom (6).

According to Howie (7) an important barrier is formed by language. The traditional diagnostic labels are no basis for meaningful discussions between consultants and general practitioners or even among general practitioners. This dilemma has also been mentioned by Scadding (8,9) and King (10). The fundamental problem is the lack of an acceptable formal definition of the diagnostic process (11). Without agreement on this point, there are no criteria whether there is such a thing like 'the diagnostic process'. To get new incentives in medicine we need to first search for and then analyse the individual work-routines. The rationalization behind this approach is that physicians, either consciously or unconsciously, make similar judgments daily; they are more or less faced with the same problems; they have practically all equal therapeutic possibilities available; and they find and understand each other throughout the world on matters of problem-solving and practice-burdens. That means that physicians can recognize common elements in their processes, that they operate along rough general guidelines distinctive to them all. But between recognisable factors and (a) formalized model(s) there is a large gap. Pauker et al (12) picture the difference between the expert in practice and the expert as often pictured in literature or folklore. The epitome of the expert in fiction is the detective who, through superior deductive powers and by sheer force of logic, organizes the facts at hand in a way that leads to a single, inevitable conclusion. By contrast, the real-world physician seem to rely much more heavily upon

"guessing", his initial hypothesis typically being based on precious little data. These "guesses" are apparently prompted by patterns of clinical findings or by specific complaints which bring to mind particular diseases. The physician then tries to demonstrate the correctness of his "guesses" moving to new hypotheses only if his initial impressions prove untenable. Apparently, judgments seem to be made by the degree to which the displayed features presented by the patient appear representative of the stereotype the physician has in mind. The more numerable and the more specific the stereotypes the physician has in mind the more he has a 'clinical view', 'clinical intuition', 'flair clinique'. But this theme only acknowledges the existence of pattern recognition as one of the elements in the diagnostic process, not its procedure. On this part models remain vague, informal.

Taylor's model

1) the physician acquires information from the patient;

2) he interprets the information, performs some kind of analysis upon it, and

3) he must make a decision concerning the individual patient(13).

But this model does not support hypotheses that allow specific predictions as required in formal theories or models. Formal models are more easily tested than informal ones. Hence their inadequacies are more apt to be exposed. The relative lack of ambiguity in a formal model encourages people to test it. The major question remaining is whether such a model can be formulated (14). The use of models in studying medical problem-solving has been scarce. The motivation for a model may come from several sources. A model may isolate and illuminate certain relationships or properties of the modeled system and hence promote an improved understanding of that system. Also, manipulation of the model may be easier than experimentation with the modelled system (14).

However, according to Gorry, there are two difficulties in modeling the medical process:

a) many medical people seem to feel that a model of diagnosis must be complete, and hence so complex as to be infeasible.

b) the potential advantages of modeling as an activity are not generally recognized.

According to Pauker et al. (12), the development of the model will require the efforts of physicians experienced in diagnosis. To a certain extent, they should consider concepts developed in other fields including cybernetics, cognitive theory, utility theory, and computer science. Regrettably, they do not make clear which elements from these disciplines are applicable to the modeling of the clinical process. Obviously, their preconception is that this process is a probabilistic one. Doctors tend to be deterministic rather than probabilistic though (3). People prefer to assume that there is a rule, rather than that there is no rule; that this rule is deterministic rather than probabilistic; that the values to be predicted from the cue values do in fact depend on these cue values rather than on other aspects such as trial number; that the rule is functional rather than nonfunctional, and that the rule is a positive linear function rather than any other function (15). Brehmer showed that when these rules fail, people tend to assume that there is no rule at all. Brehmer et al (16) provide considerable evidence that determinism, or causality, is a very basic schema used by people to make sense of the world. Wason and Johnson-Laird (17) found that individuals did not use the logical relation of "if, then", but tend to consider this relation equivalent to a double implication, "if, and only if", as would be appropriate for a causal task. The characteristic of probabilism is not manifest, but it has to be inferred. This means that the subject has to have an adequate notion of probabilism among the schemata for organizing his experience, and that he has some criterion for when a task should be considered probabilistic (15). Of course, one can, like Jesdinsky (18) argues, postulate that a deterministic model is only a specification of the probabilistic model in that it assumes the values 0 and 1. In this way, however, one denies the characteristics of the deterministic system in order to establish a probabilistic style which may be alien to the physician and his problem-solving

behaviour. Knowledge of the actual process of problem solving or descriptive process tracing should serve as a useful guide to the design of the model. This involves requirements for a knowledge of the ways in which people apply strategies in order to reach judgments (19). That means that the model must be intimately related to the actual system.

The two aspects of a system are its structure and its behaviour. The structure of a system is the totality of the interrelationships among its elements; the behaviour is composed of the interactions between the system and its environment (see also Lens-model of Brunswik, (20,14). The aim of the model must be to "represent" or to "simulate" the hidden cognitive processes of the physician as he makes his judgments.

It requires a) a search for some formal (i.e., specifiable) model(s), which (b) uses as its "input", the information (data, cues, symptoms etc.) initially presented to the judge, and (c) combines the data in an optimal manner, so as to (d) produce as accurate as possible a copy of the responses of the judge - (e) regardless of the actual validity of those judgments themselves (21). Note that such a model is always an intraindividual one; that is, it is intended as a representation of the cognitive activities of a single judge. The test of the model is not how well it works as a representation of the state of the world, but rather how well it predicts the inferential products of the judge himself (21).

If the model elucidates general principles of diagnosis, then it has both validity and value in teaching (12). Or, as Gorry states, the test of the validity of the model is the determination of whether, for each set of relevant inputs in the modelled system, the representation of these input produce output of the modeling system, which are representations of the output of the modeled system (14). In other words: predicted outcome has to be registered outcome.

To model the medical process, Elstein et al. (22) recommend the following tasks:

1) To identify the intellectual strategies and tactics characteristic of expert clinical reasoning;

2) To generate a psychological theory to explain these features;

3) To relate this theory to current theories of thinking, human information processing, decision-making and problem-solving;

4) To develop instructional methods and materials.

The decisionmaker is dealing with environments characterized by risks, uncertainty, complexity, changes over time, and conflict (19). They can lead to different styles in medical problem solving.

Paragraph 2

STYLES OF PROBLEM SOLVING

Observing people solving a problem the impression arises that everybody operates in his own style as a kind of trademark. Scrutinizing these processes, however, it appears that they have much more in common than that they differ. The question is whether the common features can be classified into a general taxonomy. In our opinion, styles of problem solving can be arranged according to three categories.

A. Personally induced;

B. Problem oriented;

C. Inferential approach;

Each of the categories has its own supporters. The first category of supporters emphasizes the psychological and emotional features in problem solving like uncertainty, stress, intuition, cognition. As we discussed in chapter II these features, undoubtedly influence people's problem solving

behaviour and information processing. In uncertainty or stress people tend to restrict information search, limit the amount of alternatives to be chosen among, avoid risky solutions etc. The Janis & Mann (23) model, as discussed in chapter II, postulates that each pattern of decision stress for coping with conflicting problems is associated with a characteristic mode of information processing. It is this mode of information processing which governs the type and amount of information the decision maker will prefer. It is suggested that these forms of ego-defenses may cause or contribute to dysfunctional behaviour (24). Mason & Mitroff (25) suggest that varying psychological personalities give rise to varying qualities of information. Their "Sensation" type relies primarily on data perceived by the sense, typified by terms like "objective, hard facts", whereas their "Intuitive" type perceives objects as they might be, and in their totality. The "Thinking" type relies primarily on cognitive processes, while the "Feeling" type focuses at affective processes. However, these global descriptions and categorizations provide little explanation for the processes under study; at the most some information about the personality of the problem solver. It fosters stereotypes like the cool, logical thinking mathematician, or the soft, affectionate nurse. It might be conceivable that physicians by choosing their profession demonstrate in most situation an information-processing behaviour mainly based on affection, intuition and hope. However, this is only one side of the task environment of the physician.

Taxonomies of personal behaviour are not always applicable to the particular physician confronted with a contingency task. This kind of taxonomies, widely used in psychology, seems to be of only remote importance to the description of particular problem-solving behaviour (19).

Problem-orientation of the problem-solving process has been repeatedly mentioned in literature (e.g.12,20,26,27). The task governing the processes involved in problem solving, sounds as a reasonable proposition. However, it does not solve our dilemma: looking for common landmarks, recognisable across persons, tasks and situations. The classification of problems appears unsatisfying. Mason & Mitroff (25) arranged problems into two categories, structured and unstructured ones. The structured ones were subdivided into:

a.) decision under certainty: all values of actions, states of nature, and outcomes are known in a deterministic relationship;

b.) decision under risk: the relationship between actions and outcomes is probabilistic;

c.) decisions under uncertainty: probabilities are not known but the possible states of the world are known.

The classification appears to be strongly related to the personal conception of the problem-solver and his processing capabilities. It does not provide independent criteria for classifying a particular task to a particular class of decision category. Moreover, the classification of the innumerable number of tasks as they are presented to the physicians does not only seem a Sisyphean labour, but also unlikely to provide the common landmarks.

Thus we turned to the inferential approach of problem solving and decision making. Decision making can be defined as the process of thought and action involving an irrevocable allocation of resources that culminates in choice behaviour (19). Strategies define the process along which the decisionmaker comes to a decision. Our aim is to discover and to elicit (a) strategy(ies) according to which a physician reaches a decision in the medical process.

The Butterworth Medical Dictionary (1965) defines diagnosis as "the art of applying scientific methods to the elucidation of the problems presented by a sick patient. This implies the collection and critical evaluation of all evidence from every possible source and of the use of any method necessary." Dudley (28) regards as scientific any purposeful human activity designed to provide tentative and reputable hypotheses about the nature of events. In diagnostic reasoning as in many other processes of rational problem solving,

explanation of the observed symptoms and signs and testing of the explaining hypotheses play a dominant role (29). Hypotheses help to overcome limitations of memory capacity and serve to narrow the size of the problem space (20). They can be seen as "the physician's formulation of possible solutions to the problem" (20). Medical problem solving has been described as a process of generating and testing hypotheses. A hypothesis can be thought of as a law of Nature, and the everyday scientific process essentially consists of examining the logical consequences of hypotheses to see if what follows from them or what they predict does in fact correspond to real life (30). It is this process that we may observe, or assume to perceive, in medical problem solving. But do "scientific methods" have the same meaning to us all? In Chapter III we discussed two ways of reasoning, deductive and inductive; probabilistic reasoning being a variant of the inductive way. According to Medawar (30) the traditional view of scientific method is the process of thought known as "induction".

"Inductivism is a complex of beliefs of which the salient points are: the Truth lies all around us out there in Nature, so that the scientists' main task is to discern and record matters of fact and then to classify and appraise them according to certain more or less well defined rules - whereupon the Truth will certainly reveal itself. Thus the scientist collects facts much as entomologists collect beetles. The process of observation can be relied upon implicitly to reveal the Truth as it lies about him in nature. The scientist, therefore, is a shrewd spectator of the world who operates a Calculus of discovery or makes use of some formulatory of intellectual behaviour which reveals the truth to him " (30).

Contrasted to the inductive inference process is the deductive reasoning, which can be described as a process of sequential and hierarchical hypothesis formation corrected by negative feedback.

Thus our research question can be reformulated as "do physicians employ a deductive or an inductive strategy in their problem solving process?" This question raises the investigational task of

1.) Modeling these strategies, in order to discover, to isolate and to illuminate landmarks in the medical processes to enable the differentiation;

2.) Providing instruments to elicit the landmarks.

Paragraph 3

MODELS

Models describe processes in gross features. The model delineates the landmarks within the process as they might be observed in the performance of the task. These landmarks can be viewed as points of emphasis as they have appeared in thought or have originated from observation or from literature. It may be clear that not all features and characteristics will be included in the model.

The more we get into detail, the more we run the risk of emphasizing the - personal - differences rather than the common characteristics.

The focus of this investigation should be the formulation of a model of the medical process and its subsequent comparing to "real" situations. The formal definition of what constitutes a model of either the behaviour or structure of a given system is not an easy task. The key issue is the elucidating of several features in the medical process in order to contribute to the understanding and - perhaps - to optimize medical problem solving and decision making.

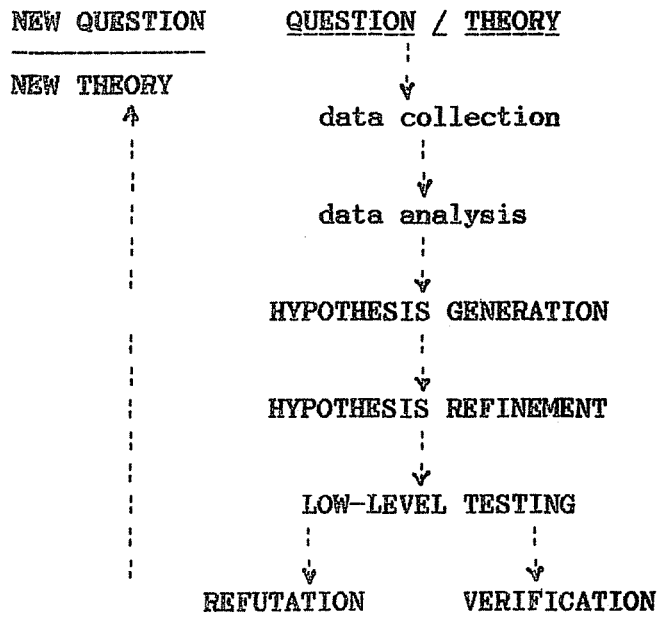
The concept of a model is intimately related to that of a system. Generally, the two aspects of a system with which modelling is concerned are its structure and its behaviour. The structure of a system is the totality of interrelationships among its elements. The behaviour of a system is composed of the interactions between the system and its environment (14). Our question is to what extent the model(s) can serve as an explanation to a system, the cognitive and decision making processes in medical judgment.

Following our research question we have to model both types of inference

methods: the deductive and the inductive one.

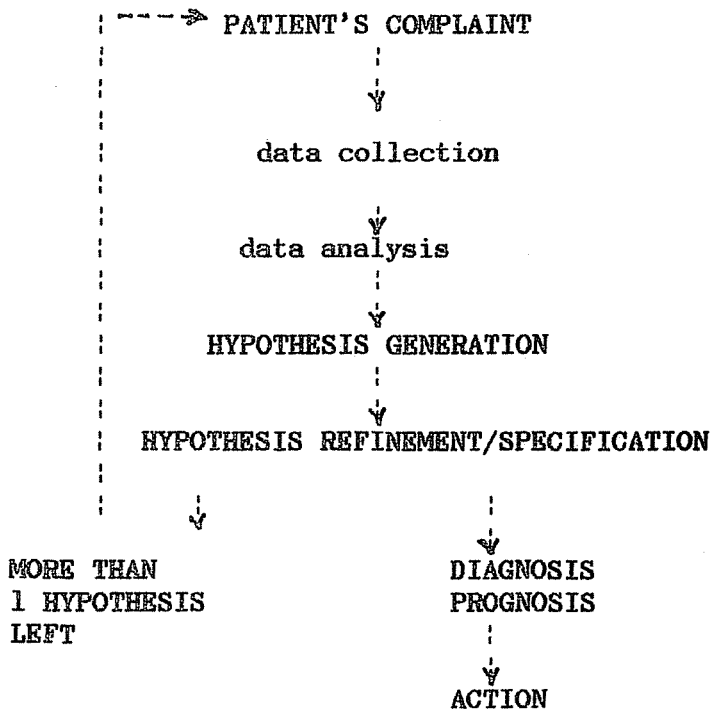
The deductive way of reasoning has been briefly discussed in Chapter III and can be depicted as in the following scheme:

FIGURE I



"Translating" this scheme to the medical nomenclature it can be sketched as follows:

FIGURE 2



This scheme reflects the physician as described by Ledley & Lusted at the beginning of this chapter.

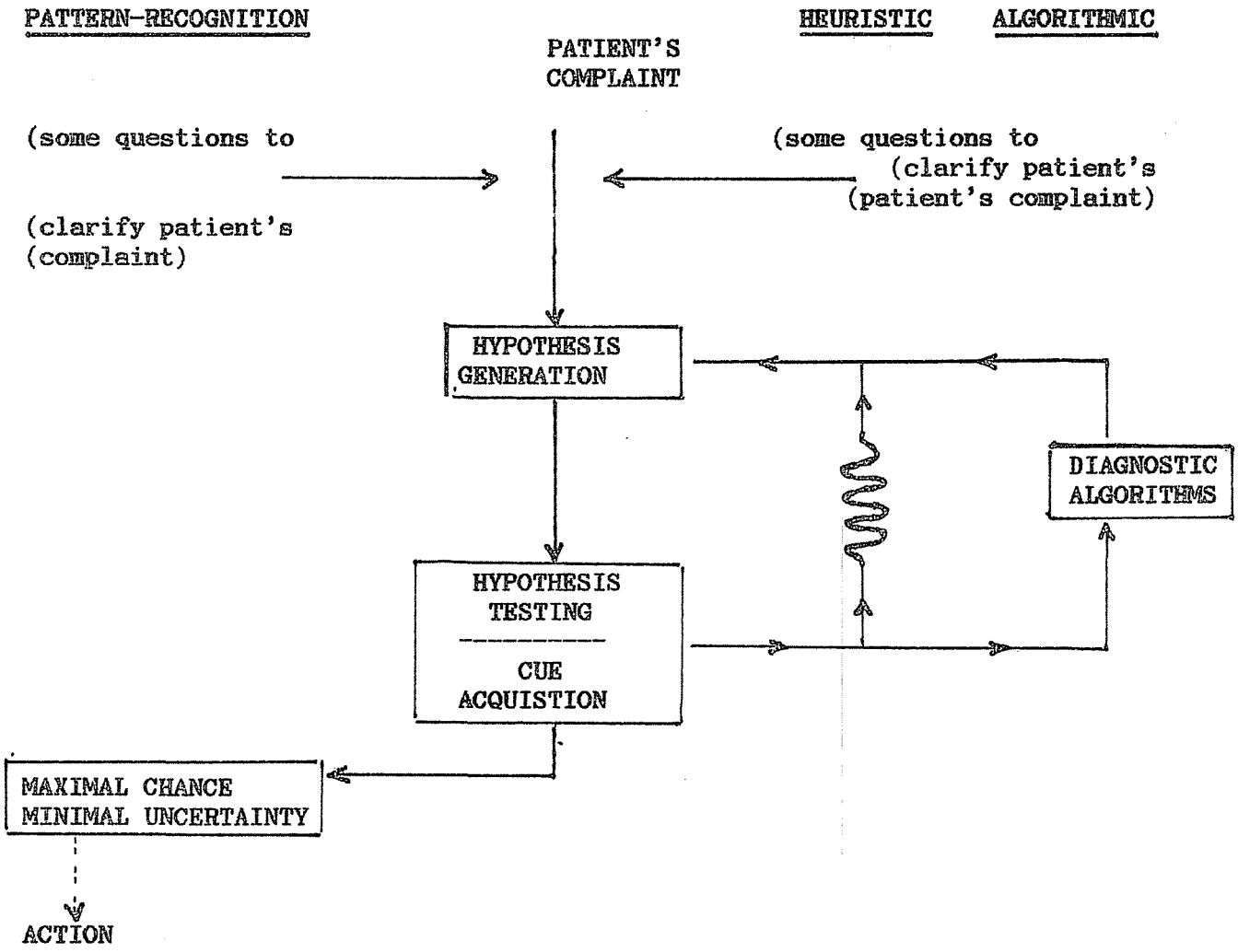
This method has long been taught and recommended to physicians even if not always practiced (31). But training programs aimed at the education of problem

solving techniques with the principles of formal logic, incur a risk of transmitting a way of thinking that hardly exists in practice and is probably quite useless (32). "Most doctors do not practice in the fashion in which they have been trained. The pressure of time, patients and competition leads the doctor to adopt a variety of short-cuts and other useful strategies which may detract from the quality he renders, but which allow him to function more effectively within the context and settings within the world"(33). The notion of restricted search, limitation of alternatives, and the use of task-simplifying heuristics have often been mentioned in literature. It is assumed that the general procedure is to develop a hypothesis about the patient's disease and then to try to confirm the hypothesis (34). These thoughts suggest a different method: an inductive strategy. Unlike the deductive method a general schema for the inductive inference process is not available, as far as we know.

Following thoughts from literature (see also chapter III) we delineated the following scheme for inductive reasoning in medicine:

FIGURE 3

INDUCTIVE



Differences with the deductive strategy are the generation of (a) hypothesis(es) based on precious little data and mainly on past experiences, and the (very) concise testing of the hypothesis(es) with elements which at the same time serve as (a small number of) data to the generation of a new hypothesis, which is stored next to the former ones(s). The ultimate testing is performed by the weighing of the hypotheses.

Again, in the centre of the model the hypothesis is placed. However, we have to remember the different meanings of both hypotheses, the one in deductive and the other in inductive reasoning, as we discussed in Chapter III.The functioning of deductive reasoning yields hypotheses from which testable consequences can be deduced; hypotheses which may be refuted by experience. In case of verification these hypotheses serve as a temporary explanation to the facts at hand. The deductive hypothesis is logically deduced from the presented evidence.

In contrast, the inductive hypothesis can be based on many sources: presented evidence, background knowledge, specific experience etc. The conception about inductive hypotheses mainly refers to hypotheses as reproductions or reflections on analogy and similarity; analogies with which we are familiar to guide us in new tasks. Since these hypotheses about the nature of relations between variables cannot come from the stimulus objects, they have to come from the subjects themselves. Therefore, this type of hypotheses involves a personal investment, a sense of commitment. The generation of inductive hypotheses is triggered by only a small number of features recognized or presumed in the presented data. Inductive hypotheses can be viewed as highly personal "patterns" like we know from pattern recognition.

Pattern-recognition may serve as the main characteristic of the behaviour of the inductive model. It can be defined as the process of matching a patient symptom configuration with those configurations that the physician has memorized either from literature or through personal experience (35).

Those physicians who ask only a few questions further (than the presented complaint), take a decision and then design a treatment plan, are making use of the pattern-recognition strategy. This strategy can be seen as the basic theme of the inductive method.

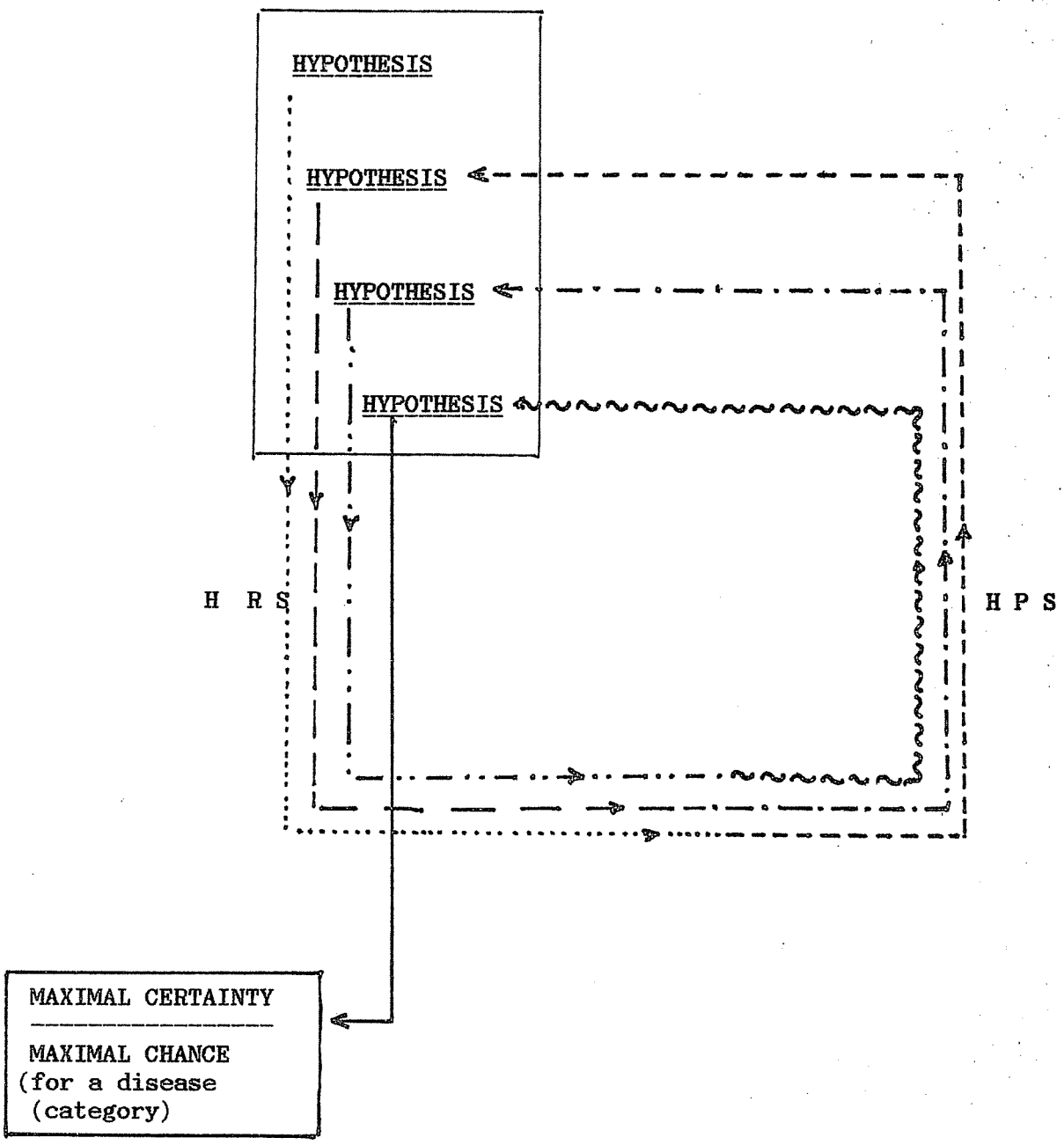
In most cases however, this elementary process does not lead to the goal, making a diagnosis or a plan of action. Second, third etc. cycles are necessary in order to arrive at a satisfactory decision. The Inductive-Heuristic strategy is the iterative process on the basic theme. It makes use of clusters of symptoms and shortcuts in a quick cycling process. Data to the cycle come mainly from the testing of the inductive hypothesis, which, in fact, is a process of filling up and confirming the pattern in mind. Any answers to a particular question of the physician can serve as a trigger to a new pattern. This conception of the cycling process has been sustained by Doroszewski (36). The term heuristic indicates that there is not a structured scheme according to which data are collected; the sequence of questions, examinations or tests is quite casual.

The third variant of the inductive method is the Inductive-Algorithmic strategy. Fundamentally it is based on the same cycling process as the preceding variant. It differs from the former one by its structured plan of strategy. In this study algorithmic has the meaning of:

- a structured plan of strategy by way of short runs of questions, examinations or tests in order to skim over the health or disease status of the patient and/or trying to avoid information-gaps. The plan mainly consists of structured questioning each organ and/or behavioural system;
- the runs have a more or less characteristic form;
- the algorithm(s) are physician - not problem-oriented (35).

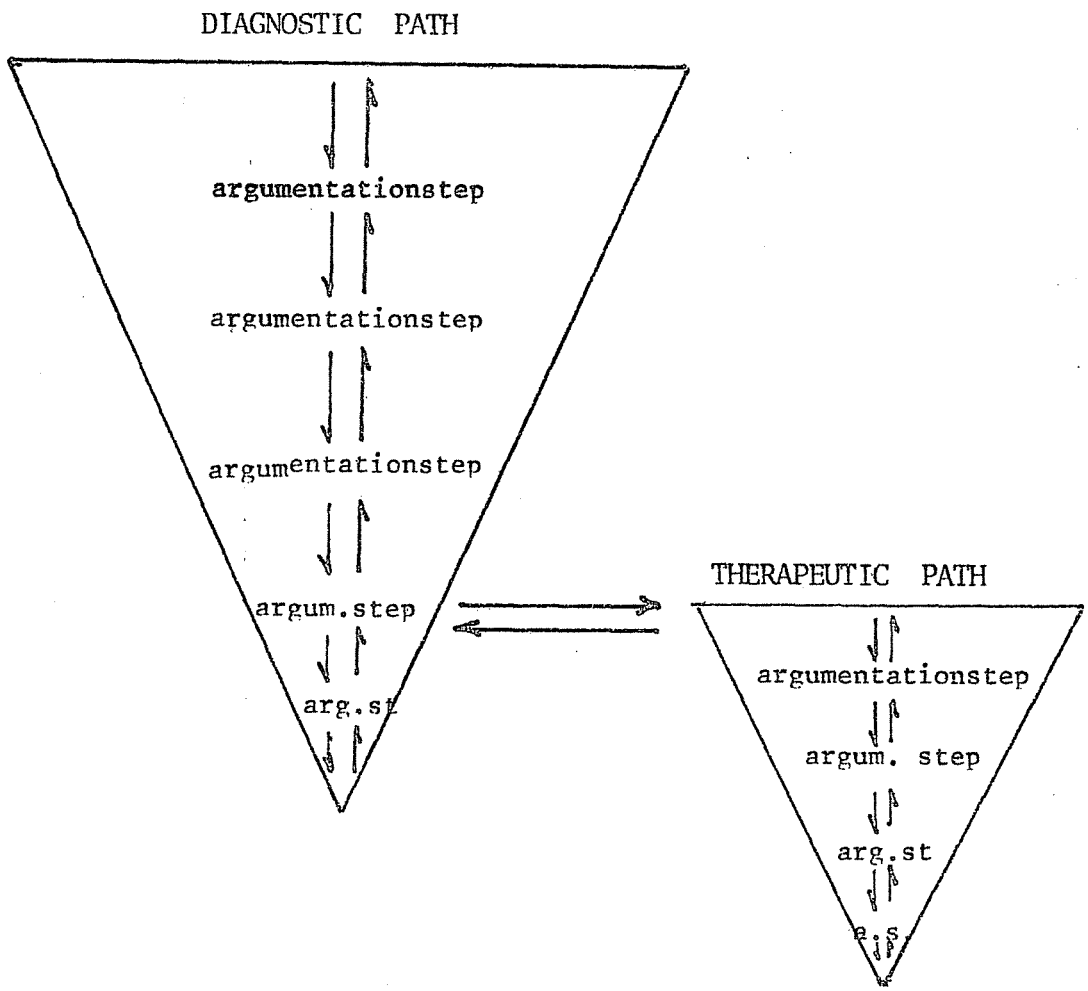
The fundamental feature of the inductive model, the cycling process, can be pictured as:

FIGURE 4



Contrasting to this cycling process is the principally linear one of the deductive strategy. The deductive strategy can be delineated as progression down a hierarchical logical sequence of hypotheses. In this system the high level hypotheses serve as universal statements about an observed phenomenon e.g. the patient is ill. In an increasing specification a sequence of hypotheses is constructed which eventually yield a hypothesis from which testable consequences can be deduced that may be refuted or verificated by the presented or acquired evidence. It can be pictured as a triangle.

FIGURE 5



The various hypotheses in the deductive method cannot be arranged in some kind of order. They are unique pictures induced from what Polanyi calls the "Tacit Knowledge". The condition of a deductive system is its logical ordering of the hypotheses. The question is whether such a logical ordering is possible in medicine.

This question led to the meditation about methods and instruments allowing the distinction between these two, deductive and inductive, strategies.

Human judgment is usually investigated to employ three types of models: PROCESS studies of how human judgments are actually made, PRESCRIPTIVE models which make use of a formal theory of how decisions ought to be made, and DESCRIPTIVE models which attempt to describe the context and factors which influence judgments. Elstein et al. (37) have recently critically reviewed in detail the strength and weaknesses of all three approaches to the investigation of clinical cognition.

Process studies can be distinguished into:

A.) Process decision models. which can be typified by the information processing studies of Newell & Simon (27), and

B.) Process tracing studies These studies have been mainly executed by subjects' reporting their cognitive processes of preceding judgmental procedures. The validity of these types of studies have been questioned by Nisbett & Wilson (38). Their main argument is that hindsight confuses reported data. We are "telling more than we can know". Some of their conclusions have been argued by Ericsson and Simon (39). These authors propose that verbal reports as data may be valid when registered at the time of action.

From the previous paragraph it may be clear that process tracing is the model to be preferred in our investigation. It includes the registration of the physician's problem solving activities in an as real as possible situation. It states the problem to find and to distinguish the landmarks which enable us to discover the deductive and inductive way of reasoning when present.

The main features of the deductive strategy are:

1.) progression down the hierarchical hypotheses sequence; and

2.) its conclusion is based upon the evidence at hand.

The latter condition appears to be the easiest one: unambiguous registration of the evidence and the eliciting of the diagnostic "contents" (cluster of symptoms and signs) provide an easy comparison.

The former condition, however, is the more difficult one. The problem is how to structure and to recognize a hierarchical hypotheses system which enables a downward analysis. Only a few attempts have been made to originate such a system.

Elstein et al (20) distinguish two levels: complex multilevel formulation and fairly specific diagnostic. This system does not allow a downward analysis. A three-level structure is proposed by Gerritsma & Smal (32): Global, unspecific and specific. These features, however, were not specified.

Doroszewski (29) distinguishes four levels:

-leading hypotheses, suggestive for the direction of the investigation;

- intermediate, explanatory to the observation;

-working, nearest explanation in a given step of reasoning; and

- main diagnostic, describing an important bio-pathological state.

These levels do not permit an independent judgment of the hypotheses across the various cases and patient problems.

For the present investigation a five-level physio-morphologic structure was developed, based on "systems view of illness and disease". (Engel, 1980, reported in 40). This multilevel general systems model ranges from Biosphere in 14 steps to subatomic particles. Within this model five levels within the human sphere can be distinguished: person, mental system, organ system, tissues and cells. From this conception the following model was constructed.

Level I: <u>Human Being</u> . Hypotheses referring to global descriptions comprising the total human being will be included within this category. It refers to descriptions like ill/not ill, psychic/ somatic, serious/not serious condition etc.

Level II: <u>Multiorganic formulation</u> . Hypotheses within this class refer to pathophysiological states comprising more than one organ system. It includes diseases of body systems and/or body parts. The class can be typified by examples like:

unspecified infectious diseases, unspecified neoplasmata, unspecified

ailments of bodyparts ("something in the chest", "abdominal pain").

Level III: Organ System . This class follows the normal medical distinction of organ systems. E.g. respiratory tract, digestive tract, musculo-skeleton tract. All hypotheses referring to a certain organ system without further specification of the illness are included in this class.

Level IV: Organ . This class refers to specified organs and tissue-entities. It includes the nearest explanation to the specified disease entity. Hypotheses nominating the specific organ or tissue-entity are included. It can be typified as: gall-bladder disease, anaemia, eczema etc.

Level V:: Cellular . This class can be compared with Doroszewski's "Main diagnostic", describing a specified bio-pathological state. It refers to al specific diseases as generally described in medical textbooks. The hypothesis provides an exact circumscription of the lesion and/or its cause. For instance, duodenal ulcer, hyperthyreoidism, acute glomerulo-nephritis, myocardial infarction.

In preliminary study the system was put to trial. Two independent judges classified 57 hypotheses according to the system. All hypotheses could be placed within a class; the interjudge variability numbered less than 5%. In the actual investigation general agreement could be reached for all 745 hypotheses.

We judged the system fair enough to distinguish a hierarchical structure of the hypotheses mentioned in the physician's problem solving activities of a patient case.

The characteristic of the inductive strategy is given by the circular process. This characteristic is typified by the peculiar arrangement between hypotheses and symptoms. As aforementioned, the induced hypothesis is followed by a number of symptoms (answers to pertinent questions) which serve as testing elements as well as information towards a new hypothesis in a succeeding cycle of the process. In this respect, we can distinguish three types of "symptoms" (including signs and tests). Symptoms as testing elements related to the preceding hypothesis (Hypothesis Related Symptoms, HRS); symptoms as information data related to a forthcoming hypothesis (Hypothesis Prerelated Symptoms, HPS), and symptoms unrelated to any of both hypotheses (Non Hypothesis Related Symptoms, NHRS). These latter symptoms may be characterized as data indicating a search behaviour in an uncertain situation. This model of symptom arrangement according to pertinent hypotheses requests a method of unambiguous allocation of symptoms to hypotheses.

To design the method two conditions have been framed:

1.) symptoms must unequivocally be recognized across situations; and

2.) symptoms must be coded* in accordance with the coding of the hypotheses.

(note: this description of the code only refers to the particular aspect of hypothesis-symptom relationship. The storing, tracing and retrieval of symptoms follows a subsequent method of coding.)

As we previously mentioned, difficulties may arise when indicating the special meanings of a symptom. These meanings can vary to a great extent according to the various characteristics of diseases, their courses and the patient. Besides, the variation of the physician's framing of questions for some symptoms can contribute to more confusion.

Required for the application of the method is:

a.) symptoms have to be recognized and coded in atomic states; and

b.) coding cannot be allocated to the physician's questions but to the "atomic symptoms" themselves.

The latter condition includes a classification of these "atomic symptoms" throughout the medical facts, allowing a physician-subject to interview the data

base freely, in his own personal style. The outcome of this task will be discussed in chapter VII.

It seems natural to code symptoms and hypotheses according to the various organ systems, being the level at which a non-committal allocation can be ascertained. Generally, hypotheses and most symptoms refer to only one or two organ systems. Those symptoms covering a larger domain were coded in conformity with all pertinent organ systems. The term 'organ system' must be conceived in a broad sense. The 'organ system' coding refers to:

a. - Circulatory tract
b. - Nervous system
c. - Skin
d. - Digestive tract
e. - General somatic illness
f. - Endocrine system
g. - Musculo-skeleton system
h. - Respiratory tract
j. - Urogenital system
k. - Blood
l. - Mental illnesses and ailments.

The method operates as follows

Suppose a hypothesis is generated referring to an organ system 'a' "testing" symptoms (as related to the physician's pertinent questions) can be traced as related (HRS) when coded for the same organ system. When the symptom's particular organ system code does not harmonize with the generated hypothesis's one, two possibilities can be distinguished. Either the symptom belongs to a non-committal search behaviour, and is unrelated to a particular hypothesis (NHRS), or the symptom is related to a forthcoming hypothesis, a hypothesis pre-related symptom (HPS). In the latter case the code of the next hypothesis, e.g. h, reflects the code of the symptom, which in that case can be allocated to the second cycle etc.

The cycling process can be easily traced from this procedure.

Needless to say, that the collected evidence does not (always) cover the diagnostic contents.

We shall now proceed to chapter VII, describing the instruments which enabled us to register and to document the data pertinent to the investigation purposes: to identify the inferential strategies in the physician's problem solving.

REFERENCES CHAPTER VI

1. Ledley R.S., Lusted L.B.,: Reasoning foundations of medical diagnosis. Symbolic logic, probability and value theory aid our understanding of how physicians reason. Science 1979, 130, 9 - 21.
2. Sober E.,: The art and science of clinical judgment : an informational approach, in: Engelhardt H.T., Spicker S.F., Towers B.,: Clinical judgment, a critical appraisal. Dordrecht,D. Reidel Publ. Co., 1976.
3. Leaper D.J., Gill P.W., Staniland J.R., Horrocks J.C., Dombal F.T. de, : Clinical diagnostic process: an analysis. Br. Med. J. 1973, 3, 569 - 574.
4. Scriven M.,: Clinical judgment. In: Engelhardt H.T., Spicker S.F., Towers B., : Clinical judgment, a critical appraisal. Dordrecht, D. Reidel Publ., Co, 1979
5. Hull F.M.,: Diagnostic Pathways in General Practice, J. Roy., Coll. Gen. Pract., 1972, 22, 241 - 257.
6. Hull F.M.,: Diagnostic pathways in rural general practice. J. Roy. Coll. Gen. Pract., 1969, 18, 148 - 155.
7. Howie J.G.R., Diagnosis - the Achilles heel? J. Roy. Coll. Gen. Pract., 1972, 22, 310 - 315.
8. Scadding J.G.,: Viewpoint. The semantics of medical diagnosis. Biomed. Comp., 1972, 3, 83 - 90.
9. Scadding J.G.,: Diagnosis: the clinician and the computer. Lancet, 1967, 877 - 882.
10. King L.S.,: What is a diagnosis? JAMA, 1967, 202, 154 - 157.
11. Stein M.A., Winter J., : Theory development in medical decision making. Int. J. Biomed. Computing, 1974, 5, 147 - 159.
12. Pauker S.G., Gorry G.A., Kassirer J.P., Schwartz W.B.,: Towards the simulation of clinical cognition. Taking a present illness by computer. Am. J. Med., 1976, 60, 981 - 996.
13. Dombal de F.T., Horrocks J.C., Staniland J.R., Gill P.W., Simulation of clinical diagnosis: A comparitive study. Br. Med. J., 1971, 2, 575 - 577.
14. Gorry G.A., Modelling the diagnostic process. J. Med. Educ. 1974, 45, 293 - 302.
15. Brehmer B., In one word: not from experience. Acta Psychologica, 1980, 45, 223 - 241.
16. Brehmer B., Kuylenstierna J.: Content and Consistency in probabilistic inference tasks. Organiz. Behav. Perf., 1980, 26, 54 - 64.
17. Wason P.C., Johnson-Laird P.N.,: Psychology of reasoning; structure and Content, Cambridge (Ma), Harvard University Press, 1972.
18. Jesdinsky H.J.,: Diagnose-Modelle in der Medizin. Meth. Inf. Med., 1972, 11, 48 - 59.
19. Sage A.P.,: Behavioral and organizational considerations in the design of information systems and processes for planning and decision support. IEEE transactions on systems, man and cybernetics, 1981, 11, 640 - 678.
20. Elstein A.S., Shulman L.S., Sprafka S.,: Medical Problem Solving, Harvard University Press. Cambridge (Mass). 1978.
21. Goldberg L.R.,: Simple Models or simple processes? American Psychologist, 1968, 23, 483 - 496.
22. Elstein A.S., Kagan N., Shulman L.S., Jason H., Loupe M.J., Methods and theory in the study of medical inquiry. J. Med. Educ. 1972, 47, 84 - 92.
23. Janis I.L., Mann L.,: Decision making, New York, Free Press, 1977.
24. Bariff M.L., Lusk E.J.,: Cognitive and personality tests for the design of management information systems. Management Science, 1977, 23, 820 - 829.
25. Mason R.O., Mitroff I.I.,: A program for research on management

information systems. Management Science, 1973, 19, 475 - 485.
26. Groot A.D., de, Methodology, Mouton & Cie., 's Gravenhage 1961.
27. Newell A., Simon H.A.,: Human problem solving. Englewood Cliffs, New Yersey, Prentice Hall, 1972.
28. Dudley H.A.F., : Clinical method. Lancet, 1971, 35 - 37.
29. Doroszewski J.: Unity and Diversity of the medical action: A review of its components and their interconnections. Metamedicine. 1981, 2, 155 - 68.
30. Medawar P.,: Scientific method in science and medicine. Persp. Biol. Med. 1975, 345 - 352.
31. Politzer P.: Decision analysis and clinical judgment. Med. Dec. Mak., 1981, 1, 361 - 389.
32. Gerritsma J.G.M., Smal J.A.,: De werkwijze van huisarts en internist. Een vergelijkend onderzoek m.b.v. een interactive patienten simulatie. Utrecht, W.U. Bunge, 1982.
33. Mechanic D. : Medical Sociology. 2nd ed. New York, Free Press, 1978.
34. Kleinmuntz D.N., Kleinmuntz B., : Systems simulation. Decision Strategies in simulated environments. Beh. Science, 1981, 26, 294 - 305.
35. Ridderikhoff J.,: Models of decision-making in the general practice: a design for a descriptive research. Med. Informatics., 1985, 10, 323 - 337.
36. Doroszewski J.,: Hypothetico-nomological aspects of medical diagnosis, part I: General Structure of the Diagnosis Process and its hypothesis-directed stage. Metamedicine, 1980, 1, 177 - 94.
37. Elstein A.S., Holmes M.M., Ravitch M.M., Rovner D.R., Holzman G.B., Rothert M.,: Medical decisions in perspectives: applied research in cognitive psychology. Perspect. Biol. Medic., 1983, 26, 486 - 501.
38. Nisbett R.E., Wilson T.D.,: Telling more than we can know: Verbal reports on mental processes. Psych. Review, 1977, 84, 231 - 259.
39. Ericsson K.A., Simon H.A.,: Verbal Reports as data. Psych. Review, 1980, 87, 215 - 251.
40. Taylor T.R., Gordon M.J.,: Understanding the clinical management process, in: Decision Making in general practice. London, McMillan, 1985.

CHAPTER VII

INSTRUMENTS TO IDENTIFY THE PROBLEM SOLVING STRATEGIES

Various methods have been described to study psychological and/or thinking processes. The implicity of the subject forces investigators to make a selection of aspects or to choose among types of research models. As sketched before, mainly three types of models can be distinguished:

1.) process studies of how human judgments are actually made; the human being, e.g. the physician, as a base for the study of professional, cognitive and/or behavioural processes;

2.) prescriptive studies, which make use of a formal theory of how decisions ought to be made. Their aim is to build or to rebuild strategies according to which the medical process, the diagnostic as well as the therapeutic part, can be shaped or reshaped into an - easy - manageable structure which allows improved performance and productivity of this process;

3.) descriptive studies which attempt to describe circumstantial and situational factors which influence judgments (1). Descriptive studies involve the problem-solver's task environment.

The process studies try to understand who the doctor is, what he is doing, how he functions and what his effectiveness as information processor is. In other words, we refer to studies that try to understand all the medical actions which are laid down in the physician's professional duties and responsibilities. Surprisingly little research has been done so far on these professional activities. Although there is a growing interest in this field, a number of methodological limitations seems to inhibit further investigations. Some of these limitations are mentioned before such as:

- the complexity of the medical system;
- the lacking of an uniform, unambiguous medical classification system, but also
- the difficulty of observation in a patient-physician encounter. This encounter encompasses two systems, the patient's and the physician's. From this the main problem arises, because studying the functioning of the physician means observing these two systems which, at the same time, influence each other in a special implicit way. It is like focusing on a certain point on a turning wheel while sitting on another wheel rotating in an opposite direction. One observes flashes only partly recognisable. Uberla (2) notices some other problems such as:
- the difficulties inherent to the formulation of feasible tasks. Substantial segments of professional medical action have never been approached in a way amenable to empirical investigation.
- the high level of inconstancy and variability. Diseases change the unstable equilibrium of the body functions. The equilibrium states change every time the organism tries to regulate itself. The variability from one patient to another is a still larger problem;
- the rarity of events. This rarity of specific phenomena is directly coupled with the huge variability inherent to medicine. Therefore, rare events can only be grasped by observing large numbers of cases, which represent another barrier to the feasibility of empirical studies;
- the incalculable totality of the phenomena involved. Dissecting the medical actions does not mean an understanding of the totality of the composing elements.
- the causal relationships are very difficult to comprehend and to prove in complex systems, of which man is certainly one;
- The inadequacy of theoretical models represents a serious limitation;
- the lacking of validated instruments for the collection of data relevant to medical dialogues and interviews, decision-making strategies, therapeutic procedures or management plans. Neither are we able to outline the utility of

the outcome of medical processes.

Uberla concludes that:

1) A science of professional medical action should be developed, centering around empirical research into precisely that action. Medical action, seen as a goal-oriented, systematic process including as elements the patterns of disease, resources, doctor-specific attributes, services and outcomes, should be more thoroughly studied than has been the case so far.

2) To make this possible, a constructive and aggressive clinical methodology is required, one for which the essential materials will come from statistics, informatics, social science, epidemiology, medical science itself as well as from the clinician's experience and knowledge.

The process studies can roughly be divided into two categories:

-studies based on the decision processes in medical problem-solving.

-studies investigating - some aspects of - the physician's procedures. They try to analyse a number of aspects inherent to the decision making behaviour of physicians.

To gain deeper insight in the working methods of physicians mainly three ways for investigation can be employed:

- observation,
- introspection, and
- simulation.

The observational methods seem to be the track with the highest validity. But they meet some major obstacles such as:

a) the influence of observation on the doctor-patient encounter. The very fact of being observed may alter the nature of the situation which is tried to be observed and documented (3). Although both partners judge this influence of minor importance we nevertheless do not know its magnitude.

b) the complexity of the process. If one does not exactly know what to look for, one tends to a superficial selection of observations not only liable to the observer's variation but also to confusion. Confusion may arise as the process consists of qualities and elements of various specialties and disciplines.

c) The two-systems mode of the physician-patient encounter.

d) The ethical issue : the guarantee of the patient's privacy and the physician's secrecy cannot be maintained (4).

Apart from this direct observational way indirect observation can be employed. We can identify three methods:

- medical records survey;
- interviewing;
- questionnaires;

Medical record survey can elicit a number of aspects in a broad perspective. The study of Noren et al. (5) gives details about referral patterns, duration of visits, use of physician time, extent of diagnostic effects, efforts at patient education, and provision of personal advice or emotional support in a comparative study of 610 family practitioners and 347 specialists in internal medicine. For gaining insight in the medical problem-solving process this type of observational research is unrewarding.

Generally, medical records do not represent the most completely documented information source. The physician takes down only a limited number of mostly positive directed facts and aspects which are of special interest to him and the special case. An unacceptable selection which does not allow to draw special or general conclusions. The attempts to standardize medical recording (e.g.Problem oriented medical record (6)) have failed to overcome a number of these problems. The strategies of interviewing and questionnaires fall mainly in the domain of introspection.

Introspection focuses on one of the two persons in the patient-physician encounter. It is based on - mostly retrospective - interviewing of the individual and verbal reports of the preceding problem-solving task. Several

studies in the medical field have been reported among which are famous studies by Kleinmuntz (7), and Wortman (8). However several psychologists question the validity of these approaches. Some cognitive psychologists (Mandler(9), Miller (10)) have proposed that we may not have direct access to higher order mental processes such as those involved in evaluation, judgment, problem-solving, and the initiation of behaviour (11). Miller (10) assumes that it is the result of thinking, not the process of thinking, that appears spontaneously in consciousness. Nisbett & Wilson (11) found that subjects, when asked about the answer they gave to report their memory of specific events "theorize" about their processes. Miller examplified this process as follows: "If a person is asked : "What is your mother's maiden name?" The answer appears swiftly in consciousness. Then if this person is asked : "How did you come up with that?" he is usually reduced to the inarticulate answer: "I don't know, it just came to me".

Nisbett & Wilson (11) came to three major conclusions:

1) people often cannot report accurately on the effects of particular stimuli on higher order, inference-based responses. They sometimes deny an inferential process of any kind. The accuracy of subjective reports is too poor to produce generally correct or reliable reports.

2) When reporting on the effects of stimuli, people may not examine a memory of the cognitive processes that operated on the stimuli; instead they base their reports on implicit, a priori theories about the causal connection between stimulus and response.

3) Even if the reports are correct, this does not mean that the instances of correct report are due to direct introspective awareness.

Slovic & Lichtenstein (12) have reviewed the literature concerning the ability of subjects to report accurately on the weights they assign to various stimulus factors in making evaluations. They come to the conclusion that self-insight was poor and that of the studies which allowed for a comparison of the perceived and actual cue utilization, all found serious discrepancies between subjective and objective relative weights (see also chapter V).

Unfortunately, most studies reviewed provide little data as to what information is heeded during the thought processes. In a number of investigations the subjects were forced to infer rather than remember their mental processes. Modern psychology has been vague about the use that can be made of these verbal reports. It may be clear that questions like: "How do you do these tasks?" produce different reports from more precise questions. The former question implicitly or explicitly requests a general, rather than specific, interpretation of how the individual was performing the task in question. In these cases the subject may be drawing on prior informations such as general knowledge on how one ought to do these tasks, to generate a verbal report describing a general procedure or strategy. In these cases, the verbal report may not bear any close relationship to the actual cognitive processes (11). In studies that use retrospective verbalization, subjects are seldom asked what they can remember about specific instances of their cognitive processes. When subjects articulate information directly that is already available to them, verbalization will not change the course and structure of the cognitive processes, nor will it slow down these processes. (13). Brehmer et al (quoted in 13) did not find any significant effect on performance when subjects directly wrote down their current hypotheses in a booklet. What is remembered, and how well, will generally depend critically on the interval between past event and the moment of recall. This interval is an important consideration in classifying verbalization procedures. If information is verbalized at the time the subject is attending to it, Ericsson & Simon (13) label the procedure concurrent verbalization. If the subject is asked about cognitive processes that occurred at an earlier point in time, it is labelled: retrospective verbalization. Ericsson & Simon found that under certain conditions concurrent verbalization does not affect essentially the processes under study. No reliable differences were found between groups of which one was being instructed being silent and the other to "think-aloud". However, the thinking-aloud group took

about 50% more time than the control group. Apart from this feature no effect on speed of performance was reported (13).

In problem-solving situations with a heavy cognitive load (as e.g. in medical processes), initial instructions were found to be disregarded by the subjects. These instructions had to be monitored by the experimentor (78). When verbalizations were asked before (blocks of) trials, effects of this verbalization were not found. Asking subjects what they thought was the correct solution before each trial in a discriminant-learning experiment general information was not found to change the proportion of correct responses. The Ericsson-Simon model assumes that only information in focal attention can be verbalized, which is in accordance with theories on human thinking.

When the time-gap between verbalization and the attending instance widens, people have to rely more and more on memory. However, memory retrieval is fallible and sometimes leads to accessing other related, though inappropriate information. It is in this light that the assumption about stable subjective estimations of probabilities and utilities is hardly tenable.

Kassirer et al. (14) question the role of "stimulated recall" in the Elstein c.s. study. Elstein and co-workers asked the physician-subjects about their previous thought processes while they watched videotapes of their earlier encounters with the simulated patient. Kassirer et al. think it possible that this retrospective verbalization may function as to suggest hypotheses that might not have been evident from the presented data.

Retrospective accounts leave much more opportunity for the subject to mix current knowledge with past knowledge, making reliable inference from the protocol difficult.All theorizing about the causes and consequences of the subject's knowledge state is carried out and validated by the experimentor, not by the subject (15). We believe that an experimental design that relies heavily on retrospective observation is potentially flawed (16).

Simulation of the medical process encompasses two persons, the patient and the physician.Studying the doctor automatically includes a "fixed" patient. A "fixed" patient can only be a simulated patient, because real patients are never identical in different situations, different times and with different physicians. It also includes:

-different interpretations upon identical information;
-different framing of queries;
-misleading answers;
-symptoms and signs changing;
-varying kinds of questions asked, and
-differing sequences of questions (17)

The simulated patient can present exactly the same clinical problem and the same clinical attributes to many different examining physicians. The variable has to be the physician, not the patient (18).

Generally speaking there are three possibilities for patient simulation:

1) 'actors' simulation:
2) written simulations, with or without the intervention of a simulator.
3) computerized simulations, with screen display and data storage.

Each of these items has its advantages and disadvantages.

Sage (19) listed a number of requirements for simulation, which, as he adds, are generally not met in practice. They include:

a) comprehensive identification of all needs, constraints, and alterables relevant to planning and decision-making;

b) determination and minimization of cost and maximization of effectiveness with regard to a);

c) detached neutrality and a calculative orientation rather than arbritrariness, conflict and coercion;

d) a unified process that will cope with interdependent decisions;

e) sufficient time to use the method;

f) sufficient information to enable use of the method;

g) sufficient cognitive capacity to use the method.

For the purpose of the present investigation we needed a device that suited the purpose of storing a large number of data. These data must be highly flexible, have to cover a broad area of symptoms and signs, and allow physicians freely to question the 'patient'. The device should be easy to produce and manage and must meet the requirements of realism on the part of the patient, and above all it must be cost-effective. It should allow to study physician-subjects in their natural environment.

With a view on these requirements a model of a "paper patient" has been constructed.

Paragraph 2

THE PAPER PATIENT

The "paper patient" can be described as a multipurpose device suitable for a number of applications such as:

- simulation in medical education;
- simulation for research purposes;
- data storage and retrieval system.

Essentially, paper patients are case histories narrated in the time sequence of the real patient. This case history is more or less formally structured by its composer(s). The degree of structuring determines partly its use. The more formally structured paper patients are mainly used in medical education for training and examination purposes. Examples of this type of simulation formats are the Modified Essay Question, a case history divided into several parts with a number of (Multiple Choice) questions at every interruption, and the widely used Patient Management Problem (PMP) formats and its variants. These types do not allow the physician (or the student) to follow his own cognitive pathway. Besides, displaying the written scenario can suggest questions, options or categories to the testee that otherwise not crossed his mind : the so-called 'cueing-problem' (20).

This 'cueing-problem' also emerges in a less rigid format, the oldest simulation method, that of Rimoldi (21,22), as based on the tab-item technique. The case history is written on a series of cards. Each card represents a patient attribute. The 'patient' is interrogated by picking cards. On the frontside of the cards preformulated questions are shown. The question that is nearest in conformity with one's own question is picked up. The relevant clinical information is printed on the reverse side of each card. Modern examples of this method are Portable Patient Problem Pack (P4) and its extension P3 (23,24), the Diagnostic Management Problem (DMP) (25), and the Minisim Mode (26).

In actor simulation an actor, mostly an interested layman, trained to simulate patients, is provided by a detailed synopsis of a patient problem i.e. the personal and medical history part. It allows the testee to follow his own pathways, and it avoids the cueing problem. Although seemingly preferable, there are a number of disadvantages such as:

- limited contents: the more facts the actor has to remember, the more liable he is to confusion and to forgetting, the more omission, the more constraint in the use of actors in simulation;
- restraint in the construction: the more complicated the patient scenario the more mistakes can be made;
- restriction of the mode: generally actors are not able to simulate the elements of physical examination or deliver the data base for their biochemical or X-ray properties;
- shift in role content: the repeated interaction with various physicians can lead to shifts in the contents and emphases of the patient scenarios
- limited use because of the availability of actors in various appointments.
- high cost: intensive training and the aforementioned restrictions count for high cost, because several actors have to be used to fulfil the demands of

more elaborate projects.

Actor simulation is examplified by the "high-fidelity" simulation used by Elstein et al. (27) in the first part of their study.

Eventually, we chose a mixed form:a verbal presentation of a written patient scenario. The patient's case history is recorded in a prestructured system (cards, books). The user asks questions verbally to an operator as though he, the operator, is the patient. The answers were provided through the operator from the - written - system. The verbal mode was found to be most enjoyable and realistic, which is consistent with the findings of de Dombal et al (26). This mode allows the testee to follow his own cognitive pathways and grants various scoring possibilities according to the individual's workup.

Formating this type of 'paper patient' a number of conditions have to be fulfilled. It requires a structure in which particularity can be transformed into a system that is general to all physicians. For the construction we need to know how data are identified, how they are selected, how many data are appropriate, how to test data, how to process data. It is essential to know the content and structure of the medical process or the various medical processes.

1) the contents: the data and symbols which compose the basis for diagnosis and therapeutical management.The various constraints of these elements have been discussed elsewhere.

2) the structure: the rules, the laws and sequences binding the data to verifiable or refutable concepts of disease or health status of the patient.

Thusfar, most simulation models are built on a number of assumptions. From the various thoughts a number of wishes, conditions and criteria for a simulation model can be formulated.

1) Case histories as obtained from medical records or composed by expert groups must be replaced by a patient's data base highly devoid of value judgments.

2) Data acquisition and storing must be uniform and unambiguous.

3) Data must come from as many sources as possible: the patient, his relatives, physicians, nurses, social background etc.

4) Data storage must not be submitted to a predetermined processing mode.

5) Data must be gathered by an independent collector from real situations.

6) The data must easily and in short a time be stored in the system (high production rate).

7) Retrieval of the data by the testee must allow an individual workup through the case.

8) No cueing is allowed.

9) Preferably the "patient" is addressed by the candidate in spoken word. This means the use of an operator between candidate and "paper patient".

10) The operator must be a trained person acquainted with the medical jargon and be able to handle a large number of scenarios (cost-aspect) without influencing the candidate.

11) The system, including the operator, must allow for a realistic provision of answers in a given time (number of questions per minute).

12) The simulation must approximate the setting and time constraints of the routine day-to-day professional activities of the physician.

13) The scoring must be based on what really happens in practice.

14) The scoring variables and their weights must be founded on a reliable and verifiable medical data base.

15) The simulation must meet the qualities of validity and reliability (4).

Trying to avoid a number of questions and limitations and to fulfil a large number of the abovementioned requirements we created a new system. One of the bottlenecks of the existing models is the discrepancy between the available data and the number of options that could be handled by a candidate. According to Gerritsma & Smal (28) a maximum of about 1600 options can be handled by a

candidate or simulator. But, according to Pauker et al. (29) the estimated total number of facts in general internal medicine and its subspecialties (nephrology, cardiology, hematology) is about two million facts as the core body of information in this specialty. For general family practice this number can be doubled, even tripled or more. There seems no way of bringing this numbers in a manageable format.

However, it crossed our mind that thusfar simulations were adapted to medicine and, as far as we know, not vice versa. But trying to inflict a more or less consistent system upon a rather unsystematic huge body of data is likely to fail. The only possibility seemed to us to systematize the medical data. Out of these thoughts this model was originated. The system is based on the elementary attributes of medicine : the symptoms and signs. The system can be viewed as a medical data storage model. Classical data storage and classification models almost invariably recommend the entering data and problems at the highest possible level of diagnostic refinement of which the user can be confident at the time. This leaves the interpretation with the physician.

It means :

a) a subjective interpretation of the observed symptoms and signs, when observed;

b) a subjective judgment about the aggregation and the level of aggregation of the symptoms and signs ('diagnosis').

In a preliminary study we found only a small conformity among physicians about the content of a diagnosis. This finding led to the construction of a system based on the level of symptoms and signs. Because there is no essential difference between symptoms and signs we shall refer to signs as symptoms. A symptom is defined as:

"Every functionally objective or subjective phenomenon which can be observed directly or indirectly and which is indicative of either an illness or condition of the patient". It also includes laboratory and x-ray reports and data.

The next problem is how to bring an estimated number of 5-10 million of facts into a manageable system. Studying various medical textbooks it occurred to us that several symptoms were mentioned more than once dependent upon the context in which they were used. They appeared in different forms and different settings depending on their causal relationships to each other or to the disease to which they can be attributed.

The symptom Pain e.g. was mentioned four, five, six or more times for one and the same disease, describing various states like: localisation, onset, irradiation etc. It seemed that the entity 'symptom' is not the basic element in medicine. Russell and then Wittgenstein have proposed that the world might be susceptible to description in terms of (atomic) statements', propositions so primitive that they require no further explanation (30). In the same sense Blois advocated a 'downward analysis': from high levels corresponding to humans, to low levels corresponding to molecules and atoms (31). Crucial to the reduction of the amount of symptoms, therefore, is the restriction of the descriptions to mere observation (leaving out causal relationships and/or inferences) and to find common denominators for equal reactions of the body or mind, functional or pathophysiological. A second objective was to create a set of attributes describing all various aspects of a symptom; a set of attributes which gives a symptom its specific significance. Preferably this set had to be pertinent and uniform to most of the symptoms.

The human body has only a limited pattern of reactions to illness provoking conditions, substances or infections. This notion has several implications. By renaming the similar symptoms the number can be reduced. Stimulation of the mucous membranes e.g. gives rise to discharge whatever the localisation may be. The aspects of the discharge are equal for all localisations: its character, colour, time of onset, course etc. When discharge is a typical functional reaction of the organism there is a strong motive to name it a symptom as such. The symptom 'Swelling' cannot be differentiated from benign or malign tumors

without causal interpretation by the physician.

The symptom 'Pain' can be fully described by a number of aspects like: localisation, character, irradiation, intensity, seizures, relapses.

Inferences, predictions or causal suggestions in the early phases of the diagnostic process can often lead to confusion and erroneous ways of reasoning. Atomization of the symptoms in medical history leads to an almost uniform set of aspects. They can be listed as follows (in parentheses some aspects that can vary according to the symptom to which they pertain):

-localisation
-character
-intensity
-irradiation (spreading)
-(colour)
-(odour)
-way of onset
-time of onset
-preceded by
-course of localisation
-course of character
-course of intensity
-course of irradiation / spreading
-(course of colour)
-(course of odour)
-(seizures)
-(duration of seizures)
-(distribution of seizures)
-relapses
-combination with other symptoms
-combination with a body function
-worsened by
-improved by
-ahead of

For the symptoms of the physical examination a slightly different set of symptom-aspects was created. Within the framework of an organ system a nearly uniform set of aspects was originated. Similarities across the various types of physical examination could be formulated. It will be examplified for the symptom (Mal) Form(ation) = the description of the Form/Shape and its deviations of existing organs or body structures.

Inspection	Palpation
-localisation	-localisation
-circumference	-circumference
-shape	-shape
-circumscription	-circumscription
-singular/multiple	-singular/multiple
-number	-number
-(im)mobility	-(im)mobility
-symmetry	-symmetry
-enlargement/reduction	-enlargement/reduction
-protrusion	-protrusion
	-pain on pressure
	-ballottement

Remarkably, also laboratory and roentgenologic data could easily be arranged to a limited number of headings, each split up into a limited number of elements. The same can be said for drugs, medical actions etc.

Furthermore, a strict categorisation of the symptoms was introduced, six classes for the diagnostic part, and three for the therapy part. These classes are:

1) Patient's personal, social, mental and environmental background; containing data about age, sex, marital state, education, occupation, finance, etc.
2) Patient's past medical history, with items like: former diseases, operations, allergies, immunisations, child-birth, etc.
3) Patient's present illness not related to the present complaint, e.g. a diabetic patient with a new complaint of backache. This class includes surveillance programs, mobility and activity scales, presently used drugs etc. These three classes combined can be described by 52 various symptoms.

4) Patient's medical history: the class of verbally communicated symptoms. This class is fully described by 68 different symptoms.
5) Patient's physical examination. Traditionally this class is divided into four subclasses: inspection, percussion, auscultation and palpation, to which, for practical reasons, is added a subclass for instrumentally obtained and observed symptoms such as: temperature, blood presssure, ECG, fundoscopy etc. This class contains 141 separate symptoms.
6) Patient's laboratory tests, including data from roentgenologic, bacteriologic, pathologic departments, containing 55 symptoms.

Each symptom and symptom-aspect was coded according to the following attributes :
- positive/negative, depending on being deviant from normal health or not.
- class (of the paper patient)
- organ system(s): see chapter VI
- a figure for the symptom
- a supplement figure for the symptom-aspect.

For the complaint 'pain in the neck' it means the code :

+ D bg 10.03.

The three classes for therapy are:

7) Description of all possible actions a physician can take in therapeutic management and treatment.
8) Description of referral destinations and types of consultation.
9) A detailed list of drugs classified according to their generic names, fysiotherapeutic, ergotherapeutic, and psychotherapeutic treatments and a not exhaustive list of alternative drugs and treatments.

The combination of these classes totals to 157 items.

The total number of items in the 'paper patient' is approximately 6000.

For the research project 9 real case histories were stored into the system. Each case history contained approximately 350 from normal health deviating items. These 3000 items could be stored for over 99%. This almost too favourable result was checked by means of a small test with 42 registered routine patient-physician encounters with three family physicians. Two co-workers (unexperienced family physicians) scored independently the registered data into the system. The result is shown in the next table :

Table 1

STORING CAPABILITIES OF "PAPER PATIENT"

Physician	1	2	3	Total
Number of consults	13	19	10	42
Number of symptoms	69	94	100	263
Number of classifiable symptoms	69	93	98	260
Number of non-classifiable symptoms	0	1	2	3
Classification rate	100%	99%	98%	99%

The described format is standard for all kinds of diseases with exception - thusfar - of the psychiatric ones. For each patient scenario a standard book is used to which the patients' data are added by means of paper slips containing patients' answers to the corresponding symptoms and symptom-aspects. Computerizing of the system can easily be accomplished.

It takes two days to construct a "paper patient" provided all material is collected. This material was acquired by thoroughly interviewing the patients, his or her relatives, family physicians, consulting specialists, nurses, social workers etc.

The choice for the nine case histories was made according to the following criteria :

- prevalence rates
- age categories
- sex
- covering most organ systems.

Excluded from the selection were:

- rare diseases (very low prevalence rates)
- age categories for children because of possible difficulties in taking the medical history directly from the patient; and for elderly people over sixty because of possible multipathology
- psychiatric diseases and disorders.

We have chosen the following categories:

1) 20 - 30 years : man : high prevalence
   woman : high prevalence
2) 30 - 40 years : man : middle-high prevalence
   woman : middle-low prevalence
3) 40 - 50 years : man : high prevalence
   woman : middle ranking prevalence
4) 50 - 60 years : man : high prevalence
   woman : middle-low prevalence

The prevalence rates are all derived from a special epidemiological study of the epidemiological department of the Erasmus University Rotterdam.

The choice of the various organ systems was made according to preliminary results from a classification project among family physicians in Rotterdam (32). These systems are, in a certain prevalence order:
Respiratory tract (11.85%), Locomotor tract (9.03%), Dermatology (6.83%), Urogenital tract (5.19%), Digestive tract (3.99%), Circulatory tract (3.54%), Endocrinologic tract (1.48%), Hematologic tract (0.42%).

From these criteria emerged the following sample of "patients" :

1) Woman, 19 years , with hypochromic anemia. (case 8)
2) Man, 21 years , with atopic eczema (case 3).
3) Woman, 31 years , with extra-uterine pregnancy.(case 4).
4) Man, 34 years , with Asthmatic Bronchitis/(Chronic non-specific lung disaese)(case 7)
5) Woman, 43 years , with gallstones(case 5).
6) Man, 41 years , with Ischialgia/pseudoradicular syndrome (case 2).
7) Woman, 53 years , with hyperthyreoidism. (case 6).
8) Man, 57 years , with myocardial infarction. (case 1).
9) Woman, 47 years , with diabetes mellitus. (case 9).

This last 'patient' was used as an instructive example for introducing the procedure to the participant-physician. The patients were included according to the (sometimes preliminary) diagnoses of their family physicians, who in Holland are the first and only entry into the health care system.

Most important to the simulation was a realistic procedure which comprises
-freely questioning by the physician according to his daily routine;
-an as complete as possible answering the physician's questions;
-answering at a realistic speed; and
-a realistic daily practice routine, which means rather strict time constraints.

The simulation procedure can be described as follows:
After an appointment has been made, the research team consisting of a young inexperienced family physician (the simulator) and a technician (a technically skilful student of social history and communication science) set up the scenery of audio- and videorecording apparatus in the physician's own consulting room.
Meanwhile the physician - participant was instructed in two ways: first, he received some printed instruction material headlining some of the items of the forthcoming procedure. Second, the simulator instructed the physician verbally in detail and answered questions which may have arisen in the candidate. Meanwhile the technical part was completed and the procedure was continued by 'doing' the instruction paper patient. All together, this instruction and the set-up of the technical equipment took 20 minutes at the average. In all cases it satisfied the instructional requirements.
The scenery was arranged in the way the physician was accustomed to. It means that the physician was seated in his ordinary place and the simulator in the place intended for the patient. The videocamera (and the student) was situated at an angle which enabled it to record the physician, his desk, his writing and a digital clock. Questions, answers, hypotheses, probabilities, uncertainty estimates, , diagnoses, prescriptions (the latter five noted on a sheet of paper) could easily be traced in their real order and in conformity with the time sequence. The student was asked to watch over the proceedings.
Three to seven days afterwards, the videotape was commented on by the physician-participants. In the presence of the investigator the physician answered questions about the (face) validity and the reliability of the simulation, the rejection of hypotheses or hypotheses unmentioned, and his estimation of probabilities and uncertainty. He was asked to break down his diagnosis into its characteristic symptoms. (signs, tests, etc.) He then was allowed to denunciate his procedure and outcome.
The candidate had to solve four patient problems within an hour. Although he was free to choose his time, he was at intervals reminded of the other "patients" waiting in the waitingroom.
Sixty-eight physicians participated in the study : 60 family physicians, 8 general internists.
Of the 60 participating family physicians 8 of them solved three patient

problems. Six out of these eight family physicians could not complete the simulation because of urgent calls of real patients: they completed the three patients in an average time of 45 minutes, which is consistent with the average time of the total group. We were conversant with the fact that generally internists take more time per patient in real life. We urged them to do at least three cases, and they all succeeded except one, again for urgent reasons.

The physician's experience of realistic simulation seems to depend on a number of features such as:surroundings, the presentation of the patient's complaint, the recognition of this complaint as well-known, time constraints, and frequency of questioning and answering.

We asked the participants to estimate to what extent the simulation met the routine practice situation. They denoted their estimation on a five point scale ranging from 1 - perfectly comparable - to 5 uncomparable. The results are shown per case history in table 2.

Table 2

PERCEIVED DEGREE OF REALISM OF THE SIMULATION PER CASE

Scale	Case1	Case2	Case3	Case4	Case5	Case6	Case7	Case8
1	17(55)	18(58)	19(58)	19(58)	24(73)	22(69)	23(77)	14(47)
2	13(42)	11(35)	13(39)	14(49)	8(25)	8(25)	7(23)	14(47)
3	1(3)	2(6)	1(3)	-	1(3)	1(3)	-	2(6)
4	-	-	-	-	-	-	-	-
5	-	-	-	-	-	1(3)	-	-
	31	31	33	33	33	32	30	30

(in parentheses the percentages).

The free questioning, the freedom to formulate the question in an accustomed way and in the order the physicians wished and were familiar with, contributed a great deal to this highly favourable result.

On the other hand, this freedom meant quite a hazardous experiment; one fool may ask more than ten wise men can answer. In an preliminary study we got reassuring results. The proof of the pudding is in the eating, especially when the simulator, the unexperienced young physician, was confronted with all kinds of questions, ranging from very complicated (asking several answers from the patient at the same time) to quite irrelevant questions (opinions about situations or about former doctors). At the end of all simulations the young colleague told me confidentially that he now knew how patients can have rather uneasy feelings when questioned by their physician. On the other hand he experienced that physicians become so involved in their task that they completely forgot that it was the simulator and not a real patient they were interviewing. The simulator was, after a hesitating start, able to answer most of the questions. The results are shown in table 3.

Table 3

RETRIEVABILITY PER "PAPER PATIENT".

Symptom-class	A	B	C	D	E	F
Nr of questions	611	431	331	3345	2419	657
Questions unanswered	7	4	2	42	1	5
% questions answered	99	99	99	99	100	99

Total non-classifiable symptoms:6.

The patient's complaint was presented in a verbal as well as in a printed form. This method prevented unmeant emphasis on some of the presented symptoms. According to Ericsson (33)

a) the instruction has to be 'bland' and must not direct the subject to produce specific kinds of information;

b) the instructions must be given in such a way that the subject assigns first priority to performing the task.

The primary complaint always consisted of two uttered symptoms. As the first clues are so important one has to procure a general comparability for all cases (16). It seems noteworthy that practically all participants rephrased the sentence and asked the same symptoms again.

After the presentation of the complaint the time started. With regard to the frequency of questioning and answering an average number of three per minute was reported in literature (34). In this project we found for family physicians an average of 2.9 questions per minute, and for general internists 4.2 questions / minute which underscores the findings of Dudley & Blanchard. A frequency distribution over the cases is showed in the next table

Table 4

FREQUENCY DISTRIBUTION OF QUESTIONS/MINUTE

	Family physicians		General internists	
Questions/minute	Number	Percentage	Number	Percentage
0 -1	4	1.7	-	-
1 -2	22	9.5	-	-
2 -3	93	40.0	2	9.5
3 -4	83	35.8	5	24.0
4 -5	26	11.3	12	57.1
5 -6	1	0.4	1	4.7
6 -7	2	0.9	1	4.7
7 -8	-	-	-	-
8 -9	1	0.4	-	-

As shown in the table the speed of questioning and answering is not influenced by the simulation-method. The method can meet speeds up to 8-9 questions per minute. We did not find any influence for first or following cases as presented to the physician. The order in which cases were presented alternated according to a prefixed scheme.

We asked the participants about their feelings of uncertainty, of being examinated, at the beginning and during the simulation. It was hypothesized that mounting feelings of uncertainty would negatively influence the task performance. They noted down this emotion on two five-point scales, one for the start-position, one for the situation during the process, ranging from 1 : highly uncertain till 5: very certain, and no reaction of this kind. The next table shows the results.

Table 5

DISTRIBUTION OF FEELINGS OF UNCERTAINTY FOR 68 PHYSICIANS

Degree of decreasing Uncertainty	Start	During
1	8	6
2	7	3
3	9	5
4	15	14
5	6	13
no reaction	23	27
Total	68	68

There is a tendency to reduced stress during the simulation session. Most participants said, that they felt almost immediately at ease during the session. An important finding was their communication that their uncertainty was not different from the feelings they experience during normal routine practice.

This leads us to the question of the validity of this 'paper patient' instrument. All simulations are based on the assumption that the workup of the simulated clinical problem is comparable to the actions physicians have taken solving a similar clinical problem with an actual patient. The validity of the simulation therefore, is the comparison between the registered values of the test and the values in the real world. The ideal validity would be:

REGISTERED VALUES = ACTUAL VALUES

One of the basic problems, however, is that a large number of qualities of the actual process is unknown or at least cannot adequately be quantified. In effect, most people assume that there are as many methods and qualities in the medical process as there are doctors. Therefore, a formal judgment about the validity of the simulation cannot be provided. A number of procedures has been originated which give indications about the validity of any measurement. These procedures are :

1. Content - or Face Validity : gives evidence about the nature of the intellectual process which the testee passes through in order to respond to a

simulation. The content, the diagnostic process, in the natural session must be (almost) the same as the one in simulation.

2. Construct Validity (35) : gives evidence about the extent to which the performance of different groups (in these exercises) corresponds with reasonable hypotheses about the degree to which such groups differ. In other words, the theory, the concept on which the construction of the model is based, is tested against empirical values in the real setting.

3. Concurrent Validity : gives evidence about the relation between scores on these exercises and the performance in other tests or settings. The comparison between the measurements of two (or more) instruments about the same elements gives an indication of the validity of (one of) these instruments (36).

4. Predictive Validity (37): gives evidence about a variable that can predict a desired criterion. The correlation between the forecasting variable and the criterion is an operational definition for predictive validity. Simon (33) defines this type of validity as:"If a simulation model of a problem-solving task predicts the relative frequency of different moves, and if the moves are observed actually to occur with approximately those relative frequencies, then to that extent the model is supported by the data - even though there has been no direct observation of the component processes of the model."

The first and the last type of validity are of interest to this project and our instruments. The former type is more or less based on the subjective opinions of the participants. Or, as de Dombal et al. (26) states, "validity" for a simulation system can be acquired by two questions: one regarding the enjoyment of the mode and one regarding its apparent realism. When a simulator can induce behaviour similar to that of real life and the subjects accept the replacement and are motivated to use it, the simulation instrument has succeeded (26).

In the current type of simulation, the relation between the theoretical quality and the empirical variable is very strong: both are based on the same data-base: the medical history of a real patient. Therefore, one must carefully see to it that, while constructing a simulated paper patient, the medical history is a true one. We asked the participants to score their subjective opinions about the mode of the simulation (with regard to the actual processes) and for each specific case. Again we used a five-point scale ranging from 0: no opinion, 1: high resemblance till 5 : no resemblance. The results for the mode of simulation are presented in table 6.

Table 6

DISTRIBUTIONS OF DEGREES OF PERCEIVED REALISM OF SIMULATION MODE FOR 60 FAMILY PHYSICIANS AND 8 GENERAL INTERNISTS

Degree of realism (decreasing)	Total	Fam. Physcn	Gen. Internist
0	1 (2)	1 (1)	-
1	27 (40)	25 (42)	2 (25)
2	37 (54)	32 (54)	5 (63)
3	3 (4)	2 (3)	1 (12)
4	-	-	-
5	-	-	-

(the percentages in parentheses).

For the various patient cases face-validity estimates are presented in Table 7.

Table 7

DISTRIBUTIONS OF PERCEIVED REALISM OF SIMULATION MODE FOR 8 PATIENT CASES

Degree of realism (decreasing)	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
1	19(61)	23(74)	16(48)	12(36)	19(58)	18(56)	22(73)	21(70)
2	12(39)	7(23)	13(40)	18(55)	13(39)	13(40)	7(23)	9(30)
3	-	1(3)	4(12)	3(9)	1(3)	1(4)	1(4)	-
4	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-
	31	31	33	33	33	32	30	30

(the percentages in parentheses).

In the predictive validity the question is whether the simulation mode permits the various moves and strategies predicted in the model(s) (see Chapter VI). The registered data and values enable us to meet our investigational goals.

THE PARTICIPANTS

For reasons of cost and feasibility we chose exclusively physicians from the city of Rotterdam (approximately 600.000 inhabitants, 220 family physicians and 80 general internists) as our population. We chose for an emphasis on family physicians for a couple of reasons:

- they are the most important, main entry in the health care system;
- they cover a - very - broad field of medicine in their daily practice;
- their daily workload and time constraints might force them to adopt (a) strategies(y) different from those of clinical specialists;
- the working methods in family medicine are largely unknown. Research into this form of health care is lacking behind;
- as a consequence of the former item postgraduate training especially modelled for family physicians is still in its infancy;
- I have been a family physician myself, and at present working at the department of family medicine at the Erasmus University of Rotterdam;
- knowledge of the local situation.

Specialists of internal medicine were chosen because of their close relation to the work of the family physician. Their position in health care (authority for consultation and referral) and their task environment (clinical and more advanced setting) place them in a different context. Our hypothesis was that these differences contribute to a different type of strategy as to be found in the problem solving behaviour of general internists. Participants for the investigation were recruted in various ways.

All family physicians in the town of Rotterdam received a letter, including two forms . one for accepting, one for declining (to be specified by particular reasons. We received 61 forms of family physicians agreeing to participate (28%). One of them had to resign because of grave illness. Forty-five (20%) family physicians returned their form of non-participation. The reasons for non-participation concentrated on 7 items:

a) no time available :20 (45%)
b) not interested . 7 (16%)
c) attending to another research : 6(13%)
d) considered the design as insufficient : 5 (11%)
e) resigning from practice . 3 (7%)
f) other reasons 2 (4%)
g) no reason 2 (4%)

The internists, from various hospitals, were personally asked to join the experiment. Eight of the ten internists we approached, agreed to participate. Because of a restricted hypothesis-testing for this group we limited its size for practical reasons.

The family physicians were stratified according to

- age and sex
- spread over the research area
- a patient population between approximately 1500 and 3000
- medical school of graduation.

Because of the total number of family physicians in Rotterdam (220) and our aim to a research sample of at least fifty family physicians we could not strictly adhere to this stratification, but tried however to strive for a representative sample of the population of family physicians.

The criterion 'age' was subdivided into four classes : young, younger middle-aged, older middle-aged, older. This leads to the next table:

Table 8

AGE DISTRIBUTION IN SAMPLE AND POPULATION

Age	Research Sample# Number	%	Fam.Physcns Popul. * Number	%
25-30y.	6	10.0	12	5.5
31-45y.	37	61.7	131	59.5
46-55y.	6	10.0	35	15.9
> 55y.	11	18.3	42	19.1

*figures kindly offered by the Dutch Institute for Family Medicine (N.H.I.), Utrecht.
#figures acquired by a questionnaire

A possible different relation between age and years of experience led to a question about the years really spent in general practice.

Table 9

DISTRIBUTION OF PHYSICIANS ACCORDING TO YEARS OF EXPERIENCE

Years in Practice	Research Sample# Number	%	Population of Fam.Physcns. * Number	%
0 - 5	15	25	47	21.4
6 - 20	32	53.3	103	46.8
21 - 30	8	13.4	44	20.0
> 30	5	8.3	26	11.8

Among the participant-family physicians were three females, representing 15% of the population of female family physicians. It must be kept in mind that not all 20 female family physicians practice in an independent practice situation on a more or less full time basis.

The family physicians were recruted from the various districts of the town.

The time consumption for each participant amounted to approximately 4 hours: 2 hours for the simulation and another two for the re-examination of the video-tape.

To describe the condition of the patient populations per practice a brief description of the Dutch health care system was needed.

Health insurance is covered by two systems : an obligatory state system (National Health) for people with an annual income below $ 14.000.- and a voluntary system of insurance companies for people with a yearly income above this level.

The obligatory insured patients (so-called Sick Fund Patients (S.F.P.) have to choose a family physician themselves. He, the family physician, earns a fixed year income for each SF individual registered into his practice administration. Customarily the private (P.P.) patient behaves in a similar way. Therefore, the Dutch family physician can count on a more or less fixed practice population. The national average of patient population per family physician is approximately 2500, which is in accordance with the figure for the town of Rotterdam. The following table shows the frequency distribution of practices for a scale mounting with 500 patients.

Table 10

DISTRIBUTION OF PRACTICE SIZES IN THE SAMPLE

Number of patients	Nr or Practices
0 - 499	-
500 - 999	1
1000 - 1499	1
1500 - 1999	13
2000 - 2499	17
2500 - 2999	21
3000 - 3499	6
> 3500	1

The mean number of patients per practice in the sample was 2352, ranging from 695 to 3550 patients. The distribution, of 'sick-fund patient' and 'private patients' over the practices is shown in the next Cross-table.

Table 11

CROSS-TABULATION FOR SICK FUND AND PRIVATE PATIENT POPULATIONS IN THE SAMPLE PRACTICES

PRIV.PAT.	0 - 499	500 - 999	1000 - 1499	1500 - 1999	2000 - 2499	2500 - 2999
SICK FUND						
0 - 499	1	-	-	-	-	-
500-999	-	1	-	2	1	-
1000-1499	4	2	-	-	-	-
1500-1999	18	7	-	-	-	-
2000-2499	13	6	2	-	-	-
2500-2999	3	-	-	-	-	-
	39	16	2	2	1	-

The physicians are working in different settings. Most of them work as a solo physician, some of them in a duumvirate, sometimes in a group of three or more family physicians and some of them in health centres. We have no statistics for the types of settings of the Rotterdam family practice situation. Experts' estimates come close to our statistics, which are shown in table 11.

Table 12

DISTRIBUTION OF PRACTICE SETTINGS IN THE SAMPLE

Setting	Number of practices	%
Solo	37	61.7
Duo	13	21.6
Group	3	5
Health Centre	7	11.7

The last stratification criterion was the medical school of graduation. The Netherlands count 8 medical schools. Although we did not estimate the influence of the medical education to be very high, former medical training might influence professional behaviour.

Table 13

DISTRIBUTION OF PARTICIPANTS ACCORDING TO SCHOOL OF GRADUATION

Medical School	Research Sample		General Population	
	Number	%	Number	%
GU	5	8.3	15	6.8
VU	2	3.3	4	1.8
RUG	6	10.0	14	6.4
RUL	15	25.0	65	29.5
KUN	1	1.7	7	3.2
EUR	21	35.0	79	35.9
RUU	10	16.7	35	15.9
Total	60	100	220	100

The codes stand for :

GU = University of Amsterdam
VU = Free University
RUG = University of Groningen
RUL = University of Leyden
KUN = Catholic University Nijmegen
EUR = Erasmus University Rotterdam
RUU = University of Utrecht.

Although acquired coincidentally, the research population appears to be rather representative for the general pertinent population.

Having sketched the tools and the participants, we now proceed to the results.

REFERENCES CHAPTER VII

1. Taylor T.R., Gordon M.J.,: Understanding the clinical management process, in:Sheldon,Brooke,Rector: Decision making in general practice, McMillan, London, 1985.
2. Uberla K.,: Methodological limitations in the analysis of medical activities. Meth. inf. Med., 1980, 19, 7 - 10.
3. Leaper D.J., Gill P.W., Staniland J.R., Horrocks J.C., Dombal F.T. de,: Clinical Diagnostic Process: an analysis. Br. Med. J., 1973, 3, 569 - 574.
4. Ridderikhoff J.,: The paper patient, in:Sheldon,M.G.,Brooke,J.,Rector,A.: Decision making in general practice, London, McMillan, 1985.
5. Noren J., Frazier T., Altman I., Delozier J.,: Ambulatory medical care. A comparison of internists and family-general practitioners. N. Eng. J. Med., 1980, 302, 11 - 16.
6. Weed L.L.,: Medical records, medical education and patient care. The problem oriented record as a basis tool. Chicago, Year Book Medic. Pub. Inc. 1970.
7. Kleinmuntz B.: The processing of clinical information by man and machine, in: Formal representation of human judgment: pages 149 - 186. New York, John Wiley & Sons, 1968.
8. Wortman P.M.,: Medical diagnosis: an information processing approach. Comp. Biomed. Research, 1972, 5, 315 - 328.
9. Mandler G.: in Spence, K.W., Taylor Spence J.T.,: The phychology of learning and motivation. New York,London,Aca.Press, 1967.pages 327 - 372.
10. Miller G.A.,: Psychology: The Science of Mental Life. New York, Harper & Row, 1962.
11. Nisbett R.E., Wilson T.D.,: Telling more than we can know: Verbal reports on mental processes. Psych. Review, 1977, 84, 231 - 259.
12. Slovic P., Lichtenstein S.,: Comparison of Bayesian and Regression approach to the study of information processing in judgment. Org. Beh. Human Perf. 1971, 6, 649 - 744.
13. Ericsson K.A., Simon H.A.,: Verbal reports as data. Psych. Review, 1980, 87, 215 - 251.
14. Kassirer J.P., Gorry G.A.,: Clinical Problem solving: a behavioral analysis. Ann. Int. Med., 1978, 89, 2 - 17.
15. Newell A., Simon H.A.,: Human problem solving, Englewood Cliffs, New Yersey, Prentice Hall, 1972.
16. Palva J.P.,: Measuring clinical problem solving. Br. J. Med. Educ. 1974, 8, 52 - 56.
17. Gill P.W., Leaper D.J., Guillou P.J., Staniland J.R., Horrocks J.C., Dombal de F.T.,: Observer variation in clinical diagnosis. A computer aided assessment of its magnitude and importance in 552 patients with abdominal pain. Meth. Inf. Med., 1973, 12, 108 - 113.
18. Barrows H.S., Bennet K., The diagnostic skill of the neurologist. Arch. Neurol., 1972. 26, 273 - 277.
19. Sage A.P.,: Behavioral and organizational considerations in the design of information systems and processes for planning and decision support. IEEE transactions on systems, man and cybernetics. 1981, 11, 640 - 678.
20. McCarthy, W.H.,: An assessment of the influence of cueing items in objective examinsation. J. Med. Educ., 1966, 41, 263 - 266.
21. Rifkin R.D., Hood W.B.,: Bayesian analysis of electrocardiographic exercise stress testing. N. Eng. J. Med., 1977, 297, 681 - 686.
22. Rimoldi, H.J.A.,: A technique for the study of problem solving. Educ. Psycholog. Measurement, 1955, 13, 450 - 461.
23. Tamblyn R., Barrows H.,: Evaluation Trial of the P 4 system. McMaster University, Monograph 4, 4/78.
24. Tamblyn R., Barrows H.,: Teaching guide for the P 4 system. McMaster

University, Monograph 5, 4/78.
25. Helfer R.E., Slater C.H.,: Measuring the process of solving clinical diagnostic problems. Br. J. Med. Educ., 1971, 5, 48 - 52.
26. Dombal de F.T., Horrocks J.C., Staniland J.R., Gill P.W., Simulation of clinical diagnosis: a comparitive study. Br. Med. J., 1971, 2, 575 - 577.
27. Elstein A.S., Shulman L.S., Sprafka S.: Medical Problem Solving, Harvard University Press. Cambridge (Mass). 1978
28. Gerritsma J.G.M., Smal J.A.,: De werkwijze van huisarts en internist. Een vergelijkend onderzoek m.b.v. een interactieve patienten simulatie, Utrecht, W.U. Bunge, 1982
29. Pauker S.G., Gorry G.A., Kassirer J.P., Schwartz W.B.: Towards the simulation of clinical cognition. Taking a present illness by computer. Am. J. Med., 1976, 60, 981 - 996.
30. Scriven M.: Clinical judgment. In: Engelhardt H.T., Spicker S.F., Towers B.,: Clinical judgment, a critical appraisal. Dordrecht, D. Reidel Publ. Co, 1979.
31. Blois M.S.,: Conceptual issues in computer aided diagnosis and the hierarchical nature of medical knowledge. J. Med. Phil., 1983, 8, 29 - 50.
32. Lamberts, H.,: Incidentie en prevalentie van gezondheidsproblemen in de huisartsenpraktijk. Huisarts en Wetenschap, 1982, 25, 401 - 405.
33. Simon H.A.,: Information processing models of cognition. Ann. Rev. Psych., 1979, 30, 363 - 396.
34. Dudley H.A.F., Blanchard E.B., Comparison of experienced and inexperienced interviewers on objectively scored interview behaviour. J. Clin. Psych., 1976, 32, 690 - 697.
35. Cronbach L.J., Meehl P.E.,: Construct Validity in Psychological tests. Psychol. Bull, 1955, 52, 281 - 302.
36. McGuire C.H.,: Simulation technique in the teaching and testing of problem solving skills. J. Research in Science Teaching, 1976, 13, 89 - 1100.
37. Groot A.D. de, : Methodology. 's Gravenhage, Mouton & cie, 1961.

CHAPTER 8

RESULTS

Introduction

The main objective of this study is to describe the problem solving of (family) physicians. In this two major questions arise:

1) Have (family) physicians (a) specific strategy(ies) along which decisions are reached ? (a decision is viewed as the ultimate point in a task process);
2) Can this(these) strategy(ies) be formalized ?

As we have delineated in chapter VI the former question can be reformulated as:

"Do (family) physicians use a deductive or an inductive method in medical problem solving ?"

The latter question concerns the model(s) and the various landmarks which mark the model(s).

In the preceding chapters we outlined the conceptual framework in order to visualize these questions (chapter VI) and the instruments to achieve the goal(chapter VII). We defined the hierarchical nesting of hypotheses to be a distinctive feature of the deductive strategy. We also defined the typical hypothesis-symptom cyclical course, and the construction of a "hypotheses-base" as a distinctive characteristic of the inductive strategy.

Within the inductive method we modeled three subtypes which might characterize the main methods employed by the various groups of physicians. Landmarks to elicit these subtypes have to be found in our material.

As a complicating factor the task environment of the two groups of physicians differs. The interaction of these two elements, strategies and task environments, could not be predicted but has to be empirically elucidated from the material.

Next to the more traditional characteristics of the medical process we originated a number of factors which may help to specify and to explain the methods and their functioning.

In their psychological model of diagnostic enquiry Elstein and colleagues (1) used three features to account for medical diagnostic reasoning: information search units, cues (symptoms which may lead to a specific, diagnostic, hypothesis), and hypotheses. Cues and hypotheses were used as cross-cutting dimensions in a two-dimensional matrix according to which each simulation was structured. To determine the quality, the effectivity and the efficiency of the problem-solving process, cues as seen related to a particular hypothesis were weighted according to a pre-established arrangement, based on expert's judments. According to McGaghie (2) the selection of the variables seem to support partially the a priori thinking of Elstein et al.(1).

To avoid interpretative variables we looked for objective criteria. Several landmarks within the medical process are submitted to subjective interpretation or can only be established in hindsight. Moreover, several of these features are interdependent. The weighing of a cue is a subjective interpretation relatived to a specific hypothesis and personal judgment; redundancy can only be determined in hindsight; whether the answer of a patient to a particular question will be positive or negative is quite coincidental.

The physician has to operate in an uncertain situation. He does not know the answers beforehand; he can only be aware of the limitations in patient enquiry. Most of the time, however, the information does not come from

interrogation or physical examination but is generated by the hypotheses(3). "The sequence the physician assigns to the variables may originate from hypothesis rather than from observation"(4).

Physicians' initial steps are aimed at a striking reduction of uncertainty(3). Uncertainty and the pressure of the clock urge physicians to look for simplifying and time-saving methods in their daily work. Within certain time-limits the physician has to collect his information,make a diagnosis and decide upon action.

The amount of information is partially based upon the number of positive answers the physician collects (as positive is defined any symptom being deviant from normal function and form). However, whether an answer is positive or negative cannot be foretold. Inclusion of such an element for analysis means introduction of interpretation.

Eventually, we decided on three basic features of analysis:
- hypotheses
- symptoms
- questions.

Apart from the main question of deductive and inductive methods,for each paragraph we shall focus on:
- the subtyping of the inductive strategies;
- the characteristics of these strategies and the inductive method in general;
- the characteristics of the deductive strategy;
- the specifications for the two groups of physicians;and
- the problem-orientation of the problem-solving process.

The order of the following paragraphs are:
1) strategies
2) hypotheses
3) questions,answers,time.
4) probabilities and (un)certainties.
5) patient management and treatment.
6) experience and consequences.

Paragraph 1

THE STRATEGIES

In chapter III we outlined the various ways of reasoning, concluding that only the deductive strategy can bring us the force of proof. In chapter VI, the deductive strategy for the medical process is schematized. Within this schema two major elements can be distinguished. A major characteristic of the hypotheses in the deductive method has been discussed: its progression down the levels towards a testifiable low-level hypothesis.

Two other dimensions may be mentioned: hypotheses and the rejection of redundant hypotheses.

It may be clear that a progression down the levels requires not only the level recognition but also a classification towards the levels. Hierarchical nesting requires at least two hypotheses. This criterion excludes all work-ups of patient cases in which only one hypothesis was generated.

In deductive reasoning the proof is given by the verification or falsification of the lowest-level hypothesis. It excludes the existence of hypotheses equally possible.

At the end of each work-up the physician was asked if he maintained hypotheses different from the diagnosis. All cases with more than 1 hypothesis left after the conclusion of the diagnostic process must be excluded from deductive reasoning strategy.

Generation of hypotheses is assumed to be based on acquired evidence. Elstein et al. (1) found an early hypothesis generation. Although time and

evidence,as a variable of the number of questions, are closely interconnected, the generation of hypotheses cannot be variable of time. (Early) hypothesis generation,therefore, can only be based on acquired information, which will be translated in a number of questions.

Recapitulating the criteria for deductive strategy:

1) DATA GATHERING. Deductive reasoning implies inference from the evidence;it implies the subordination of the real to the realm of the possible.Therefore,an indication of the true nature of the presented problem resulting from acquired information is a necessary condition to the generation of a hypothesis.On empirical grounds we defined the amount of information as the equivalent of at least 10 questions of search through the patient's data base.

2) HYPOTHESIS SPECIFICATION. The progression down the levels (level I the highest, most unspecified;level V the lowest,most specified) has been discussed as the main distinctive feature in deductive reasoning.This criterion is defined as a one-way progression down level I or II to level V.

3) REDUNDANT HYPOTHESES. As a consequence of the deductive reasoning the testing of a final ((lowest-level) hypothesis can only result in a verification or a falsification.A process which results in a number of hypotheses equally possible is incompatible with deductive reasoning.

Applying these criteria to the data base of 253 cases deductive strategy could not be detected. As a consequence of our main research question all cases must fall within the domain of the inductive strategy.

We may remember that the inductive strategy can be distinguished into three types:

a.) pattern-recognition
b.) inductive-heuristic
c.) inductive-algorithmic

As aforementioned the criteria fall within the domains of questions, hypotheses and symptoms.

a.) PATTERN-RECOGNITION

Distinctive to this type is the sudden flash of recognition of a pattern of symptoms on observing only one or two items. It is the 'Aha-Erlebnis' of the recollection of a quite familiar picture one has in mind. Whether this picture is true or false to the facts is not important. It is true to the man or woman who experiences it. We can deduce from this description the following distinctive features:

- The - very - limited amount of data that trigger the 'Aha-Erlebnis';
- the conviction of correctness of the hypothesis;

We define the following criteria for Pattern Recognition:

1) NO QUESTIONS may be asked before the generation of the first (and only) hypothesis. We may remember from the description of the "paper patient" that each presented complaint contained two symptoms. The 'zero' value in the tables refers to active enquiry by the physician;

2) ONE HYPOTHESIS. As a consequence of the description only one hypothesis is generated. The subject-physician does not consider alternatives.

Applying these criteria we found 27 records of it in our data base: 26 referring to family physicians and 1 to a general internist.

To our surprise we found a group of records closely related to the former one, which we called:

HESITATING PATTERN RECOGNITION

This group is related insofar that it is also characterized by the sudden recollection of a familiar pattern. It differs from the former type insofar that the preamble data gathering is more elaborate. Participants ascertained that this cautious character was due to the situation: uneasiness in an observed position. Whether this category must be considered as an artefact of the study cannot be established, because each descriptive study had to rely, one way or another, on some kind of observation. In our opinion, the sudden flash of recognition, the 'Aha-Erlebnis', largely determines the strategy of Pattern-Recognition. Therefore,this type of strategy might be considered as part of the first type.

It contains 26 records, all attributed to family physicians.

The next subtype to be distinguished within the inductive way of reasoning is the:

b.) INDUCTIVE-HEURISTIC

As discussed previously, this type of strategy can be considered as an iterative process on the former theme. In the course of the task performance several patterns as hypotheses jump to mind.These recollections are induced by some data acquired during the problem solving. Information and testing cues are completely mixed up and cannot be separated. The hypotheses collected, gradually, form a data-base from which in the end the physician picks up one of the hypotheses as a diagnosis. The problem-solver regularly visits his 'data-base of hypotheses' in order to add a new one or to reconsider the existing ones. In this sense we can speak of a cycling process as schematized in chapter VI. The hypotheses do not form a hierarchical structure of levels within a certain organ system, but are randomly generated regardless of levels and organ systems. In the end of the diagnostic process a number of hypotheses is maintained, perhaps in order to fall back on alternative hypotheses in the 'data-base' in case of failing predictive outcomes. From these thoughts we can note down the discriminative criteria for this strategy.

1.) NO QUESTIONS ahead of the generation of the first hypothesis.The first step in this strategy reflects the Pattern Recognition mode.In a small minority of cases (six) a limited number of questions was asked in this phase.Because of the following criteria they are enclosed in this subtype.

2) HYPOTHESES. As described,a 'hypotheses-base',a set of hypotheses,is constructed during the work-up containing a varying number of hypotheses;

3) HYPOTHESES-LEVELS. Because the hypotheses were generated randomly the levels do not follow a particular line or structure;

4) DIAGNOSTIC LEVEL. Because the diagnosis is picked up from the 'hypotheses-base' the diagnostic level can take all values from I to V.

We found 172 records, of which 6 records with a more elaborate preliminary data gathering.

It constitutes the largest group in our material, substantially attributed to the working-pattern of family physicians.

c.) INDUCTIVE-ALGORITHMIC

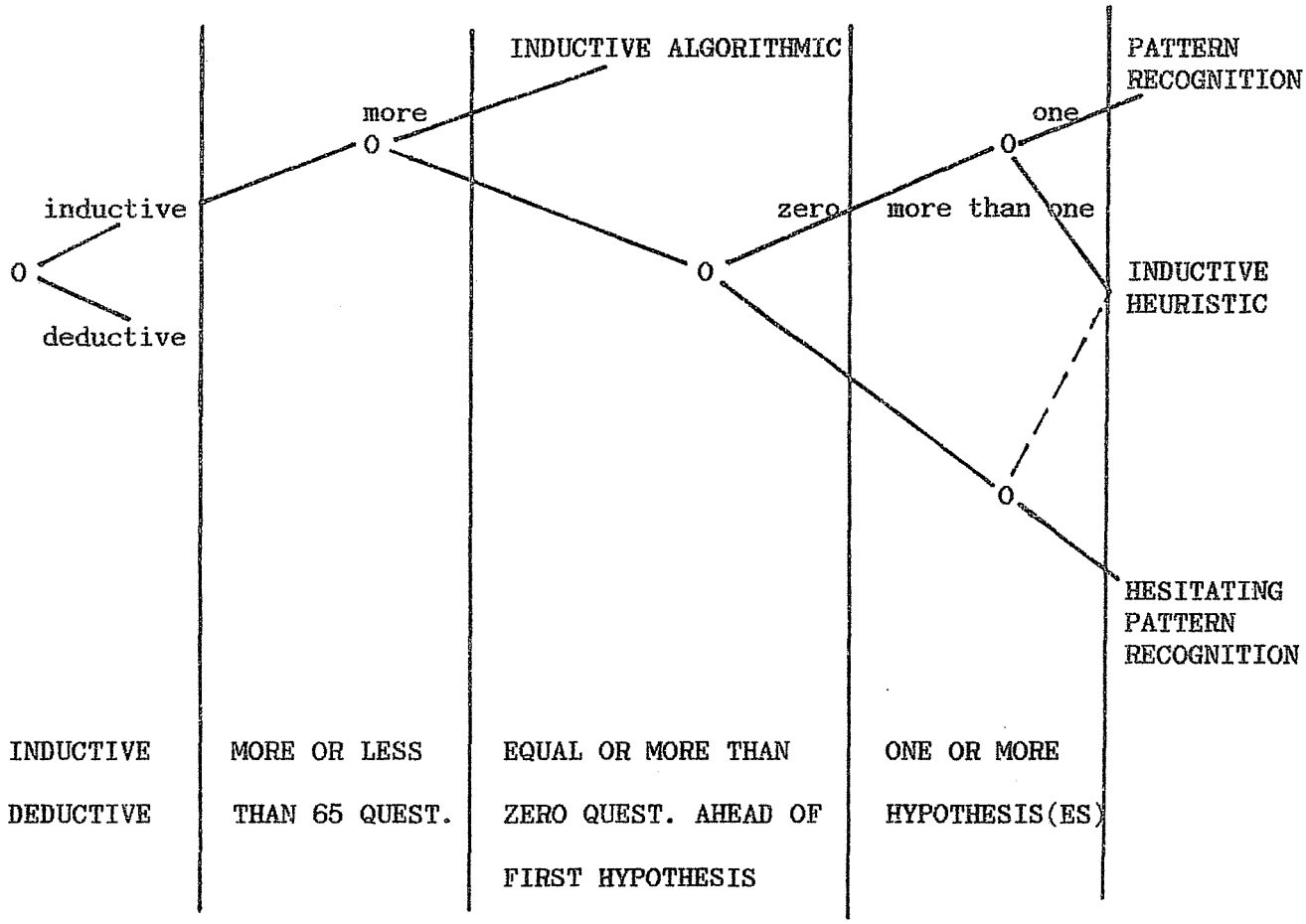
The last type of our model is the Inductive-Algorithmic strategy. In many respects it resembles the former strategy. It differs in the method of data collection. While in the former mode the data collection is directed by intuition, the casual answers of the patient, questions arising at the spur of the moment, in this strategy more or less fixed runs of questions are asked; runs of questions covering most of the organ systems regardless to the organ system to which the patient's complaint refers (and most of the hypotheses can be attributed). It is a way of superficial screening of the patient. It is a kind of habit in which these more or less fixed runs of questions are applied across the various cases. These runs are typically person-bound and can differ in contents and order from physician to physician. Distinctive to this type of

strategy is the total number of questions, which surpasses greatly the number of questions in the former strategies. On empirical grounds, we found the number of 65 questions a decisive point in the determination of the criteria, which can be summarized as follows:

1.) no questions ahead of the first hypothesis;
2.) more than 1 hypothesis;
3.) hypotheses levels ranging randomly from I to V;
4.) diagnostic level may range from I to V;
5.) more than 65 questions asked.

Applying these criteria to our material we found 28 records of which the majority must be attributed to the group of general internists.

The research material was searched in a branching fashion. After the search for records in which a deductive strategy might have been applied, we followed the next schema:



With this procedure we could easily trace the various strategies. Only one record caused a dilemma. It contained 1 hypothesis and 95 questions. This record was selected in the first round and automatically attributed to the Pattern Recognition mode.

After studying this record we decided to classify this record into the Inductive Algorithmic mode mainly because of its nature of sequentially interviewing the various organ systems. This record belonged to the problem-solving process of a general internist.

In paragraph 3, on Hypotheses, we observe 54 'one-hypothesis' cases instead of 53 (the sum of both Pattern Recognition and Hesitating Pattern Recognition strategies).

Table 1

FREQUENCY DISTRIBUTION OF THE INDUCTIVE STRATEGIES

SUBTYPES	ABSOLUTE FREQUENCY	FAMILY PHYSCNS	GENERAL INTERNISTS	RELATIVE FREQUENCY
PATTERN RECOGNITION	27	26	1	10.7
HES.PATT.RECOGNITION	26	26	-	10.3
INDUCTIVE HEURISTIC	172	168	4	68.0
INDUCTIVE ALGORITHMIC	28	12	16	11.0
	253	232	21	100

Subtyping of the Inductive strategy shows a predominance of the Inductive Heuristic strategy, largely attributted to family physicians. The Inductive Algorithmic strategy is largely assigned to the group of the general internists. The Pattern Recognition method is presented in its dual form.

These figures show an application of the Inductive Algorithmic mode by general internists in three-quarters of the cases; family physicians applied the Inductive-Heuristic strategy in three-quarters of the cases. Pattern Recognition seems an almost exclusive prerogative of family physicians.

The strategies distribution over the patient cases is shown in table 2.

The numbers of the patient cases represent the following "diagnoses":

1 Myocardial Infarction

2 Ischialgia

3 Atopic Eczema

4 Extra Uterine Pregnancy

5 Cholilithiasis

6 Hyperthyreoidism

7 Asthmatic Bronchitis

8 Iron-Deficiency Anaemia

Table 2

ABSOLUTE AND RELATIVE FREQUENCY DISTRIBUTION OF THE STRATEGIES OVER THE CASES

PATIENT CASE	PATTERN RECOGNITION	HES.PATTERN RECOGNITION	INDUCTIVE HEURISTIC	INDUCTIVE ALGORITHMIC.	NUMBER
1 row%	7 (22.6)	1 (3.2)	22 (71.0)	31 (3.2)	(100)
2 row%	2 (6.5)	6 (19.4)	18 (58.1)	5 (16.1)	31 (100)
3 row%	6 (18.2)	4 (12.1)	21 (63.8)	2 (6.1)	33 (100)
4 row%	1 (3.0)	0 -	31 (93.9)	1 (3.0)	33 (100)
5 row%	2 (6.1)	11 (33.3)	16 (48.5)	4 (12.1)	33 (100)
6 row%	0 -	1 (3.1)	25 (78.1)	6 (18.8)	32 (100)
7 row%	6 (20.0)	2 (6.7)	19 (63.3)	3 (10.0)	30 (100)
8 row%	3 (10.0)	1 (3.3)	20 (66.7)	6 (20.0)	30 (100)
	27	26	172	28	253

The distribution of the strategies over the cases indicates the variabilitty across the various patient's problems. The case of the problem orientation of the problem solving processes has been substantiated by the available literature. Specific case preference could not be traced.

From table 2 no specific of a certain strategy for a particular case could be detected.Or, reversely, a special case apparently does not require a specific strategy.

Apparently, next to the problem orientation personal preferences play a role.Within the field of personal problem solving behaviour strategy-preferences towards the cases must be assumed.

We listed the occurrences in which a physician used 1 or more strategies for the (4) presented patient cases.

Table 2a

FREQUENCY DISTRIBUTION OF THE PHYSICIANS ACCORDING TO THE NUMBER OF STRATEGIES EMPLOYED.

	NUMBER OF STRATEGIES EMPLOYED			
	1X	2X	3X	4X
NR of PHYSCNS	26	34	7	1
	38%	50%	10%	1%

Of the 68 physicians 38% used consequently 1 strategy for the 4 presented problems.More than 60% of them used 2 or more different strategies.

When the task were mainly determined by the problem, we could have expected a predominance of a particular strategy for that particular case. The variability,however, suggests that personal preferences and capabilities play as much a role as problem orientation. As we shall see in the following sections, the personal touch often dominates the contents and the structures of the medical process.

The cyclic process of the Inductive strategies can be visualized using the feature of the qualitative hypothesis-symptom relationship. Knowing the organ system of the hypothesis, it was easy to attribute symptoms with the same code of an organ system to this hypothesis.

A symptom e.g. coded according to the organ system say 'f' can be attributed to a hypothesis coded with the same letter of the organ system.

A hypothesis is circled by Hypothesis-Related Symptoms(HRS), Hypothesis-Prerelated-Symptoms(HPS), and Non-hypothesis-Related-Symptoms (NHRS)

On the basis of the hypothesis-related symptoms the cyclic procedure can be pictured (by linear stretching of the circle) with the help of nine examples of hypotheses-symptoms sequences.

Table 3

THREE EXAMPLES OF SYMPTOM - HYPOTHESES SEQUENCES

The first example is a case of Myocardial Infarction.

The second example is a case of Extra-Uterine pregnancy.

The third example delineates the process in a case of iron-deficiency anaemia.

CASE 1: MYOCARDIAL INFARCTION

time/minutes	Physician 1			Physician 2			Physician 3		
	HPS	HRS	NHRS	HPS	HRS	NHRS	HPS	HRS	NHRS
1	–	5	–	–	4	–	–	5	1
2	–	5	–	2	3	–	1	1	1
3	–	4	–	3	2	–	3	–	–
4	2	1	–	–	–	2	–	1	2
5	3	–	1	–	2	2	–	–	3
6	2	–	1	1	–	5	3	2	–
7	–	–	3	–	3	–	3	3	–
8	–	–	4	–	4	2	1	1	2
9	1	3	1	2	2	–	5	3	–
10	4	–	–	–	3	–	–	1	–
11	2	–	1	–	4	–	3	–	–
12	1	1	–	3	–	1	2	–	2
13							3	2	–
14							2	1	–
15							5	–	–
16							2	–	–
17							3	–	–
18							3	–	–
19							2	–	–
20									

---- =hypothesis

= diagnosis

= diagnostic hypothesis, i.e. the hypothesis picked from the "hypothesis-base" as the diagnosis.

The entries contain numbers of questions as related to a symptom.

CASE 2: EXTRA - UTERINE PREGNANCY

time/minutes	Physician 10			Physician 11			Physician 12		
	HPS	HRS	NHRS	HPS	HRS	NHRS	HPS	HRS	NHRS
1	2	-	1	2	-	-	-	4	-
2	2	-	-	3	-	-	-	7	1
3	4	-	-	5	-	-	3	1	-
4	3	1	3	-	-	-	3	-	
5	4	5	-	3	3	1	2	2	1
6	-	1	2	-	-	1	-	2	-
7	4	4	-	-	-	3	4	4	-
9	3	3	-	3	3	-	1	-	1
10	4	4	-	1	-	-	-	-	-
11	2	2	1	2	2	-			
12	-	2	-						

------ = Hypothesis

▥ = Diagnosis

▨ = Diagnostic Hypothesis, i.e. the hypothesis picked from the "hypothesis-base" as the diagnosis.

The entries contain the number of questions as related to a symptom.

CASE 3: IRON - DEFICIENCY ANAEMIA

time/minutes	Physician 20			Physician 21			Physician 22		
	HPS	HRS	NHRS	HPS	HRS	NHRS	HPS	HRS	NHRS
1	4	-	-	3	-	1	3	1	-
2	-	2	1	1	2	-	2	-	-
3	1	2	2	-	2	2	-	-	
4	2	2	-	-	-	3	2	1	2
5	1	1	1	-	-	2	2	1	-
6	-	2	-	1	2	2	-	-	5
7	-	3	-	-	1	3	-	-	2
8	-	1	-	1	3	1	3	-	1
9	-	-	3	4	-	-	-	2	2
10	-	-	1	2	-	3	-	2	1
11			-	-	5	-	-	6	
12				-	-	1	-	4	1
13				1	-	1	-	3	2
14				2	-	-	-	5	-
15				1	-	-	-	-	1
16					1	-	-	-	1
17				-	-	3			
18					4	1			
19				-	1	4			
20				-	2	2			
21				-	1	3			

----- = Hypothesis

= Diagnosis

= Diagnostic Hypothesis

The entries contain numbers of questions as related to a symptom.

These examples show a number of characteristics of the inductive reasoning processes in medical problem solving, like:

- hypotheses are sometimes generated upon a small number of symptoms within very limited time-intervals(1 - 3 minutes);
- most first hypotheses were generated immediately after the presentation of the complaint;
- generation of more than one hypothesis at a time appears relatively rare;
- in four of these nine cases the physicians picked a hypothesis as diagnosis from previously generated hypotheses (the physician's base of hypotheses);
- in the majority of cases testing was not performed.
- the pace of questioning differed greatly during the work-up.
- the hypothesis-prerelated symptoms and the hypothesis-related symptoms represent the relevancy of elements in a problem solving task; the non-hypothesis-related symptoms typify the redundancy.

Approximately three-quarters of the total number of symptoms per cycle refer to a particular hypothesis. It indicates a rather stable data acquisition, underlining Feinstein's (4) observation that variables come from hypothesis rather than from observation. The "screening-method" by means of short runs of questions in the Inductive Algorithmic mode contrasts more or less with the other ones.

The next table shows the distribution of the HRS, HPS, NHRS typology over the strategies.

Table 4

FREQUENCY DISTRIBUTION OF HYPOTHESIS(PRE)RELATED SYMPTOMS (HPS, HRS) AND NON-HYPOTHESIS RELATED SYMPTOMS (NHRS) PER CYCLE OVER THE INDUCTIVE STRATEGIES. (MEAN VALUES)

Strategy	Hyp.Prerelated Symptoms		Hypoth.related Symptoms		Total Relat. Symptoms	Non Hypothesis Symptoms	
	Number	%	Number	%	%	Number	%
Patt.Recogn.			16.26	71	71	6.67	29
Hes.Patt.Rec.	6.69	23	17.38	60	83	5.04	17
Induct.Heur.	3.76	30	5.47	44	74	3.24	26
Ind.Algorithm.	5.92	25	8.01	34	59	9.90	41
Total (*)	16.37	19	47.12	53	72	24.85	28

The frequency distribution of symptoms as related to a particular hypothesis, as contrasted to non-relation, shows a rather stable picture for the Pattern Recognition and the Inductive Heuristic strategies (approximately three-quarter of the symptoms being related). The Inductive Algorithmic strategy presents a slightly different picture.
(*): symptoms may be assigned to more than one hypothesis, the preceding and the forthcoming one. In that case the same symptom can be HPS as well as HRS. Addition may surpass the total number of questions.

Redundancy of symptoms is regularly mentioned in literature on medical decision making (5,6,7,8,), and is often claimed for the propagation of information science in medical problem-solving (see also chapter V). In hindsight the authors may be right, but the medical process considered as a search process with an uncertain outcome, redundancy cannot be determined by considering the outcome alone. Redundancy can only be established by denoting the symptoms being relevant or irrelevant to a particular hypothesis. From the examples it may be clear that redundancy, denoted by non-hypothesis-related symptoms, is not an all too pregnant feature in the problem solving process.

We delineated the distribution of the three types of Hypothesis(pre)(non)related symptoms for 7 frequently mentioned hypotheses/diagnoses, standardizing the number of questions to 100, expressing the numbers of the various symptoms in percentages.

Table 5

FREQUENCY DISTRIBUTION OF HYPOTHESIS (PRE)(NON) RELATED SYMPTOMS FOR SEVEN DISEASES IN THE INDUCTIVE-HEURISTIC STRATEGY

DIAGNOSIS	Nr of questions standardized	Mean HPS	Mean HRS	Mean Related	Mean NHRS
Myoc.Infarction	100	52	35	87	13
Ischialgia	100	10	73	83	17
Atopic Eczema	100	8	53	61	39
Gall-stones	100	40	49	89	11
Hyperthyreoidism	100	26	35	61	39
Iron Def.Anaemia	100	28	37	65	35
Extra-Uter.Pregn.	100	56	30	86	14
Total Average		29.4	45.3	74.7	25.3

Redundancy defined as Non-Hypothesis-Related-Symptoms ranges from 11 to 39 %, mean value 25.3%. Three disease entities score markedly higher.

Approaching the question of redundancy by means of the NHRS method, we see for the main hypothesis/diagnosis a redundancy of 25%. However, the markedly different values for Atopic Eczema, Hyperthyreoidism, and Iron-defiency anaemia indicate that redundancy may very well be case-dependent. Redundancy, therefore, can be an attribute of problem orientation. From table 2 we can see that in the cases of hyperthyreoidism (case 6) and iron-defiency anaemia (case 8) 18.8 resp. 20.0% of the chosen strategy is the Inductive Algorithmic one. From its criteria follow an excess of non-hypothesis-related-symptoms.

For the Inductive Algorithmic strategy the figures for the main mentioned hypotheses/diagnoses are:

Table 6

FREQUENCY DISTRIBUTION OF HYPOTHESIS (PRE)(NON) RELATED SYMPTOMS FOR THREE DISEASES IN THE INDUCTIVE ALGORITHMIC STRATEGY.

DIAGNOSIS	Nr of questions standardized	Mean HPS	Mean HRS	Mean Related	Mean NHRS
Gall-stones	100	61	20	81	19
Hyperthyreoidism	100	19	33	52	48
Iron-defic.Anaemia	100	23	47	70	30
Total average		34.3	33.3	67.6	32.4

Similar findings as in the preceding table are to be found in the Inductive Algorithmic strategy. The striking deviant values for Hyperthyreoidism, and to a lesser extent, for Iron-deficiency anaemia, underlines the suggestion of problem-dependent solving processes.

For the Pattern-recognition strategy only onne disease was mentioned a sufficient number of times to allow statistics:

Table 7

FREQUENCY DISTRIBUTION OF HYPOTHESIS(PRE)(NON)RELATED SYMPTOMS FOR ONE DISEASE IN THE PATTERN RECOGNITION STRATEGY.

DIAGNOSIS	Nr of questions standardized	Mean HPS	Mean HRS	Mean Related	Mean NHRS
Myocard. Infarct.	100	-	80	80	20

This result, although of restricted value, underlines the predicted functioning of the Pattern Recognition method.

Opposed to redundancy is Relevancy. The relation of a particular symptom to a hypothesis (a disease pattern) is essential to the eliciting of a diagnosis from the patient's data base. The more symptoms which can be related to a particular hypothesis in the least number of questions the more relevant the medical enquiry is.

We designed a new variable, the Relevancy Factor , which can be drawn as:

$$\frac{HPS + HRS}{HPS + HRS + NHRS} * 100$$

With the help fo this factor we are able to test two hypotheses:

1) The factor scores higher in risk-avoiding than in risk-seeking people; risk-avoiding typified by Inductive Heuristic strategy, risk seeking by Pattern Recognition.

2) The factor scores lower in relatively less frequent, less prevalent, diseases.

The risk-avoiding physician cautiously gathers information, heuristically searching for "filling the pattern he has in mind", and for plausible explanations of the presented symptoms. The risk-avoiding search method evades lingering in less promising directions.

This behaviour must be differentiated from the Inductive Algorithmic mode in which 'redundancy' is a 'built-in' characteristic. In contrast, the physician using 'real' pattern-recognition is quite determined and does not avoid more 'risky' questions. He knows that he is on the right track, and less favourable answers do not shake him in his conviction.

The results are shown in Table 8.

Table 8

RELEVANCY FACTORS FOR THE INDUCTIVE SUBTYPES.

STRATEGY	RELEVANCY FACTOR*
Pattern Recognition	71.65
Hes.Pattern Recognition	86.16
Inductive Heuristic	77.30
Inductive Algorithmic	69.75

* percentage of symptoms related to a hypothesis. The relevancy of the medical enquiry process does not significantly differ for the various Inductive strategies. The runs of 'unrelated' questions in the Inductive Algorithmic strategy apparently produce in the end a level of relevancy equal to that of the other strategies.

The results scarcely corroborate the first hypothesis. Evidently, the relevancy factor does not suit an appropriate distinction among the strategies. But on the other hand, a relevancy of the diagnostic enquiry process of 76% on the average is an encouraging finding for the physician's ability and efficiency in data acquisition. In the Pattern-Recognition method and in the Inductive Heuristic mode this result stems from alert reaction of the physician to the patient's answers. In the Inductive Algorithmic way this result is accomplished by asking more or less fixed runs of questions in a fast pace. In the end both methods score approximately equal.

Less prevalent diseases are assumed to be less familiar to physicians. Physicians may be handicapped by their lack of ready knowledge or gaps in memory. It is asserted that the data collection in these cases adopts a less determined character. The search through the patient's 'data-base' is wavering, often leading to less relevant directions. In case of familiar types of symptom-arrangements physicians know how to react, how to ask, how to combine to a seemingly valid hypothesis. These are what physicians call the 'easy cases'.

The results are shown in table 9.

Table 9

RELEVANCY FACTORS FOR THE DIFFERENT PATIENT CASES.

PATIENT CASE	RELEVANCY FACTOR
1) Myocardial Infarction	79.22
2) Ischialgia	73.68
3) Atopic Eczema	76.28
4) Extra-uterine Pregnancy	79.83
5) Gall-stones	83.21
6) Hyperthyreoidism	61.29
7) Asthmatic Bronchitis	85.29
8) Iron-deficiency Anaemia	69.37

The relevancy factor shows little variation across the cases. Apparently, this factor cannot be related to the (difficulty of the) problem.

The Relevancy Factor can be viewed as a consistent characteristic of the hypothesis - symptom relationship within a medical problem solving process. It underlines the intimacy of this relationship and its meaning for the reasoning process of the physician.

The results of table 9 also indicate that when variability of medical enquiry is due to problem difficulty, this difficulty is hard to establish. Among other problems difficulty seems to depend on factors like prevalences or incidences, the life-threatening or incapacitating nature of the disease, the physician's personal knowledge, and (recent) attention which has been paid to a particular disease or case.

In the research design we judged the cases 2 (low back pain, Ischialgia), 4 (extra-uterine pregnancy), 6 (hyperthyreoidism) and 8 (iron-deficiency anaemia) as the more difficult problems, in comparison to the odd-figured ones. Apart from the cases 6 and 8 the relevancy factor does not fully discriminate for this aspect. An explanation for case 2 might be that, although the diagnostics for diseases and ailments in this body region are rather difficult, the physician is quite acquainted with these cases and knows how to ask relevant questions. As we shall see in paragraph 3 of this chapter the mean number of questions and the average time until diagnosis was reached (per case) creates a better fit for our preliminary judgment of problem difficulty.

Although the Relevancy Factor does not discriminate for problem difficulty, it can elegantly determine the physician's effectivity in medical data collection.

Pararagraph 2

HYPOTHESES, DIAGNOSES, AND SYMPTOMS

Hypotheses are considered to be the key-stones in reasoning processes. In the diagnostic reasoning the explanation of the observed symptoms and signs and testing of the explaining hypotheses play a predominant role (9). They help to overcome limitations of memory capacity and serve to narrow the size of the problem space (1). Questions asked by a physician taking the history of the present illness are prompted by specific hypotheses, which are prompted in turn by patient's data. But the statement does not explain what hypotheses are. Are they the 'bold statement' in the 'Popperian sense? Are they the logical causal explanation of the observed symptoms? Or, are they the 'chunks': the cluster of bits, as Miller hypothesized in his theory on the limited channel capacity of the human memory? (10). We do not know, although the Miller theory is a tempting one which may explain a number of qualities we found in our study. It might explain the variance within the problem-solving performances of our case and the highly personal character of the hypotheses. Defining hypotheses in a general way does not permit an explanation of their variation in medical problem solving. It is my intention to present a number of characteristics and qualities as they emerged in the problem solving behaviour of the participating physicians.

In the 253 presented problem-solving tasks 745 hypotheses were generated, including the diagnoses. Obviously, one patient scenario stimulated the generation of hypotheses more than another. The less prevalent illnesses (case 4 and 6) incited more hypotheses than more frequently occurring disease entities.

Table 10

ABSOLUTE AND AVERAGE NUMBER OF HYPOTHESES PER CASE FOR THE INDUCTIVE HEURISTIC AND ALGORITHMIC STRATEGY.

Cases	1	2	3	4	5	6	7	8
Number of Hypotheses	85	96	74	129	86	109	67	99
Average Number per case	3.7	4.2	3.2	4.0	4.3	3.5	3.0	3.8

The number of hypotheses differs across the cases. Realizing that the cases 2, 4, 6, and 8 are assumed to be the more difficult ones, a relationship between the average number of hypotheses and problem difficulty cannot be observed.

(*for explanation of the figures (1-8)see page 179).

In deductive reasoning the hypotheses successively generated during the inference process form a hierarchical structure.The hypotheses can be deduced from higher to lower levels of specification in a strict logical way (see the example of the free-falling objects in chapter III).

For our research purpose we constructed a multi-level scheme in order to detect hierarchical nesting and thus the possible existence of a deductive strategy.The multi-level scheme is recapitulated in a concise form:

Level I :all matters affecting the total human being;

Level II :multiorganic formulation:diseases and illnesses concerning body systems and body parts;

LevelIII :organ system,the nomenclature following the usual medical classification as defined for these structures;

Level IV :organ:diseases and illnesses of specific organs;

LevelV :"cellular level":the main diagnostic entity,the most specified bio-pathological state.

Although we could not find a deductive strategy according to our criteria, we nevertheless were interested in the distribution of the hypotheses over the levels.(N.B. hierarchical nesting can only be established by screening each record).

Table 11

FREQUENCY DISTRIBUTION OF HYPOTHESES OVER THE LEVELS.

LEVEL	NR of HYPOTHESES	PERCENTAGE
I	30	4
II	100	13
III	110	15
IV	229	31
V	276	37
Total	745	100

From this table we can see a gradual increase in numbers of hypotheses the lower the level, as can be predicted from the theory.

Generally,during the work-up of a case a - considerable - shift from high level (I,II) to low-level (IV,V) hypotheses is expected.In the medical deductive way the problem solving procedure can be viewed as an increasing specification of observed (clusters of)symptoms to a specific clinical entity.

As we argued before,the sequence of inductive hypotheses cannot be viewed in this way.Although a certain specification of the generated hypotheses may take place during the diagnostic procedure,these hypotheses must be seen as patterns in mind,regardless of their state of specification.

Focusing on the first and last hypotheses we found,related to the type of strategy:

Table 12.

DISTRIBUTION OF THE LEVELS FOR THE FIRST AND LAST HYPOTHESIS IN THE INDUCTIVE STRATEGIES.

STRATEGY	FIRST HYPOTHESIS					LAST HYPOTHESIS				
	LEVEL					LEVEL				
	I	II	III	IV	V	I	II	III	IV	V
Patt.Rec.	-	4	3	9	11	-	4	3	9	11
Hes.Patt.Rec.	-	2	-	16	8	-	2	-	16	8
Induct.Heur	12	32	29	56	43	1	11	20	55	85
Ind.Algorith.	3	6	3	4	12	1	1	4	8	14

A slight shift from the higher (I,II) levels to lower (IV,V) levels of hypotheses is shown in this table. Seemingly, the hypotheses does not follow the rules of deductive reasoning.

Because of the one-hypothesis style there is no shift towards higher level hypotheses in the pattern-recognition methods. When we take the number of the hypotheses of the less specific levels (I, II and III) together, a shift towards specificity can be observed. The levels I, II and III decreasing in numbers from 42% to 21%, and more the specific hypotheses increasing from 58% to 79%. Because of the relatively small numbers of the pattern recognition method it is difficult to draw some conclusions from the differences between the less (P.R: 26%, HPR: 8%) and the more specific levels (PR: 74% - which is rather consistent with the the inductive types -, and HPR: 92%). We may assume that the preliminary questioning before generating the only hypothesis in the work-up leads to more specification of the disease in question.

The conception of "level-less" patterns as inductive hypotheses is shown in table 13.The levels of the diagnoses for the eight cases,which may vary within each patient work-up,show a marked distribution over the levels.The figures represent the numbers of diagnoses named for the patient cases.

The numbers between the brackets represent the heterogenous diagnoses for that particular case within that particular level.The table gives the impression of diagnoses more or less scattered over the levels. In total, 77 different diagnoses in 253 records have been found; in 9 cases more than 1 diagnosis was mentioned.

Table 13.

DISTRIBUTION OF THE LEVELS FOR THE DIAGNOSES OF THE EIGHT CASES

PATIENT CASE	DIAGNOSTIC LEVELS				
	I	II	III	IV	V
1	-	-	-	21 (3)	10 (3)
2	1 (1)	2 (2)	9 (3)	19 (8)	-
3	-	12 (7)	5 (3)	16 (6)	-
4	-	1 (1)	2 (1)	2 (1)	28 (5)
5	-	-	-	10 (4)	23 (3)
6	-	1 (1)	2 (2)	2 (2)	27 (1)
7	-	-	9 (9)	6 (1)	15 (2)
8	1 (1)	2 (2)	-	24 (4)	3 (1)
Total	2 (2)	18 (14)	27 (18)	100 (28)	106 (15)
Percentage	0.8	7	11	40	42

The figures show the numbers of diagnoses made for each patient case per level. The numbers between brackets represent the difference of diagnostic nomenclature. E.g. 21 diagnoses, level IV, case 1 stand for 3 different diagnostic names; within level III 27 diagnoses are composed of 18 diagnostic names.

Even on level V there are twice as many various diagnoses as there are patient cases. The one diagnosis for level I stands for "some somatic disease". For cases 2 and 3 a highly specified diagnosis could not be reached.
The recognition of dermatological symptoms (case 3) placed many family physicians for unsoluble problems. Indeed, many of them did not recognize the case and made a diagnosis on level II: skin eruption.
The low back pain case (2) placed many physicians in a serious dilemma. This kinds of complaints is so often heard unjustly, that many physicians mistrust the patient's answers, and limit their examinations to some obligatory tests. But it may be possible, as one of the participants observed, that deciding upon a diagnosis is a very hard job indeed,especially in this part of the body, as the complaints are also often caused by mental states.The more "difficult" (less prevalent) diseases (cases 4 and 6) led to more highly specified diagnoses. It might be that the detailed description of these diseases, their highly specific tests and their 'logical' structure may more easily lead to 'high-leveled diagnoses'. In the case of the iron-deficiency anaemia (case 8) the participating physicians often diagnosed 'anaemia' but hes tated to choose the type of anemia. The asthmatic bronchitis case (7) is one of the most remarkable ones. Although this disease gets very high prevalence rate (in literature percentages of 10-15% of the total population are mentioned), ca 50% of the diagnoses fall within the level III i.e. rather unspecified An explanation for this phenomenon is hard to give, although it may be that the consequence of further diagnostics does not outweigh the scope of treatment. Broad-spectrum antibiotics replaced specification of diagnosis. Obviously, the cases 1 (myocardial infarction) and 5 (gallstones) are so straightforward that there are hardly any difficulties for the physicians.
Preliminary to the 77 various diagnoses a large and varying collection of hypotheses was generated. As mentioned before, each "patient case" gave rise to a large number of hypotheses. Although each patient scenario was designed within the context of one organ system, hypotheses pertaining to a case were generated in the framework of several organ systems. The - almost confusing - distribution of hypotheses over the levels, is also true for the distribution over the various organ systems.

The next tables show the frequency distribution of hypotheses over the organ systems for each of the patient cases.
The hypotheses were arranged into the order in which they were generated (1-7):

Table 14:

HYPOTHESES ARRANGEMENT OVER THE ORGAN SYSTEMS PER PATIENT CASE.

CASE 1: MYOCARDIAL INFARCTION

ORGAN SYSTEM	SEQUENCE OF GENERATION							
	1	2	3	4	5	6	7	%
Circ.system	28	12	10	8	2	-	-	71
Digest.system	3	4	2	-	-	-	-	11
Locomotor syst.	-	-	-	-	1	-	-	1
Respir.system	-	7	4	2	-	1	-	16
Total	31	23	17	10	3	1	-	85/100%

CASE 2: ISCHIALGIA

ORGAN SYSTEM	SEQUENCE OF GENERATION							
	1	2	3	4	5	6	7	%
Circul.system	-	2	-	1	1	-	-	4
Centr.Nerv.system	12	9	10	8	3	3	-	47
Dermatol.system	1	-	-	-	-	-	-	1
Digest.system	-	-	1	-	-	-	-	1
Somatic disease	1	1	2	-	1	-	-	5
Locomotor syst.	17	10	6	2	1	1	-	39
Psychosocial	-	1	1	-	1	-	-	3
Total	31	23	20	11	7	4	-	96/100%

CASE 3: ATOPIC ECZEMA

ORGAN SYSTEM	SEQUENCE OF GENERATION							
	1	2	3	4	5	6	7	%
Dermat.system	30	17	8	6	1	1	-	85
Somatic disease	-	1	-	-	-	-	-	1
Endocrin.system	1	-	-	-	-	-	-	1
Respir.system	2	5	2	-	-	-	-	12
Total	33	23	10	6	1	1	-	74/100%

CASE 4: EXTRA-UTERINE PREGNANCY

ORGAN SYSTEM	SEQUENCE OF GENERATION							
	1	2	3	4	5	6	7	%
Digest.system	15	6	2	1	1	-	-	19
Somatic disease	3	1	-	-	-	-	-	3
Respir.system	-	-	1	-	-	-	-	1
Uro-genital syst.	15	25	25	18	10	4	1	76
Total	33	32	28	19	11	5	1	129/100%

CASE 5: CHOLELITHIASIS (GALL-STONES)

ORGAN SYSTEM	SEQUENCE OF GENERATION							
	1	2	3	4	5	6	7	%
Digest.system	32	19	14	9	3	2	-	92
Somatic disease	-	-	1	11	1	-	-	3
Uro-genit.system	1	1	2	-	-	-	-	5
Total	33	20	17	10	4	2	-	86/100%

CASE 6: HYPERTHYREOIDISM

ORGAN SYSTEM	SEQUENCE OF GENERATION							
	1	2	3	4	5	6	7	%
Circul.system	-	-	-	1	-	-	-	1
Digest.system	1	6	2	1	-	-	-	9
Somatic disease	18	3	4	1	1	-	-	25
Endocrinological	9	18	11	9	7	-	-	50
Respir.system	-	1	-	-	-	-	-	1
Uro-genital system	1	-	1	-	-	-	-	2
Blood	-	-	2	2	-	-	-	4
Psychosocial	3	2	3	2	-	-	-	9
Total	32	30	23	16	8	-	-	109/100%

CASE 7: ASTHMATIC BRONCHITIS

ORGAN SYSTEM	SEQUENCE OF GENERATION							
	1	2	3	4	5	6	7	%
Circul.system	-	2	-	-	-	-	-	3
Somatic disease	1	2	-	-	-	-	-	4
Locomotor syst.	-	1	-	-	-	-	-	1
Respiratory syst.	29	17	11	3	1	-	-	91
Total	30	22	11	3	1	-	-	67/100%

CASE 8: IRON DEFICIENCY ANAEMIA

ORGAN SYSTEM	SEQUENCE OF GENERATION							
	1	2	3	4	5	6	7	%
Circul.system	1	4	1	1	-	1	-	8
Centr.Nerv.syst.	-	1	1	-	-	-	-	2
Digest.system	1	2	3	3	1	-	-	10
Somatic disease	5	3	2	1	1	-	-	12
Endocrinol.syst.	2	3	3	-	1	-	-	9
Respirat.system	-	1	1	1	-	-	-	3
Blood	14	8	7	6	3	-	1	39
Psycho-social	7	4	1	1	-	-	-	13
Total	30	26	21	14	6	1	1	99/100%

The physician must often choose among - clusters of - symptoms overlapping more than one organ system. The symptom Dyspnea (breathlessness) can refer to cardiac as well as respiratory dysfunctioning. The presented tables might erroneously give the impression that physicians are 'wild guessers'. Apart from the cases 6 and 8 this impression is unjustified. Taking into consideration the pertaining organ system and its adjacent one we can delineate the next table, showing the relevancy of the hypotheses to the presented case (as percentages of the total number of hypotheses generated for the case):

Table 15:

RELEVANCY OF THE HYPOTHESES PER CASE.
(proportional to the total number of hypotheses per case)

PATIENT CASE	1	2	3	4	5	6	7	8
Hypotheses pertaining	68%	44%	82%	74%	90%	48%	90%	35%
Hypotheses of adjacent organ systems	18%	40%	15%	21%	7%	9%	4%	4%
Total	87%	84%	97%	95%	97%	57%	94%	39%

Hypotheses can pertain to more than one organ system. The sum of the hypotheses within a certain organ system and in its adjacent one gives a more realistic impression of the hypothesis-relevancy towards a problem solving process.

Again, the cases 6 and 8 represent the exception. In the case of hyperthyreoidism the presented symptoms sometimes led the physician to erroneous but sensible directions such as a more common endocrinologiccal disease, diabetes mellitus, or the more frightening perception of carcinoma. The latter direction represents almost 25% of all hypotheses mentioned in this case. Far more difficult to explain is the case of hypochromic anaemia: a young girl with iron deficiency mainly caused by profuse menstruation. Many physicians failed to inquire specifically after the menstrual cycle. These physicians missed the clue which often led to - almost desperately - extensive searching through a great variety of organ systems.

The complaints also play a part in the generation of relevant hypotheses. Some symptoms evoke great uncertainty. Symptoms like tiredness (case 8) or vague pain in the leg (case 2) can point to various directions. A physician choosing the right direction immediately can accomplish his task in a minimum of time. However, when he chooses some blinded-ended course, which can easily be the case when confronted with multidirectional symptoms, the problem solving task can become bothersome and less motivated.

In the inductive schema we modelled a cycling process circling round (a) hypothesis(es).In our material we found a total of 7 rounds.After the first round the physicians who employed the Pattern-Recognition methods (plus one of the Inductive Algorithmic strategy:see preceding paragraph) had finished their task.With each following round a number of physicians finished their task.This feature can be more or less read from the decreasing number of hypotheses in the next cycles.Table 16 pictures the decreasing number of hypotheses,reflecting the physicians who had accomplished their work-ups.

Table 16.

ABSOLUTE AND RELATIVE NUMBER OF HYPOTHESES PER CYCLE OF THE SOLVING PROCESS.

CYCLE	NUMBER OF HYPOTHESES	PERCENTAGE
1	253	34
2	199	27
3	147	20
4	89	12
5	41	5
6	14	2
7	2	0.2
TOTAL	745	100

With the increase of the number of cycles the number of physicians still working at their tasks decreases. This is reflected by the decreasing number (and percentage) of the hypotheses per cycle (of the inductive inference).

As a consequence of the definition Pattern Recognition strategies contain only one hypothesis. The remaining Inductive strategies contain more than three hypotheses on the average. The figures are presented in table 17.

Table 17.

ABSOLUTE AND MEAN NUMBER OF HYPOTHESES PER SUBTYPE OF THE INDUCTIVE STRATEGY.

STRATEGY	NR of HYPOTHESES	MEAN NUMBER PER WORK-UP	STANDARD DEVIATION
PATTERN RECOGNIT.	27	1.0	0.0
HES.PATT.RECOGN.	26	1.0	0.0
INDUCT.HEURIST.	585	3.4	1.2
INDUCT.ALGORITHM.	107	3.8	1.4

On the average 3.5 hypotheses were generated in the "multi-cycled" inductive strategies. A significant difference between the heuristic and the algorithmic strategy could not be found.

As mentioned in our introductory paragraph we linked hypotheses to questions as non-interpretative data. In our opinion a certain amount of data is required to produce a picture in the mind. When this assertion is true, each cycle in the process should contain a more or less constant number of questions. The next table shows the results from our material.

Table 18

	Mean number of questions
Cycle 1	17,20
Cycle 2	16,64
Cycle 3	15,46
Cycle 4	15,47
Cycle 5	17,44
Cycle 6	18,67
Cycle 7	1,00

A more or less constant amount of data is required to generate a pattern in the mind. This amount of data can be represented by the number of questions.

Cycle 7 cannot be reckoned as a full-scale cycle, while it concludes the diagnostic process with one question testing a reached diagnosis. The figures seem to confirm the statement. The question whether the amount of clinical data can really be represented by the simple counting of questions, is a matter of future research.

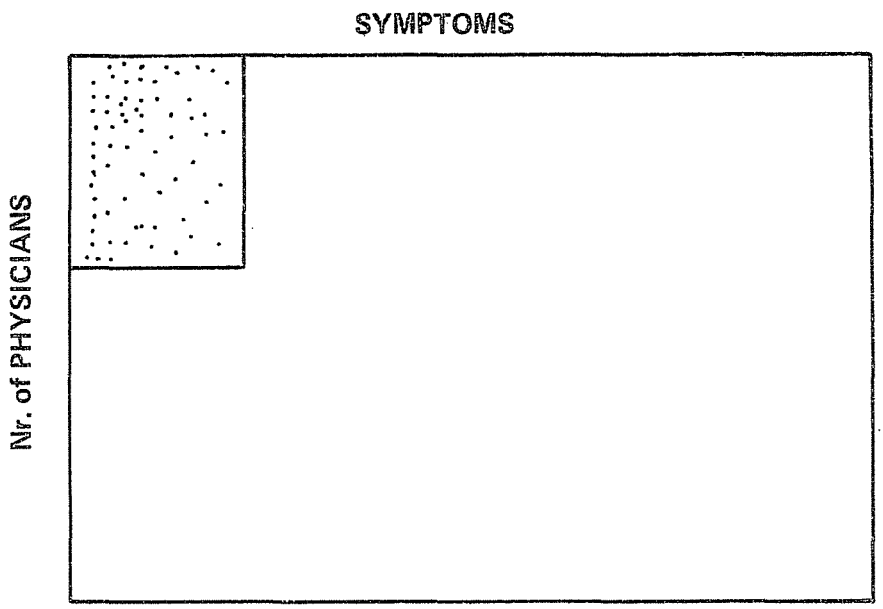
This does not include the question whether a certain number of symptoms can raise a hypothesis. One symptom can be questioned several times. Also, a symptom can contain more than 1 attribute.

If a hypothesis is conceived as a cluster of symptoms it is interesting to know whether these clusters are consistent across the various physicians who generated the same hypothesis. With the help of a computer-program the symptoms (and symptom-aspects) as mentioned in the problem solving tasks were clustered for a number of more frequently named hypotheses.

When symptoms were asked more than once, e.g. by asking different aspects of the

same symptom, they were denoted as 1 symptom. Figures less than five in the scatterdiagrams refer to these situations.
The figures in the schemes refer to the various symptoms, the letters to the physicians who asked these symptoms. We shall present four examples from different patient cases.
First example: myocardial infarction.
This hypothesis (diagnosis) was mentioned by 23 subjects, asking 283 symptoms of which 58 were different. We included in the cluster-program all symptoms mentioned more than 5 times in the data-base for this hypothesis, which was the case with 12 various symptoms. These 12 symptoms were asked by 12 out of 23 physicians. This means that in the space of symptoms asked by all 23 physicians, the 'cluster-space' represents only 11% of the total area.

FIGURE 1



The following scattergrams show the distribution of the symptoms within the 'cluster-space', each cross representing a symptom. None of the symptoms was unanimously named by the physicians, and for only 3 symptoms agreement of more than 80% was reached. For all other symptoms agreement among physicians counted for 50% or less. The three main symptoms were: pain, blood pressure and pulse rate.

SCATTERDIAGRAM 1

```
     abcedfghijk
1    +++++ ++++
2    ++++ +++  +
3    +++++++ +
4    ++++++++
5    ++++++    +
6    +++ ++ + +
7    +++    +++
8    ++++      +
9    +++ +
10   +         ++
11   ++
12    ++
```

Second example : Bronchitis.
18 physicians asked 229 symptoms of which 56 were variant. 14 symptoms (mentioned more than 5 times in a case) were found in 15 physician-work-ups. The 'cluster-space' represented 20% of the total symptom-space. None of the symptoms reached unanimity.
On two symptoms 66% of the physicians agreed, the rest counted for an agreement of 50% or less.
The two main symptoms were: quality of sputum and coughing.
SCATTERDIAGRAM 2

```
     abcdefghijklmn
1    +++  +++++
2    +++++++ +
3    + + ++   + ++
4    ++ +++   ++
5      + +++++ +
6     +++   +     +
7    ++       ++
8    ++ +      +
9      ++  + +
10   +++    +
11   ++         ++
12   ++ +   +
13   + + + +
14         +      +
15    + +
```

Example 3 : Hyperthyreoidism.
27 Physicians asked 543 symptoms of which 115 symptoms were different. 28 symptoms (mentioned more than 5 times in a case) were found in 23 physician-work-ups. The 'cluster-space' was 20% of the total symptom-space. One symptom reached an agreement among the physicians of 70%, the rest of the symptoms 50% or far less. The main symptom was: the hormone thyroxine level in blood.
SCATTERDIAGRAM 3

```
     abcdefghijklmnopqrstuvwxyz#*
1    +++ ++++++  +++ ++++   + +
2    ++++++ +++ +++   +++++ +
3    +++++++ + ++++++++++  ++
4    ++++ ++++  + ++++++
5    +++  +++  +++    +  ++  ++
6      +   +  ++           +  +++
7    +++ ++++++ +        + + +
8    ++  + + ++  +  +   +
9          ++   +        + +   ++
10   ++++ +     +  + +  ++
11   ++ + +        +++    +
12   +  +         ++ +     ++
13   ++++     +                +
14       +   +              +++
15   ++++           ++
16   +  +         +          ++
17   +   +    ++ +
18       +  +  +
19       +  +    +
20          +   +              +
21   +  +
22   +  +
23         +
```

Example 4 : Hypochromic Anaemia.

25 physicians asked 407 symptoms of which 93 symptoms were variant. 22 symptoms (mentioned more than 5 times in a case) were found in 22 physician-work-ups. The 'cluster-space' was 20%. One symptom reached an agreement of 63%, the remaining of the symptoms came to 50% or far less (9%). The main symptoms is - as expected - the blood-haemoglobin level.

SCATTERDIAGRAM 4.

```
     abcdefghijklmnopqrstuv
1    +++++++++++ +
2    +++++        +++++
3     +++ + +   + + +  ++
4    ++++ + + ++        +
5    +++ ++++        +   +
6    +++ +      +++ +  +
7    +   ++++    +      + +
8    +   + +       + ++  ++
9     +++ ++ +    +
10   +++++             +
11   +     ++ +++
12    +++ + ++
13   +  +  +    +     ++
14   +   + +  ++
15    ++      ++  +
16      + + ++      +
17           +   ++   +
18   ++            +
19   +   +                +
20     +        +   +
21   +
22           +
```

These examples were chosen because they represent different types of diseases, because they represent diagnostic hypotheses, and because they contain symptoms mentioned more than 5 times.
In the majority of cases the symptoms varied so much that no real clusters could be found.

In contrast with the rather consistent method of working,the contents of diagnostics seem rather inconsistent. Physicians seem to agree upon only one or two symptoms within an extensive space of symptoms. A possible explanation for this phenomenon might be that most physicians are really looking for a (small number of) symptom(s) to discriminate with a high degree of certainty between a small number of hypotheses. In example 3 e.g., 70% consensus is observed one symptom, the hormon level in serum. Apparently this so-called T4 test is viewed by the participants as a test with a very high sensitivity (inclusion of all diseased people who really have the disease: hyperthyreoidism) and a very high specifity (excluding all persons who do not have the disease). In the opinion of the participants this test can practically certainly discriminate between the hypothesis Hyperthyreoidism and the hypothesis Non-hyperthyreoidism. Obviously physicians need tests or symptoms that help them sorting out hypotheses with a high degree of confidence which provide them maximum certainty towards the

solution of the problem. But at the same time we can read from these scattergrams that some general consensus about these highly discriminating (sets of) tests or symptoms does not exist in practice. Rather, every physician uses his own set of tests and symptoms. Each set is used by the physician with a strong conviction of properness.
It would be of assistence to physicians to elicit and originate (sets of) proper discriminating tests and symptoms. The variation within the set of hypotheses resembles the variation within the set of symptoms, as can be observed from table 14. Recapitulating the variation, table 19 shows the wide differentiation.
Realizing the inconsistency of hypothesis contents we scrutinized the hypotheses as they were generated within the work-up of the case. As we already know the relevancy of the hypotheses to the pertaining and adjacent organ system is high (table 15), but they show a lot of variation over the physicians as appears from the next table.

Table 19.

VARIATION OF HYPOTHESES - NOMENCLATURE ACROSS THE CASES

PATIENT CASES	1	2	3	4	5	6	7	8	TOTAL
Total number of hypotheses	85	96	74	129	86	109	67	99	745
Number of different hypotheses	22	43	30	37	21	33	27	37	250
Average number of homonymous hypotheses per case	4.0	2.2	2.5	3.4	4.2	3.4	2.6	2.5	2.9
Number of records	31	31	33	33	33	32	30	30	253

Difference of hypotheses is rather the rule than the exception in medical problem solving.On the averrage not more than 10% of a case work-up contains homonymous hypotheses.

Table 19 examplifies the personal work-up of the problem solving procedure as we have pictured it as a characteristic of inductive reasoning.It also indicates the personal character of the inductive hypotheses as patterns originated from personal experienced cases and "practice-acquired" knowledge.

Rejection of redundant hypotheses is a criterion for the deductive strategy. We asserted a less rigorous behaviour in inductive strategy. Apart from the pattern-recognistion mode, more than one hypothesis was left at the closure of the diagnostic process. The table shows the results.

Table 20.

REJECTED AND REMAINING HYPOTHESES.
(Total number / strategy; mean number / case)

STRATEGY	TOTAL NUMBER HYPOTHESES	REJECTED HYPOTHESES	MEAN NUMBER PER CASE	STANDARD DEVIATION	REMAINING PER CASE
INDUCT.HEUR.	585	142	0.83	0.95	2.6
INDUCT.ALGOR.	107	21	0.75	0.89	3.0

On the average less than 1 hypothesis per case work-up is rejected. These results confirm the conception of hypotheses - end-point - redundancy in the inductive strategy.

These results sustain the theory.
Rejection in inductive strategy is not the result of logical reasoning but of the loss of conviction in a particular hypothesis.Its probability estimate falls below a - personally defined - threshold.This threshold can vary from physician to physician,from case to case,from day to day.
The function of the remaining hypotheses does not become clear in this study. Generally,we may conceive it as a matter of uncertainty. It could be hypothesized that the rejection rate was problem dependent. The next table shows the results:

Table 21

REJECTED HYPOTHESES PER PATIENT CASE

PATIENT CASE	HYPOTHESES GENERATED	HYPOTHESES REJECTED	PERCENTAGE OF REJECTION
1	85	11	13
2	96	23	24
3	74	7	9
4	129	34	26
5	86	20	23
6	109	38	34
7	67	12	18
8	99	18	18
Total	745	163	22

The table shows a variability of rejection across the cases between 9 and 34%. Apparently, rejection is as much case-dependent as the generation of hypotheses.

The figures and the percentages show a rather large variability of rejection across the cases. The percentages ranges from 9 to 34. The average rejection rate for hypotheses is approx. 20% for both the Inductive Heuristic and the Inductive Algorithmic strategy. Both deviations from this average exist for various cases: in the low region for cases 3 and 1 (9 resp. 13%); in the high region for cases 4 and 6 (27 resp. 34%). The variability indicates also problem-dependency for hypothesis-rejection.

We suggest that the lack of a proper discriminative factor or factors or the presence of it largely determines this rejection. The existence of a T4 test may contribute to the larger percentage of rejection of redundant hypotheses.

Another explanation may come from hypothesis-specification. Low-level hypotheses are more liable to falsification than the more universal statements. The levels of the rejected hypotheses are shown in table 21. Evidently, the latter explanation fails.

Table 22

LEVELS OF HYPOTHESES REJECTED : NUMBERS AND PERCENTAGES.

LEVEL	TOTAL NUMBER OF HYPOTHESES	NUMBER REJECTED HYPOTHESES	PERCENTAGE of REJECTION
I	30	14	47
II	100	23	23
III	110	23	21
IV	229	36	16
V	276	67	24

Rejection of hypotheses is distributed over all levels. For the levels II to V a significant difference for the rejection rate could not be found.

A hypothesis can be rejected during the problem-solving task or afterwards. Hindsight makes it easier to retrace on one's steps and reconsider the generated hypotheses. Another question is which hypothesis, in order of its generation, is rejected.

It could be hypothesized that the first hypotheses were more or less "wild guesses" and more liable to be overthrown by more appropriate hypotheses. With the progression down the cycles the rejection rate might decrease. The results are shown in table 23.

Table 23

HYPOTHESIS - REJECTION RATE PER CYCLE.

CYCLE	HYPOTHESES GENERATED	HYPOTHESES REJECTED	PERCENTAGE of REJECTION
1	253	92	36
2	199	30	15
3	147	11	7.5
4	89	1	1
5	41	-	-
6	14	-	-
7	2	-	-

The primary guesses are more liable to be overthrown than the following ones. Remember that the first cycle also contains 53 cases of Pattern Recognition. Calculation for the remaining cases in cycle 1 results in an increase for the rejection in cycle 1 to 46 %.

Thirty hypotheses were rejected after the process, representing 18,5% of the total sum of rejected hypotheses.

On a global scale the diagnostic process of the participating physicians shows a harmonious, rather uniform picture. Observed more in detail however, this

picture shows a number of very personal and,perhaps therefore,unexplained features.

Paragraph 3.

QUESTIONS,ANSWERS,TIME

Physicians work mostly under pressure of time. "The essence of the physician's art actually is to make decisions, a tremendous number every day, often the on basis of insufficient evidence, under pressure of time (...) and to make them with - at least outwardly - the appearance of a calm, dedicated and warmly human personality "(11). It is essential to the physician to choose carefully the right questions in order to get the desired answers within a reasonable time. The simulation standardized the 'patient' and, therefore, standardized the speed of answering, the reliability of the answers and their adequacy. On the other hand, there was awareness of being observed, the noting down on forms of several elements, like hypotheses, estimates of probabilities and uncertainty, prevalence- and incidence rates etc.. As was agreed in "Medical Problem Solving" (1) time constraint has to be part of the simulation. We estimated the average time for a routine patient-physician contract at approximately 10 minutes on the average for the family physician. Adding another 5 minutes for the noting down, explanation and different situation seemed reasonable. On this basis we presented 4 patient cases to be solved within an hour. Although general internists generally claim half an hour for each patient, we tried to persuade them to perform in high speed, which mostly meant three patients within the hour.

In a questionnaire each participating physician estimated the mean actual time for each patient contact and the mean overall time in office (excluding visits to patients, hospitals etc.) each day. The next table gives the results.

Table 24

PARTICIPANT'S ESTIMATIONS FOR THE AVERAGE TIME USED FOR PATIENT CONTACTS AND AVERAGE TOTAL OFFICE HOURS IN DAY-TO-DAY PRACTICE.

		MEAN TIME	ST.DEV.	MIN - MAX
PATIENT CONTACTS	FAM.PHYSCNS	9.27'	2.19	5' - 15'
	GEN.INTERNISTS	17.5'	5.98	10' - 30'
OFFICE HOURS	FAM.PHYSCNS	5.19 h	1.53	1 h - 9 h
	GEN.INTERNISTS	2.29 h	1.25	1 h - 4 h.

Our approximation was fairly consistent with these results. As we described in chapter 7 the participants considered the simulation a good substitute for real life situation. We also learned that the speed with which the general internists

ask their questions is higher than of family physicians. From the above-mentioned table we learn that internists use almost twice as much time as their colleagues in primary health care for a patient contact. This leads to the assumption that general internists need more than twice as many questions in solving the same problems as family physicians.
The next table shows the results:

Table 25

TOTAL AND MEAN NUMBER OF QUESTIONS PER CASE FOR FAMILY PHYSICIANS AND GENERAL INTERNISTS.

	FAMILY PHYSICIANS			GENERAL INTERNISTS		
PATIENT CASE	NUMBER of QUESTIONS	MEAN	NUMBER of CASES	NUMBER of QUESTIONS	MEAN	NUMBER of CASES
1	940	33.6	28	168	56	3
2	1151	39.7	29	167	83.5	2
3	717	23.1	31	196	98	2
4	1025	34.2	30	184	61.3	3
5	868	29.0	30	256	85.3	3
6	1301	45.0	29	235	78.3	3
7	659	24.4	27	277	92.3	3
8	1319	47.1	28	184	92	2

The table shows marked differences in numbers of questions over the various patient cases. The (mean) values for the family physicians differ to a larger extent than those of the general internists.

The values for the number of questions differ greatly within the group of family physicians.More than within the group of general internists, the figures indicate a case (problem)-dependent process.As aforementioned,the cases 2,4,6,and 8 were assumed to be the more difficult problems, especially for the family physicians.The general internists seem (their number of records makes clear-cut conclusions hazardous) to be rather consistent in the number of questions which sustain the conception of the inductive algorithmic strategy,which is largely attributed to this type of clinical specialists.

On the average the results are affirmative to the hypothesis, which leads to the conclusion that the relation between questions and time, albeit existing, must be differentiated for the various professions. The next tables give some overall results of the number of questions per group of physicians related to time. Time is divided into two categories: time until the diagnosis (the time course between the presentation of the complaint and the making of a diagnosis), and total time, including the preceding time course and the time necessary to establish, describe and explain a course of action to the patient. The first table is delineated in plain figures for the total numbers, the next shows the averages and their standard deviation.

Table 26

TOTAL AND MEAN NUMBER OF QUESTIONS PER PARTICIPANT
TOTAL AND MEAN TIME TILL DIAGNOSIS PER PARTICIPANT
TOTAL AND MEAN TIME FOR THE MEDICAL PROCESS

PHYSICIANS	TOT.NR.QUESTIONS	MEAN	TIME TILL DIAGN.	MEAN	TOTAL TIME	MEAN
FAM.PHYSCNS	7.796	130	2.437'	40.6'	2.602'	43.4'
GEN.INTERN.	1.667	208	383'	47.9'	398'	49.8'
TOTAL	9.463	139	2.820'	41.5'	3.000'	44.1'

The mean values refer to the averages of the total work-up for the 4 (with an exceptional case of 3) presented patient problems. General internists generally solved three cases. Family physicians asked fewer questions in less time than the other physicians.

Although comparisons are hard to make (family physicians generally solved 4, general internists 3 problems) it is be clear that the average number of questions asked by the family physicians during the total session is markedly lower (60%). Family physicians ask fewer questions in less time.This conception is consistent with the heuristic behaviour in interviewing the patient,as we have described in chapter VI.

Table 27

AVERAGE NUMBER OF QUESTIONS - AVERAGE TIMES PER PATIENT CASE.
FOR FAMILY PHYSICIANS AND GENERAL INTERNISTS.

	QUESTIONS	TIME t.DIAGNOSIS	TOTAL TIME
FAMILY PHYSICIANS	34.1 + 12.1	10.7' + 3.87'	11.25 + 3.90
GENERAL INTERNISTS	80.7 + 20.6	18.55' + 5.55'	19.25' + 5.8
Total	37.4 + 20.66	11.15' + 5.53'	11.86' + 5.8

This table describes per patient case what could be deduced from the former table.The differences between the two groups of physicians in theses respeects are obvious.
(Time is counted in minutes)

Comparing the results with the physician's own estimates (table 24) we learn from table 26 that for the simulation scene the physicians remained within the same limits for consulting hours as in routine practice. The same observation holds true for patient contacts.

The differences between the group of family physicians and the group of general internists can also be traced to differences between the strategies.

Table 28

NUMBER OF QUESTIONS AND TOTAL TIME EMPLOYED PER STRATEGY

STRATEGY	Nr of QUESTIONS	MEAN	ST.DEV.	TIME (min)	MEAN	ST.DEV.
PATT.RECOGN.	592	21.92	8.54	182	6.74	2.61
HES.PATT.REC.	794	30.54	13.98	269	10.35	4.68
INDUCT.HEUR.	5765	33.52	12.67	1943	11.30	4.14
INDUCT.ALGOR.	2312	82.57	14.77	606	21.64	5.29

For these items the Inductive Algorithmic strategy largely deviates from the other methods.

As might be expected pattern recognition scores lowest in questions and in time, while the Inductive Algorithmic strategy strikes by its large number of questions (more than 80) asked in a twenty minutes' time-span.

It has been hypothesized that similar ranges and mean lengths of patient interview suggest similar distributions of case difficulty and complexity (21). This suggestion consists of two conceptions: a relationship between case-complexity and length of work-up, and the relationship between questions and time. Standardizing the problem case on the one hand to the least number of questions and on the other to the least of time used in the problem-solving, we could construct correction factors for each of the eight patient cases.

Table 29

DISTRIBUTION OF THE NUMBER OF QUESTIONS AND TIME TILL DIAGNOSIS PER CASE.

PATIENT CASE	NUMBER of PHYSICIANS	NUMBER OF QUESTIONS MEAN	NUMBER OF QUESTIONS ST.DEV.	TIME TILL DIAGNOSIS MEAN	TIME TILL DIAGNOSIS ST.DEV.
1	31	34.77	13.46	10.13'	4.37'
2	31	39.39	18.19	13.16'	4.60'
3	33	25.49	20.77	8.67'	4.47'
4	33	30.70	12.93	12.18'	4.62'
5	33	30.24	20.79	10.09'	4.61'
6	32	42.88	16.85	13.97'	5.28'
7	30	29.27	23.31	7.70'	4.57'
8	30	36.67	20.81	13.30'	7.97'

The number of questions across the various cases fall within the range of approx. 30 - 40 questions per case work-up. The large standard deviation indicates a great variability for this item. The time till diagnosis predicts about the "more difficult" cases (2,4,6,8). The - rather - stable standard deviation indicates that this variable is a more consistent one across the cases.

The following correction factors can be calculated:

Table 30

DIFFICULTY OF THE CASES AS A QUESTIONS - TIME RELATIONSHIP.
(Correction factors for the questions and time with regard to the difficulty factor of the presented problem

PATIENT CASE	CORRECTION FACTOR QUESTION	CORRECTION FACTOR TIME
1	1.364	1.315
2	1.545	1.709
3	1.000	1.126
4	1.205	1.582
5	1.187	1.311
6	1.682	1.810
7	1.148	1.000
8	1.439	1.727

The case with the least value was considered the standard

Although the 'standard cases' were different for each of the factors, they approximate the suggestions of the hypothesis mentioned above. Especially the time factor underlines what we already know from the preceding descriptions, namely that the patient cases 2,4,6 and 8 were the complex and difficult ones.

The relation between questions and time can be calculated crosswise.When the mean number of questions is corrected for the time factor and the mean time until diagnosis (the time used for patient interviewing) is corrected for the amount of questions,we may expect an ironing out of the differences in case of complete relationship.The result of this operation is shown in table 31.

Table 31

INTERRELATIONSHIP QUESTIONS - TIME PER PATIENT CASE

Patient	Questions corrected	Time corrected
1	26,44	7,43
2	23,07	8,52
3	22,63	8,67
4	19,40	10,11
5	23,07	8,50
6	23,69	8,30
7	29,27	6,71
8	21,23	9,24

The variables seem to relate to each other, although the differences make clear, that the relationship is much more complex than we might have expected.

The two variables seem to relate to each other, although some unexplained deviations remain (especially cases 4 and 7). The number on which these calculations is based does not allow for exact conclusions on this subject; it suggests for a serious relationship.

There may always be an important point which might distort this relationship between questions and time. De Dombal exclaimed: " The doctor has to ask the right questions, he also has to ask the questions right" (13). Whether 'our' doctors asked their questions right was beyond the scope of the investigation; they all faced an intelligent patient responding to all questions, even the unclear ones, or in case of doubt, rephrasing the doctor's question to a clear one. An ideal situation, doctors will say, and certainly not always met in daily practice. But the first part of De Dombal's exclamation pertains to the effectiveness of the physician's questioning. This effectiveness includes two characteristics:

- the relation of the questions to the pertaining hypothesis (as we described in paragraph 2),
- how many times a positive symptom is scored by a question: a positive symptom being a characteristic of the patient deviating from normal (as perceived by patient and/or physician) functioning or form . When the asked symptom is present in the patient's 'data-base' it is coded with +, when not it is coded with -.

With the help of this coding, we could find this type of effectiveness of physician's questioning. The next table shows the overall results:

Table 32

PROPORTIONAL NUMBER OF POSITIVE SYMPTOMS PER TOTAL NUMBER OF QUESTIONS.

	FAMILY PHYSCNS	GENERAL INTERNISTS
Nr of Pos.Symptoms	4242	596
Nr of Questions	7796	1667
Percentage	54.4%	35.8%

The acquisition of positive symptoms appears more effective for family physicians than for general internists

The smaller percentage for general internists follows from the type of strategy he mainly applies: the inductive-algorithmic. The effectiveness for this type of strategy must come from the more extensive covering of the organ systems and the greater speed with which the users ask their questions.

And these people are rewarded! They collect a high count of positive questions at the cost of a large number of questions, as shown in table 33.

Table 33

PROPORTIONAL NUMBER OF POSITIVE QUESTIONS PER STRATEGY

Strategy	Pos.Sympt. p.physcn	Nr of physcns	Proportional to total Nr of questions	Total Nr. of questions
Patt.Recogn.	13.2	27	60%	592
Hes.Patt.Recogn.	17.3	26	56.5%	794
Induct.Heur.	18.7	172	55.9%	5765
Induct.Algorithm.	28.9	28	35.0	2312

The proportional numbers for the acquisition of positive symptoms are favourable for the pattern recognition and the heuristic methods.

The effectiveness as the ratio of positive symptoms divided by the total number of questions, i.e. the RATIO effectiveness , is highest in the pattern-recognition strategy and lowest in the inductive-algorithmic. But the amount of elicited positive symptoms, the CONTENT effectiveness , is highest in the inductive-algorithmic and reverse in pattern-recognition. It seems that RATIO-effectiveness and CONTENT-effectiveness operate in opposite directions. It looks as if a physician has to make a choice: a high speed of questioning in order to collect randomly a sufficient number of positive symptoms or carefully choosing the questions that will probably elicit the highest amount of positive symptoms. In my opinion physicians are hardly ever aware of this choice. In fact they act in conformity of their natural behaviour and at the spur of the moment.

However, we have to realize that another feature of the process can be involved. It might be possible that the first steps in the process are more difficult than later ones. In that case, the physician with the least number of questions runs the chance to elicit less positive symptoms than the persisting physician, who asks questions at great length. We calculated the RATIO-effectiveness for each block of 10 questions in the order of their questioning.

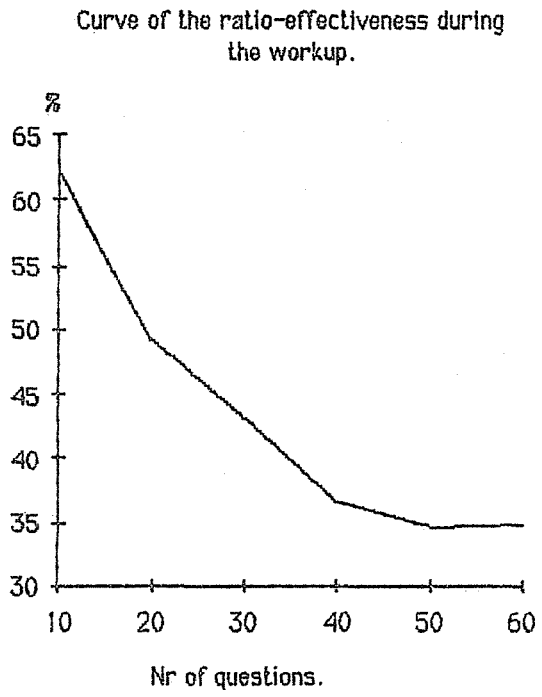
Table 34.

THE COURSE OF THE RATIO-EFFECTIVENESS DURING THE PROBLEM SOLVING

BLOCK OF 10 QUESTIONS	TOTAL NUMBER OF QUESTIONS	NUMBER OF POSITIVE SYMPTOMS	RATIO-EFFECTIVENESS %
1 - 10	4916	3060	62.2
11 - 20	2265	1115	49.2
21 - 30	1098	475	43.2
31 - 40	444	162	36.5
41 - 50	228	79	34.6
51 - 60*	135	47	34.8

* beyond the 60 questions calculations become unreliable.
The decreasing numbers of questions is overtaken by the rate of decrease of the positive symptoms. In fact, ratio-effectiveness gradually diminishes during the work-up.

FIGURE 2



Evidently, the first questions are the 'easiest' ones to elicit positive symptoms. Gradually the effectiveness levels down to a range between 30 and 35%. This figure is essentially important to everyone planning to produce a medical data base. This base must be up to its task for a number of approximately three times the attributes necessary to make a diagnosis, regardless whether it is right or wrong.

The CONTENT-effectiveness is remarkably harmonious across the patient cases.

Table 35

CONTENT-EFFECTIVENESS PER PHYSICIAN PER PATIENT CASE.

PATIENT CASE	NUMBER OF POSITIVE SYMPTOMS	CONTENT-EFFECTIVENESS p.PHYSICIAN (mean values)
1	547	17.7
2	598	19.3
3	565	17.1
4	804	24.4
5	619	18.7
6	635	19.8
7	517	17.2
8	553	18.4

The Content-Effectiveness shows rather harmonious values across the patient cases.

In contrast to most variables the Content Effectiveness seems rather unaffected by the problem variation.

In general,these results suggest that strategy and problem variation do not influence the acquisition of positive symptoms.Whether the amount of positive symptoms affects the strategy and the diagnostic outcome is a question for future research.

Each patient scenario has a more or less standardized number of positive symptoms and symptom-aspects.We did not differentiate for symptom aspects.Clusters of symptom-aspects were too accidentally achieved and, therefore,cannot contribute to more general features of the process.

When a symptom was asked more than once,partly because of the different aspects,partly because the physician obviously had forgotten the answer,we noted as positive the times the physician asked this special item.

It is the physician's task to elicit effectively the positive symptoms and symptom-aspects from the 'data-base',which can be a real-life patient as well as a simulated one.The effectiveness of this eliciting leads to the retrieval rate of the questioning process.

Table 36

RETRIEVAL RATE FOR POSITIVE SYMPTOMS FROM A PATIENT DATA BASE.

PATIENT CASE	PATIENT DATA BASE SYMPTOMS	SYMPTOM-ASPECTS	RETRIEVAL RATE for SYMPTOMS	SYMPTOM-ASPECTS
1	47	278	37.7%	6.5%
2	49	281	39.4%	6.9%
3	54	306	31.7%	5.6%
4	53	316	46.0%	7.7%
5	46	274	40.7%	6.8%
6	48	291	41.3%	6.8%
7	51	293	33.7%	5.9%
8	49	299	37.6%	6.2%

Column II displays the standardized numbers of symptoms as they are incorporated into each patient scenario. Column III displays the symptom-aspects. The retrieval rate indicates that a relatively large portion of the symptoms in the data base is visited, but in a superficial way, as we can see from the last column.

The Retrieval rate relates the acquisition of positive symptoms to the standardized patient scenario.It establishes the thoroughness of searching through this symptom/symptom-aspects data base.Like the content-effectiveness it shows a rather harmonious picture across the cases with the exception of the famous case 4.

The elicitation of symptom-aspects is,according to these results,not a major characteristic of the search procedure.

Evidently, a large number of symptoms is visited during the questioning and/or physical examination or laboratory data, but only a very few symptoms are more deeply interrogated. It gives the impression of skimming-off the territory.

Recapitulating the variables for the positive symptom elicitation we can distinguish three types:

1) RATIO-EFFECTIVENESS ,which relates the acquired number of positive symptoms per case to the total number of questions asked within this case work-up.

2) CONTENT-EFFECTIVENESS ,which relates the number of positive symptoms to the physician who collected them.The figures are given as mean values for a group of physicians solving the same problem.This variable allows comparisons between physicians in this respect.

3) RETRIEVAL RATE ,which relates the collected number of positive symptoms per case to the standardized patient data base.It may measure the effectiveness of the search procedure.

Each "paper patient model" is subdivided into a number of classes.(for the description see chapter VII). With the help of these classes comparisons can be made between the styles of the family physician and the general internist.

Several hypotheses have been made on this subject.
It was hypothesized that family physicians asked more questions about social background and mental status than general internists (14, 15). Although family physicians may differ in their search methods, questions about social and emotional troubles constitute a considerate part of the history-taking (16). Physical examination constitutes a rather varying part of the diagnostic procedure(16,17). It was hypothesized that general internists spent more time on this feature of the diagnostic process than family physicians (14, 15). The same trend was present with regard to laboratory and X-ray data. Smith & McWhinney (14) could verify these hypotheses in two of their three simulated patients, while Gerritsma & Smal (15) did not found significant differences.
Our results for the various classes were:

Table 37

VARIATION OF SEARCH THROUGH THE DIFFERENT CLASSES OF THE "PAPER PATIENT"

CLASS	FAMILY PHYSCNS %	GENERAL INTERNISTS %
Social background	7.8	3.5
Medical background	5.5	4.9
Current diseases	4.3	2.1
Medical history	42.9	37.9
Physical examination	31.0	43.7
Laboratory etc.	8.4	7.9

Family physicians asked more questions about the social background of the patient and performed less physical examinations than the general internists.

These results partly corroborate the Smith & McWhinney hypothesis. Obviously, family physicians asked more questions about the social and emotional background of the patient, but history-taking was almost as elaborate for family physicians as for general internists. Physical examination (which was performed by asking specific questions) played a slightly larger part in the diagnostic process for general internists than for family physicians. We found no difference for the laboratory and X-ray class between the two categories of physicians. However, these results should be cautiously interpreted, because of the relatively small number of participating internists. To interpret differences in style between these physicians on the basis of class variations seems, however a dubious foundation.

Paragraph 4

PROBABILITIES AND (UN)CERTAINTIES

"Decision-making is reducing uncertainty in a problem situation" (18). Initially, before a patient enters the health care system, uncertainty as to the true state of the individual is maximal to the people in the system, but by means of a certain set of maneuvers on the part of the health care system and the patient, information is obtained which reduces the uncertainty, to a point, it is hoped, where this information is prescriptive of a course of action (3). In terms of information science, each bit is able to attenuate the problem-solver's mental state of uncertainty as to the possible solution of the problem. When 'cues' are defined as "items of meaningful information", they may point to certain or probable diagnoses. A cue (or a number of cues) may enable the physician to say with certainty what is wrong with the patient (19). But this conception implicates that the problem solver recognizes this item of information as a 'cue' or a 'bit'. It means that it must match some prior information, or some picture in mind. Crombi & Pinsent (20) focus our attention on the observation that the acquisition of clinical information is more often than not dependent on prior information of a highly selected kind available only to a personal doctor. But being aware of the enormous amount of information beyond the grasp of the physician and/or the patient uncertainty can be maximal.Given the available data, the physician estimates the probability that the patient is suffering from disease A, disease B, disease C, and so on (19). But not every datum is a cue. When cues reflect the pictures in mind, probability estimates might mirror the personal opinions of physicians in diagnosing diseases.

However, we have to be sure which disease pattern one has in mind. Each disease might evoke its own personal pattern from the physician's memory. Family physicians, e.g., observe diseases in an early stage before the full clinical picture has developed. Information on which to base a precise diagnosis - the kind of information discussed in textbooks - is often not available to the family physician when he first sees the patient.

According to McWhinney(19) knowledge of the patient's background may be the most distinguishing feature of family medicine,and may positively contribute to the family physician's decision-making.But we must clearly be aware of the disadvantages of this foreknowledge.

Not only can it lead family physicians into prejudiced and erroneous directions,but - because of the personal and implicit nature of this foreknowledge - also to observation errors and misjudgment of presented complaints and signs.In their turn these observations and judgments lead physicians to implicit and personal apprehension of the frequency of occurrence of particular ailments and complaints within their pertinent practice population.

The question is whether the morbidity and the symptoms presented reflect the prevalence of disease or symptoms in the community? (21,22). Decision-making approaches (e.g. Bayesian, Discriminant analysis) have to rely on this assumption of reflecting prevalences.

As we discussed before (chapter V), there are at least two main obstacles: reliable morbidity figures do not exist, and, relying on subjective estimates means introducing biases in the form of a number of heuristics people use in judging frequencies and their distributions (Tversky & Kahnemann, 23). We may wonder whether the personal estimates of prevalences of diseases and symptoms are consistent among physicians. For that reason we asked the participating physicians to estimate the figures of prevalence, incidence and

disease-consultation for the particular disease they diagnose during the session.

Table 38.

PREVALENCE, INCIDENCE, DISEASE-CONSULTATION RATE ESTIMATES FOR 8 DISEASE - ENTITIES

DISEASE	PREVALENCE*		INCIDENCE		DISEASE-CONSULTATION	
	MEAN	ST.DEV.	MEAN	ST.DEV.	MEAN	ST.DEV
Myoc.infarct.	2.21	4.69	48.32	23.95	114.32	128.75
Ischialgia	133.80	203.77	460.30	981.96	707.60	1070.04
Gall-stones	147.61	247.10	21.72	17.23	83.72	90.70
Hyperthyreoid.	50.63	52.38	12.33	17.50	90.67	166.99
Bronchitis	337.88	645.84	1287.38	1372.84	2918.63	3485.97
E.U.pregnancy	3.05	10.89	8.38	6.26	9.67	9.45
Iron-def.an.	206.35	510.83	64.35	165.41	194.90	254.77

*the figures are standardized for a population of 10.000.

The huge figures for the standard deviation challenge the assumption that subjective probability estimates could reliably be used for statistical approaches in medical decision making.

*(note):
Prevalence : number of patients with a certain disease in a circumscribed population.
Incidence : number of new patients with a certain disease seen in a circumscribed time interval (customarily one year).
Disease consultation rate : number of patients with a certain disease visiting the physician during a circumscribed time interval.)

The huge figures for standard deviation challenge the assumption that subjective probability estimates can reliably be used for statistical approaches in medical decision making. (see Gustafson et al, 24)

They underscore instead, Croft's (25) statements about the real obstacles to practical computer-aided medical diagnosis:

1.) lack of standard medical definitions; the large variability among the physician's estimates can be explained by their different concepts about the diseases; and

2.) lack of large, reliable medical data bases.

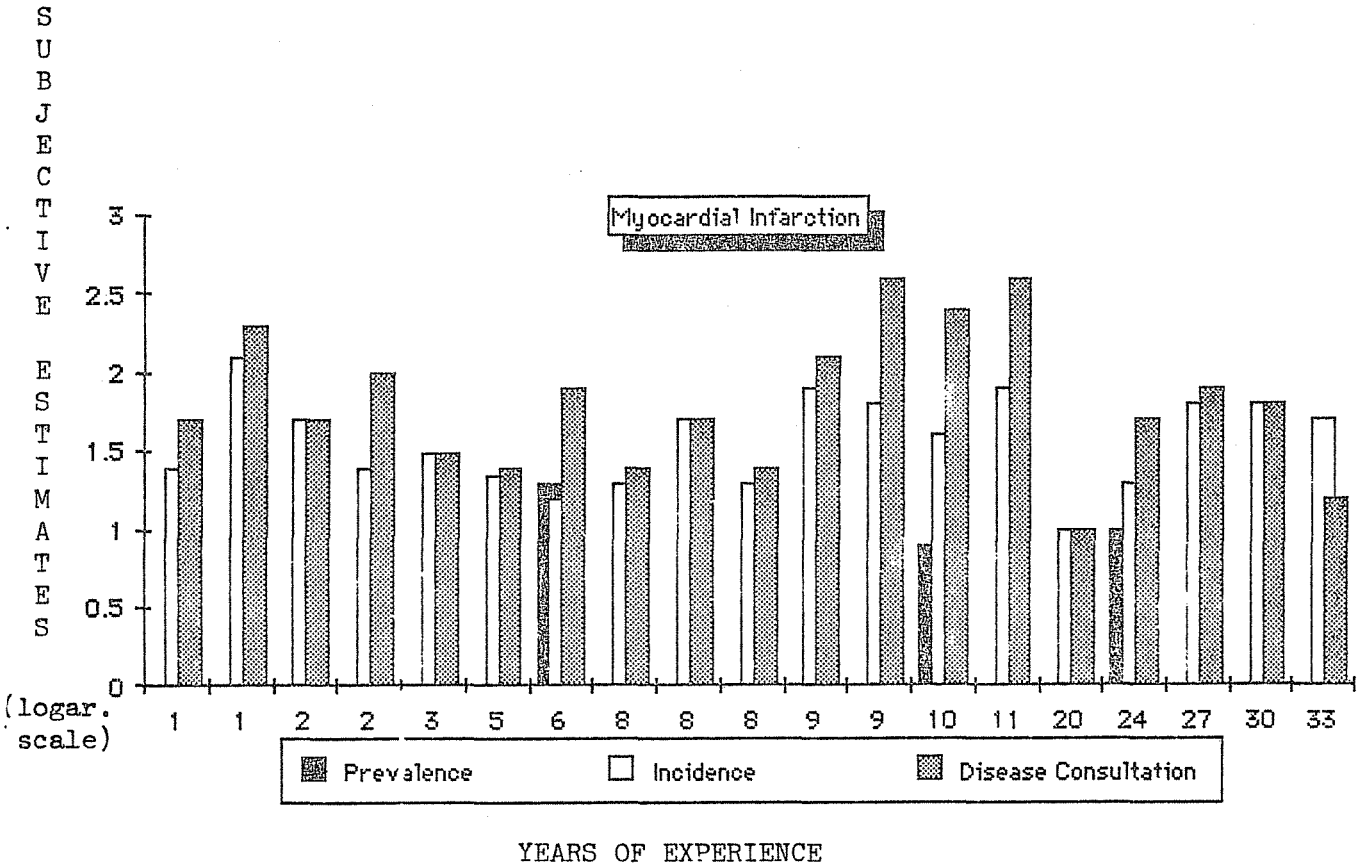
The participating physicians obviously related their estimates to recent experiences. Recent calls for extra-uterine pregnancy made figures jump to a 20 - 30 fold of the average. Because the investigation took place during the months of February till July, seasonal influences were evident. Figures for bronchitis were highest in the first two months and lowest in the last month. On a large scale, the variations of the estimates could be contributed to highly personal, situational and/or seasonal influences,and only partly related to the disease in question.

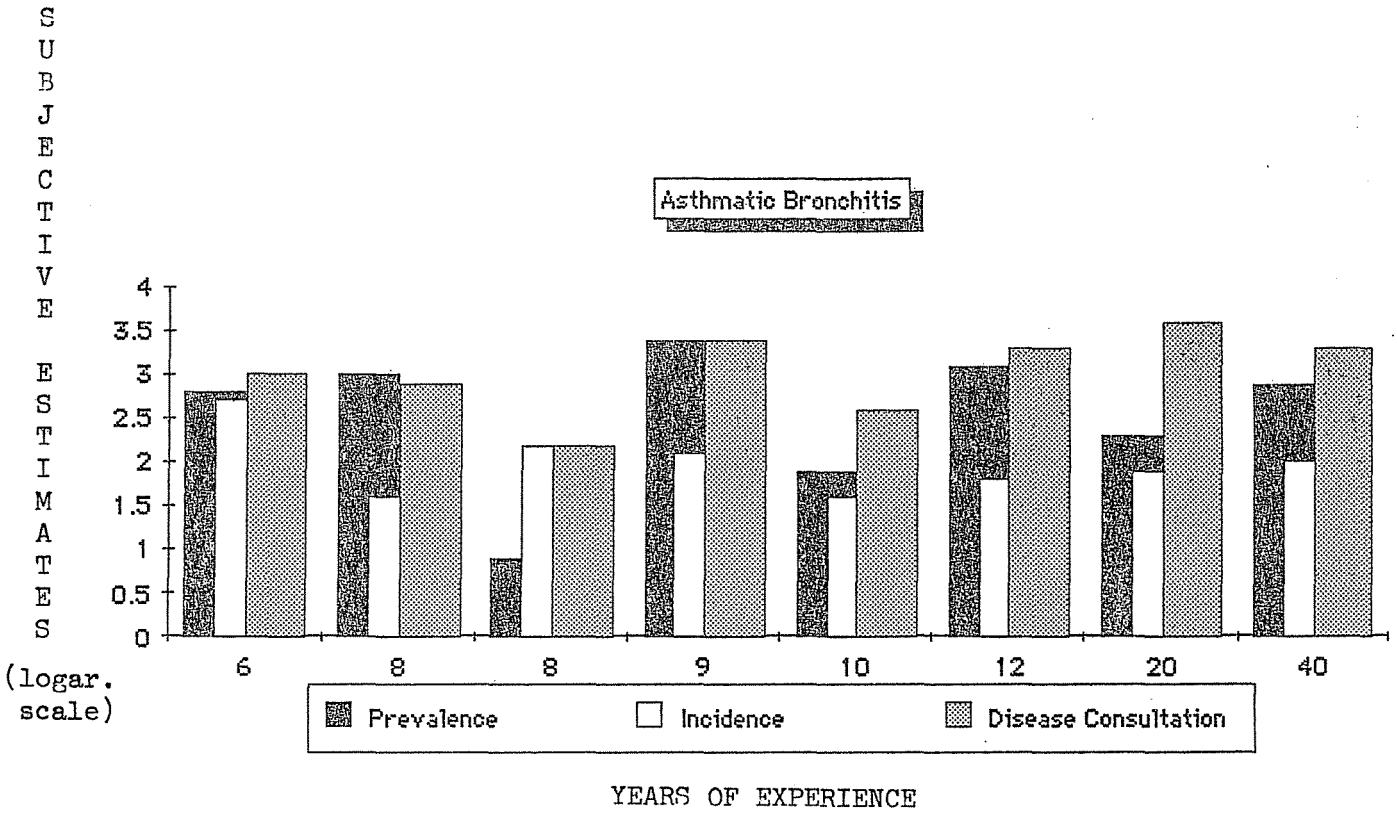
From these figures it may be clear that several assumptions as they are made in medical decision-making like symptom independence, normal probability distributions, tightly clustered values for all patterns in one class, reliable morbidity figures etc. are not always met in a number of various diseases. And, as Croft remarks, "the more grossly the assumptions are violated, the less accurate the diagnoses are expected to be."

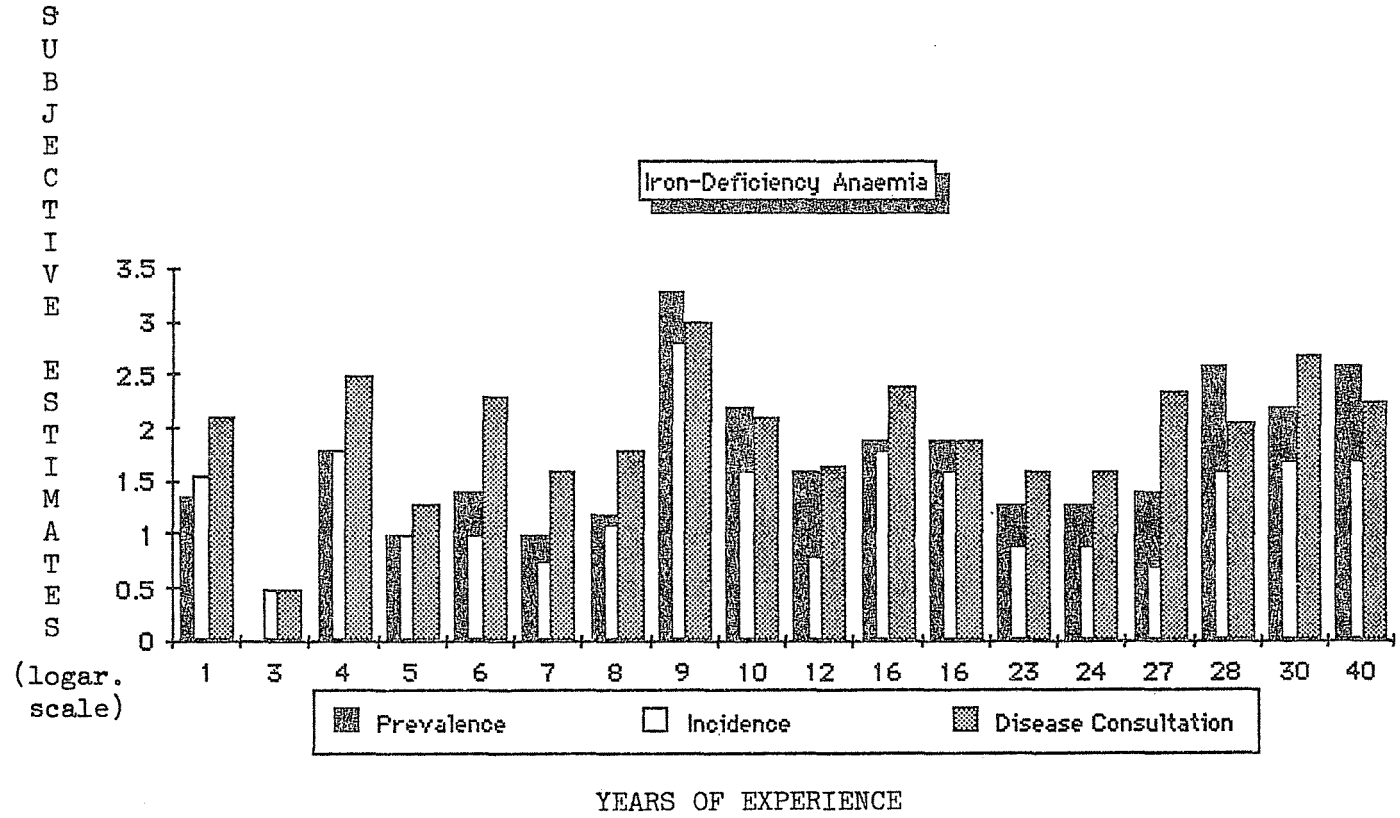
An assumption regularly sustained in literature is the relation of these estimates, especially their validity and reliability of them, to the experience of the physician. It is expected that the more years the physician has spent in real practice the more reliable and valid his estimations. Although we shall discuss the relation of experience to some elements of the medical process in paragraph 7 of this chapter, we shall in this case make an exception for the estimates of prevalence, incidence and disease-consultation. We shall present the results for three diseases, myocardial infarction, bronchitis and iron-defiency anemia in the following figures.

On the horizontal axis the years of experience of the physicians is presented. Because of the wide variation of the estimates the vertical axis had to be designed in a logarithmic scale. The black columns represent the prevalence rates, the white ones the incidence and the pointed ones the disease-consultation rates. For each physician a series of three columns is drawn. The very tiny columns approximate the figure zero.

Figure 3:




Asthmatic Bronchitis
SUBJECTIVE ESTIMATES
(logar. scale)
4
3.5
3
2.5
2
1.5
1
0.5
0
6
8
8
9
10
12
20
40
Prevalence
Incidence
Disease Consultation
YEARS OF EXPERIENCE


Iron-Deficiency Anaemia
SUBJECTIVE ESTIMATES
(logar. scale)
3.5
3
2.5
2
1.5
1
0.5
0
1
3
4
5
6
7
8
9
10
12
16
16
23
24
27
28
30
40
Prevalence
Incidence
Disease Consultation
YEARS OF EXPERIENCE

The notion of expertise rests on the presumption that medical competence is an innate characteristic of experienced physicians that is generalized across clinical situations. If we assume that more than 10 years of regular clinical practice represents some kind of expertise, we may expect some conformity among the experienced participants regarding the various estimates.We remember that most physicians regarded the simulations as highly valid for real-life situations!

Instead, we found tremendous differences. Moreover, we found approximately the same variations among the experienced and the non (or less) - experienced physicians, and between these groups.Any influence of experience or expertise upon the estimates of prevalence, incidence and disease consultation rate could not be detected.

In decision theory it is hypothesized that generated hypotheses are weighed or ordered according to their prior probabilities reflecting the prevalence (or incidence) of the disease. Each bit of information can change these probabilities. When the probability of a certain hypothesis surpasses a certain, predefined, threshold the decision-maker can consider the presence of a disease to be confirmed (26,27). We asked the participants to identify the degree of weight, or probability, to each hypothesis immediately after its generation. There were two possibilities for (re)adjustment of the subjective probability of a hypothesis.The participants could adjust their probability estimates in the light of an important cue, and they were asked to (re)consider their estimation of the previously generated hypotheses in the work-up,any time a new hypothesis came to their mind.To that purpose he could give a mark on an open horizontal bar, measuring 10 centimetres; each centimetre thus indicated 10% (from left - 0 % - to right, 100 %). For each hypothesis three bars were available:one for a prior probability estimation,one for an important cue adjustment,and one for a posterior probability estimation.In case of a need for more scoring possibilities,the next bars were available.

Neither of the physicians used the possibility of 'cue-adjustment',or the extension of another three bars.They all made their estimations on basis of the generation of another hypothesis,thus confirming the concept of the inductive nature of these hypotheses.

When the physician wanted to re-establish a hypothesis,which he (temporarily) quit during the work-up,he was able to note it again as a newly arrived hypothesis with its own prior probability,which could be quite different from the one the physician generated at an earlier time in the course.

Our assumption was that in case of a - rather - consentient work-up the probability estimates could tell something about the physicians' subjective estimations of the pertaining patient population.

If the physicians' estimates reflect the morbidity states in their pertaining patient populations, careful office registration cannot only provide each doctor with his disease-symptom data bases but also with the probability estimates necessary to process patients' data in a decision- theoretical way. As the reader already knows (paragraph 3) the variety of hypotheses within each case's work-up made testing of this assumption impossible.

The ultimate point of the diagnostic process in inductive strategy has been defined as a maximalization of the disease probability and a minimization of uncertainty.

When we first take a look at the probability estimates, we are able to determine the subjective probabilities across the hypotheses of the work-ups: from the prior probabilities of the first hypothesis till the posterior probability of the last (diagnosis) hypothesis.

For the four types of inductive reasoning some assumptions about their respective behaviour can be made. Physicians using the pattern-recognisition strategy must be quite determined; a relatively high increase of probability estimate and a rather high maximum. The hesitating pattern-recognition method is symbolized by a small increase and a rather moderate maximum probability estimate. The "inductive-heuristic physician" starts easily, making a slow but

certain increase and reaches a slightly lower maximum than the quite determined first physician. The inductive-algorithmic method, in spite of the elaborate work-up, was presumed to stay on the safe side of the maximum: the more information the more possibilities one has in mind, the less determined the judgment. The next table shows the results:

Table 39

FROM THE FIRST (HYPOTHESIS) PROBABILITY ESTIMATE TO THE LAST (HYPOTHESIS) PROBABILITY.

INCREASE OF PROBABILITY ESTIMATES PER STRATEGY.

STRATEGY	FIRST PROBABILITY		LAST PROBABILITY		INCREASE
	MEAN VALUE	ST.DEV	MEAN VALUE	ST.DEV	
PATT.RECOGN.	47.48	20.25	81.41	22.97	33.63
HES.PATT.RECOGN.	56.12	19.11	73.38	26.00	17.26
INDUCT.HEURIST.	41.28	21.67	73.74	23.94	32.46
INDUCT.ALGORITHM.	47.86	15.92	69.96	28.11	22.10
MEAN TOTAL	44.22	21.16	73.91	24.47	29.69

(figures are given in percentages)

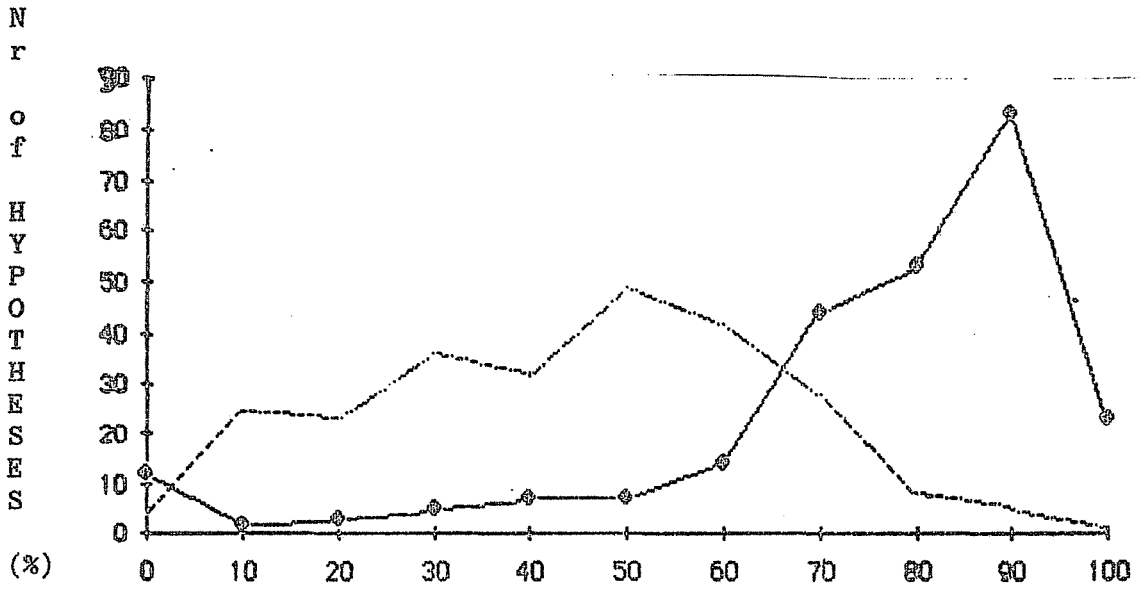
The estimates represent the weights which is given to a particular hypothesis. This weight is an expression of the physician's confidence being on the right track towards a solution. Across the various hypotheses which are generated in the course of the solution these weights follow the line of the physician's search and reasoning. These estimates give us indications about the physician's problem solving behaviour, not about precise calculations.

The great variability among the participating physicians is represented by the large figures for the standard deviation. Our assumptions about the behaviour of the various types of inductive strategy do not surpass the requirements of scientific proof. They are a mere indication, perhaps speculation, worthwhile to take it a more specific look in future.

The variability can be sketched by the frequency distribution of the prior and posterior probabilities. In figure 6 the figures on the horizontal axis represent the distribution of the probabilities, and on the vertical axis the number of hypotheses (in percentages) to which the particular probability pertains. The grey dotted line stands for the prior probabilities, the black line for the posterior probabilities.

Figure 6

COURSE OF THE PROBABILITIES



Subjective Estimations

Distribution of the estimated probabilities for prior and posterior values. The dotted line stands for the prior probabilities and the black one for the posterior probabilities.

The broad and shallow curve of the prior probabilities reflects the wide variability of the primary hypotheses. As the last hypothesis does not always represent the diagnosis, e relocated for reasons of comparability for this occasion all diagnostic hypotheses to the last place in the sequence of hypotheses.Therefore,all posterior probabilities can be considered to be the estimates for the diagnosis.From figure 6 we can find the suggestion that the end of the diagnostic process is partly characterized by a maximization of the - subjectively estimated - probability to the final hypothesis.

The assumption that people when uncertain about the direction or the outcome of a gamble generally choose a 'mid-position' (see also chapter III) is demonstrated in the grey curve. As the inductive-algorithmic strategy, the type with the largest number of specialists, gives the idea to be a more cautious strategy, we were interested whether the 'starting-position' and 'end-position' differed between family physicians and general internists. It was presumed that both estimates (prior and posterior probabilities) scored lower for the group of general internists than for the group of family physicians. This presumption could not be sustained by the figures as presented in table 40. The tendency to take a midposition in gambling situations with an uncertain outcome,as was described in chapter III,could be demonstrated.

Table 40

MEAN PROBABILITY ESTIMATES FOR THE TWO GROUPS OF PHYSICIANS

	FIRST PROBABILITY	LAST PROBABILITY
FAMILY PHYSICIANS	44.2	74.4
GENERAL INTERNISTS	44.3	68.6

Both groups of physicians start with equal prior probability estimates. In the - diagnostic - posterior probability estimates a slight difference can be observed. Family physicians give their predictions a higher value.

We were also interested in the difference of behaviour between physicians primarily estimating a low-level prior probability, (less than 0.5), contrasted with those physicians who estimated at a higher level, (more than 0.5).In 169 out of 253 records the prior probabilities were estimated at a less than 0.5 level. 142 records of this group (84%) scored at a more than 0.5 posterior probability level.The group of records starting at a more than 0.5 level equally (88%) ended up at a more than 0.5 posterior probability level.More or less confidence in the first hypothesis does not lead to significant differences of the probability estimations of the outcomes.Prediction on basis of subjectively estimated prior probability values is quite hazardous.

Table 41

COURSES OF PROBABILITY ESTIMATES FOR 253 PHYSICIAN CASES

	PRIOR PROBAB.<0.5		PRIOR PROBAB.>0.5		TOTAL
	number	col %	number	col%	
POST.PROBABILITY <0.5	27	16	10	12	37
row %	73		27		100
POST.PROBABILITY >0.5	142	84	74	88	216
row %	66		34		100
TOTAL	169	100	84	100	253

Comparisons between two groups of physicians, one starting at a less than 0.5 probability level, one starting at a more than 0.5 level, have been made. Both groups end up in equal extent in the high level probabilities. 85% scored in the end at the more than 0.5 posterior probability level.

These results underscore the graphics in figure 6. No matter at which level the physician started his estimation, he generally ends up in the high regions. The question was raised whether this result was an overall effect or had to be differentiated for the various problems. In the next table (42) the mean values for the prior probability of the first hypothesis (Pr.Prob.I) and for the posterior probability of the last hypothesis (Po.Prob.L) are given. The difference between these two values can be viewed as the increase of confidence over the succesive hypotheses, and expressed as a percentage of prior probability I.It may be denoted as the rate of increase.

Table 42

FROM PRIOR TO POSTERIOR PROBABILITY AS A MEASURE FOR INCREASE IN CONFIDENCE
(per case)

PATIENT CASE	PRIOR PROBAB.1 MEAN VALUES	POST.PROBAB.last MEAN VALUES	INCREASE	RATE of INCREASE %
1	42.6	72.8	30.2	71
2	42.9	73.5	30.6	71
3	54.2	60.8	6.6	12
4	36.7	76.9	40.2	110
5	45.2	74.8	29.6	65
6	34.4	81.2	46.8	136
7	54.0	69.6	15.6	29
8	44.3	82.1	37.8	85

The increase of probability estimates during the work-up gives some insight into the physician's conception about the case. It indicates his increase of confidence to find a solution for the problem. The "more difficult" cases (2,4,6,8) generally start at a low level, and reach the highest increases.

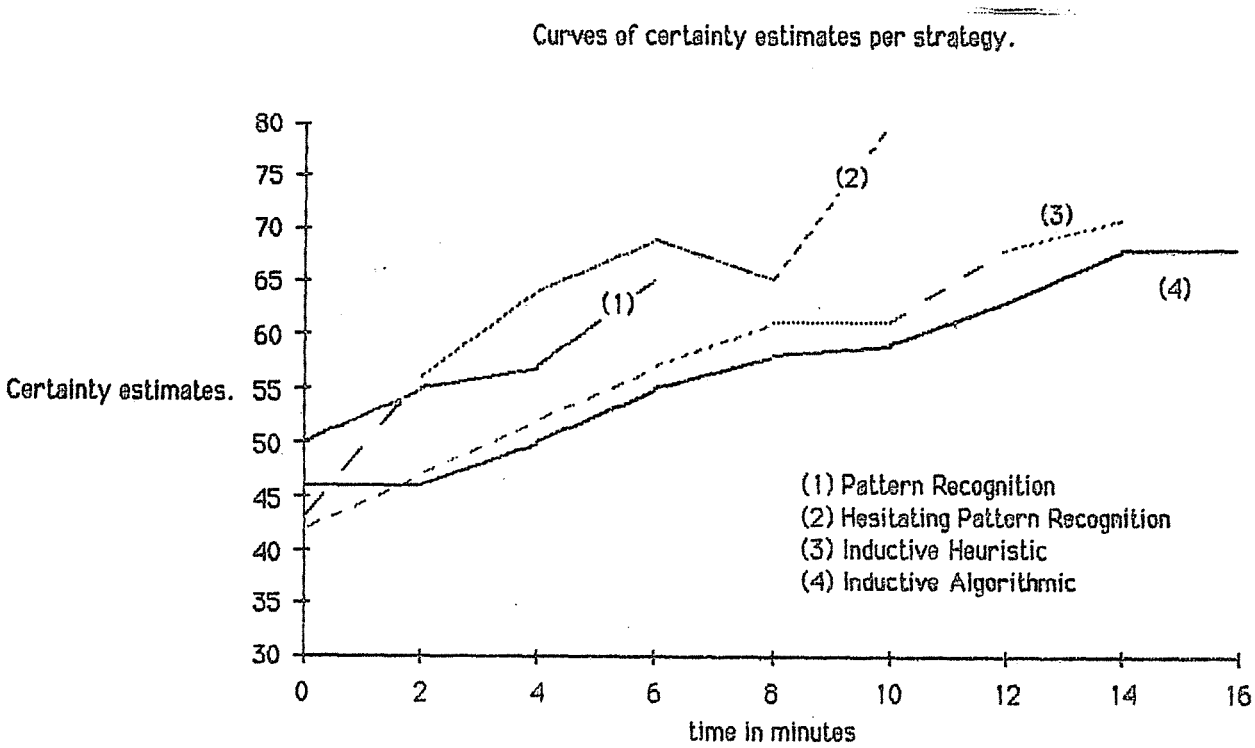
Four cases stand out: cases 4 and 6, for their high increase, and cases 3 and 7 for their low increase. The first ones started at a comparatively low level. But, obviously, the physicians found strong evidence to enhance confidence in their last hypothesis. As was discussed before, to find (or to invent) a strongly discriminating factor, a test or symptom with a high sensitivity (inclusion of real patients) and/or a high specificity (exclusion of non-patients) considerably reinforces the doctor's confidence in the proposed hypothesis. As we have seen (par.3) the T4 (thyroxine hormone serum level) test is such a symptom with regard to hyperthyreoidism (case 6). The same is true for vaginal examination in the case of extra uterine pregnancy (case 4). Apparently, such strongly discriminating factors could not be found in the cases of atopic eczema (case 3) or asthmatic-bronchitis (case 7). Although symptoms and tests with reasonable sensitivity and specificity values are mentioned in literature, they evidently are not in the forefront of awareness of the physicians.

More or less parellel with the rise (or fall) of confidence in the hypotheses is the physician's confidence in the problem-solving itself. The values for this type of confidence were called the uncertainty estimates. They express the estimation of one's own clinical competence to solve a particular problem. The values are expressions of one's subjective, emotional feelings of (un)certainty at a given phase of the process. These values were measured in the same manner as the previous (probability) ones: giving an estimation by means of a line in an open bar. In contrast with the probability estimates, people preferred to draw their lines of (un)certainty in a vertical bar, moving upwards with the rise of certainty and vice versa. The bar, again was 10 cm and a ruler could easily measure the values in percentages. The participating physician was urged to express his feelings of (un)certainty at two minutes intervals during the work-up, starting immediately after listening to the patient's complaint.

According to decision theory one of the essential aims of the decision-maker is to reduce uncertainty (18). According to Bohlinger & Ahlers (3) it is really one of the first steps in the decision making process.

So our first step was to calculate the mean values of the certainty estimate in the successive time score for each of the various types of inductive strategy. We hoped that the point of onset and the slope of the curve could tell something about the behaviour of the physicians applying the successive strategies. The results of this exercise are shown in Figure 6:

Figure 7



The curves show a gradual progression to a maximum. The main difference between the subtypes of the inductive strategy is the rate of increase as expressed by the slopes of the curves. The Pattern-recognition methods show a more steeper slope (45 degrees) than the other ones (both approx. 25 degrees).

In general,the curves show a gradual, rather uniform, progression to a maximum.The main difference between the subtypes of the Inductive strategy is the slope of the curves.For the Pattern-Recognition methods the slope (45 degrees) is steeper than those of both the Heuristic and the Algorithmic strategies(25 degrees).These curves sustain the predicted behaviour of the strategies:the more determined pattern for the Pattern Recognition modes,and the gradual spiralling to a maximum of the other two.

Table 43

CHANGES OF PHYSICIANS' ESTIMATES DURING THE WORKUP

	FIRST CERTAINTY ESTIMATES	LAST CERTAINTY ESTIMATES
PHYSICIANS ESTIMATING LESS THAN 0.5	64 %	14 %
PHYSICIANS ESTIMATING MORE THAN 0.5	36 %	86 %
TOTAL	100 %	100 %

In the beginning of the process 64 % of the physicians scored a certainty of less than 0.5. In the end 86 % of the physicians felt rather certain about the outcome of the process.

Decision making is reducing uncertainty in a problem situation (18). The physician starts almost always in an uncertain situation. Two-thirds of the physicians started at a 'certainty-level' of less than 50%. The confidence of one-third of the physicians (with certainty estimates of more than 50%) at the start of the problem solving procedure cannot be explained from the situational circumstances.We expect that these physicians have personal reasons to be confident in finding a solution.

In the end 86% of the physicians are rather certain about the achieved solution.This result corroborates our statement about the conclusive point of the inductive process:maximum probability with minimum uncertainty.

However, as we know from figure 6, complete certainty does not exist.We can only speak of a minimization of the uncertainty.Table 43 shows that 14 % of the physicians still remained pretty uncertain at the end of the road.

These figures support the opinions about physicians' work as they are laid down in literature: in the end uncertainty remains. The doctor in his daily practice does not know the feeling of complete certainty,because there is not an absolute proof of a single hypothesis nor can the possibility of another explanation be excluded with certainty. Perhaps this is more true for family physicians than for general internists, whose presented problems are, at least in Holland, a selection of the patients first seen by his colleague in primary care. However, the figures as presented in table 44 do not sustain this suggestion.

Table 44

FIRST AND LAST CERTAINTY ESTIMATES : FAMILY PHYSICIANS - GENERAL INTERNISTS.
(mean values)

CATEGORY	FIRST CERTAINTY ESTIMATE	ST.DEV.	LAST CERTAINTY ESTIMATE	ST.DEV.
FAMILY PHYSCNS	42.1	28.5	67.04	22.58
GENERAL INTERN.	45.3	33.7	75.71	19.64

Both groups of physicians score in the same ranges, for the first certainty estimates as well as for the last ones. According to the figures (un)certainty seems to be an innate characteristic of all kinds of physicians.

For the various kinds of inductive strategy the values are:

Table 45

CERTAINTY ESTIMATES PER STRATEGY
(mean values)

STRATEGY	FIRST CERTAINTY ESTIMATE	LAST CERTAINTY ESTIMATE
PATTERN RECOGNITION	48.9	67.4
HES.PATT.RECOGNIT.	41.1	70.3
INDUCTIVE HEURISTIC	41.1	67.0
INDUCTIVE ALGORITHM.	44.6	70.3

The boldness of the physicians employing the Pattern Recognition method in se contrasts with the more or less uniformly lower first certainty estimates of the other strategies. In the end they all come together.

We have to realize that the values of the second column (C.E.Last) do not represent the maximum within their category. Several wavering physicians were found at the end of the decision making process, and their estimates could influence the results.

Eventually, as we did before, we wondered as to the effect of case dependency with regard to the values of certainty.

Table 46

FIRST AND LAST CERTAINTY ESTIMATES PER PATIENT CASE

PATIENT CASE	FIRST CERTAINTY ESTIMATE	LAST CERTAINTY ESTIMATE	RATE of INCREASE %
1	41.9	56.0	34
2	34.8	65.8	89
3	37.6	69.6	85
4	36.4	64.8	78
5	45.4	70.9	56
6	41.5	69.9	69
7	54.0	77.2	43
8	48.0	68.0	42

The certainty estimates for the various cases show a heterogenous picture, as could be expected from a notation of the physician's personal feelings. Therefore, interpretations from these figures can only have a personal meaning. The more recognizable (prevalent) diseases as 5 and 7 score highest for the last certainty.

Expectation and confidence play a mysterious game. Highly confident physicians chose high estimates in cases 5, 7 and 8 but evidently their expectations were not fulfilled, according to their final certainty rate. Especially in case 1 (myocardial infarction) a moderate start resulted in a medium high estimation giving the lowest increase of all the cases. Is myocardial infarction indeed such a horror to a physician that even in simulation he trembles? We do not know and we cannot know from these clustered figures and values. We have seen people slaving away, perspiring, we have seen them happy and desperate, drifting from one end of the certainty scale to the other. We have seen them pleading and shouting to the "patient" (the poor simulator), but also begging, advising with all conviction they had. They addressed the young (male) simulator as 'madam' seriously inquiring about her marriage, her experience in childbirth; they argued with the 'patient', they sent the patient away with an advisory list to be fulfilled for the next week or else.....! In short, we have seen the physicians at work, with all their emotions and feelings, their hesitations and uncertainties. It was a tremendous experience. But all these things cannot be caught in figures, values, curves or statistics. And perhaps so much the better. The human being is so complex and so unique that dismembering that being into small parts would deprive him of all his charm.

Paragraph 5

PATIENT-MANAGEMENT AND THERAPY

Patient management and therapy are recognisable entities within the medical process, but these items are treated as Cinderella.

On the average, less than one minute of attention is given to this subject once a diagnosis is arrived at.

In contrast with the patients' views on the physician's professional activities, the decision maker himself apparently considers it as a negligible task.

Nevertheless, in this very minute a judgment and a decision is made with sometimes far-reaching consequences to patients and physicians. Within the diagnostic process, as it were, the therapy has already been decided upon. It may be considered as putting a full stop at the end of a sentence. In this conception each therapy and treatment can only be understood within the context of the whole medical process. It is not this one minute that counts but the preceding process. As we discussed in chapter IV it is the prognosis, the prediction of future development of the disease and the diseased state of the patient that really matters. Treatment might be seen as suggestive for a direction towards a better or preferably, the best health condition. Because predictive arguments can only be borrowed from the results of the diagnostic process, it is self-evident that treatment only takes up a small and brief part of the medical process. In the next paragraph we shall see that the utter personal character of the diagnostic process - and as a consequence the prediction - is highly suggestive for this concept.

In this paragraph we shall discuss a number of aspects which are regularly met within this "treatment - part" of the medical process. To elicit these aspects we subdivided this part of the "paper patient" into three classes (see also chapter VII):

a.) the description of all possible actions a physician can take in therapeutic management;

b.) the description of referral destinations and types of consultation;

c.) a detailed list of drugs, classified according to their generic names. This class also includes all kinds of fysiotherapeutic, ergotherapeutic and psychotherapeutic treatment and a not exhaustive list of alternative drugs and treatments.

To explain the first two classes we have to recollect that in the Netherlands the family physician is the sole entrance to health and or social care.

Therefore, a number of actions alludes to referrals and their destinations and to consultations of various practitioners and professionals in health and/or social care. Several actions in this class are self-evident: advising, medication, dietary advices, surgical or psychotherapeutical treatment, but also action to gain information from laboratories, X-ray departments, institutions etc. The destinations of referrals and consultations are listed and arranged according to a pre-established list of medical and paramedical workers.

The third class is mainly composed of a list of medicaments and treatments annually published by the Royal Dutch Association of Pharmacists. Because we used a limited number of patient scenarios, (and diseases) only a small portion of the 157 items of these classes were employed.

It may be clear from the preceding description that we distinguish two features within the treatment stage, a phase of determination of the action, the direction of the treatment, the management of the disease process; and a phase of completion of the contents of the treatment, the medication. It is essential

to make this distinction because some actions cannot be understood without this division.

In the investigation we made a distinction between the time till diagnosis and the time of completion of each problem solving task, the 'total time'. "Therapy time", can be established by its subtraction. As we recorded time in minutes, we are only able to measure into this unit of measurement.

Out of 253 records 128 come to a 'therapy time' of less than one minute; that means that no difference was recorded in the data base between time till diagnosis and total time. The remaining 125 records scored as follows:

Table 47

"THERAPY - TIME" FOR THE MORE ELABORATE THERAPEUTICAL CASES (125)

	MEAN TIMES	ST.DEV.
TIME TILL DIAGNOSIS	11.02'	5.41
TOTAL TIME	12.46'	5.50
THERAPY TIME	1.44'	

one minute and three-quarters are appropriated for the establishment of patient management and treatment for cases for which a more elaborate deliberation was needed.

From our conception of a problem-oriented task environment we could assume a variation of therapy-time over the cases. First we presented a table with a distribution over the cases for the two main categories: one with a therapy-time of less than one minute, the other with more than that.

Table 48

"THERAPY - TIME" : MORE OR LESS THAN ONE MINUTE PER CASE

PATIENT	LESS THAN ONE MINUTE		MORE THAN ONE MINUTE	
CASE	NUMBER	%	NUMBER	%
1	22	71	9	29
2	12	39	19	61
3	10	30	23	70
4	21	64	12	36
5	23	70	10	30
6	16	50	16	50
7	10	33	20	67
8	14	47	16	53

Cases requiring an (extensive) prescription take more time than cases which demand a quick referral to a hospital. Cases like Myocardial Infarction (1) and Extra-Uterine Pregnancy (4) which ask for an immediate referral score less than one minute, Atopic Eczema (3) and Asthmatic Bronchitis (7) score higher because of more elaborate prescriptions.

The mean therapy-time for the more-than-one--minute group is presented in table 49

Table 49

MEAN "THERAPY - TIME" FOR THE MORE-THAN-ONE-MINUTE GROUP PER CASE.
(125 records)(mean values)

PATIENT CASE	TIME TILL DIAGNOSIS	TOTAL TIME	THERAPY TIME
1	10.8	12.1	1.3
2	12.0	13.6	1.6
3	8.5	10.3	1.8
4	11.8	13.2	1.4
5	11.1	12.4	1.3
6	15.8	16.9	1.1
7	7.6	8.9	1.3
8	12.5	13.9	1.4

Apart from some slight variations across the cases significant differences were not found, or expected.

Apart from some slight variation, real differences were not found.
We also matched 'therapy-time' with the various subtypes of the Inductive strategy.

Table 50

"THERAPY - TIME" PER STRATEGY

STRATEGY	TIME TILL DIAGNOSIS	TOTAL TIME	MORE THAN 1 MINUTE		THERAPY TIME
	mean values	mean values	cases	%	mean values
PATTERN RECOGNIT.	5.00	6.13	15	56	1.13
HES.PATT.RECOGN.	10.92	12.69	13	50	1.77
INDUCT.HEURIST.	10.70	12.14	84	49	1.44
INDUCT.ALGORITHM.	20.08	21.54	13	46	1.46

Fast working people seem also to have the quickest therapy time (pattern recognition) in contrast with their more hesitating colleagues who react conform their (chosen) title. The other strategies come approximately to the same score.

While the inductive-heuristic and inductive algorithmic strategies occupy a mid-position, both Pattern-Recognisition strategies are at the extremes. The P.R.-strategy in se, according to its overall outlook, results in the shortest time and, apparently, it takes the quickest decision. On the other hand, the hesitating pattern-recognitioner honours its name by taking the largest therapy-time of all strategies. However, we must bear in mind that these figures

stand for the more deliberate group of physicians representing only half of the total group of physicians.

What do these physicians do during this short period? Are they asking questions about the patient's condition, e.g. his tolerance of certain drugs, allergies to certain medicaments, former drug reactions etc.?

We found that four out of 68 participants asked questions after having arrived at the diagnosis with a total of 15 questions. Of these 15 questions 10 were about the patient's social condition, like his experience of labour or living situation, 2 about drug allergies and three were not classifiable within this context.

The largest part of the therapy-time is destined for the decision which action has to be taken and to determine the content of this action.

Of the 31 possible actions listed in the "paper patient" the participants used 8 options. These options are:

1. Advising (Adv) e.g. about life-style, staying in bed, stop smoking etc.

2. Prescription (Pr) : prescribing medicaments in a broad sense,

3. Dietary Advice (Die) : advising and/or prescribing a more or less specific diet.

4. Laboratory Information (Lab) : to gain more specific laboratory information not immediately available to the physician e.g. serological, bacteriological or very complex biochemical tests.

5. X-ray Information (XR) : asking more complex and specific information from X-ray departments.

6. Referral to another physician , usually a specialist. (Ref).

7. Referral to a paramedical worker (RpW) e.g. a fysiotherapist, district nurse, midwife etc.

8. Consultation of another physician (Con) without referring the patient and/or the physician's responsibility to his colleague.

We listed the kind of actions for each patient's case in the next table:

Table 51

KINDS OF ACTIONS

CASE	nr	ADV	PR	DIE	LAB	XR	REF	RPW	CON	Nr of cases
1	Nr	2	2	-	1	-	27	-	-	31
	%	6	6	-	3	-	87	-	-	
2	Nr	18	20	-	-	-	4	8	1	31
	%	58	65	-	-	-	13	26	3	
3	Nr	3	30	-	1	-	4	-	-	33
	%	9	91	-	3	-	12	-	-	
4	Nr	2	3	-	-	-	29	-	-	33
	%%	6	9	-	-	-	88	-	-	
5	Nr	3	7	7	2	3	22	-	1	33
	%	9	13	-	9	3	78	-	6	32
6	Nr	3	4	-	3	1	25	-	2	32
	%	9	13	-	9	3	78	-	5	
7	Nr	4	30	-	-	-	-	-	1	30
	%	13	100	-	-	-	-	-	3	30
8	Nr	5	19	-	2	-	8	-	-	30
	%	17	63	-	7	-	27	-	-	

adv= advice; pr= prescription; die= diet; lab= laboratory tests; xr= x-ray; ref= referral to a physician; rpw= referral to a paramedical worker; con= consultation.

Preferences of therapeutic actions are clearly case dependent: cases 1,4,5,6 for referral, cases 2,3,7,8 for prescription

We must realize that several actions can be taken for the same case. E.g. advising, prescribing and dietary measures can perfectly go together, as well as advice together with laboratory and/or Xray information etc.

Therefore, in all cases the number of actions amount to more than 100%.

The figures largely indicate that the action taken is problem-oriented, which is supports our conception that the therapeutic action is more linked to the preceding process than to the outcome of it. The two most striking actions, prescription and referral, are, as a matter-of-course, opposing.

Consistency of action across the physicians can be found in cases 1, 4 and 6 (referral) and 3 and 7 (prescription). The physician's confidence in his competence to handle and manage the situation seems to reflect the choice of action. In case 4 it is obvious that the needed surgical treatment is beyond the possibilities and capabilities of family physician and general internist. For case 1, the myocardial infarction case, there is some general discussion in medical literature about the desirability to refer these patients to a hospital or to treat them in their own environment at home.Still, the overall trend is to refer the patient to a coronary care unit. However, in case 6, hyperthyreoidism, it seems that the family physician's lack of knowledge and confidence determined his decision.

Cases 3 (atopic aczema) and 7 (asthmatic bronchitis) are obviously within the range of the (family) physician's competence. We have to admit that the consequences of misdiagnosis and/or indequate therapy are less severe than, for instance, in cases 1 and 4. Possibly these considerations play a role in the physician's decision. In the other cases less consensus of the various actions is reached across the different physicians. However, we have to admit that not all physicians arrived at the same diagnosis. So disagreement of diagnosis can also contribute to inconsistency.

The majority of treatment plans concerns two types of actions: prescription and referral. A referral (to another physician) was established in 129 (out of 253) cases, 51%. In 10 cases the patient was referred to two specialists at the same time. In case of atopic eczema e.g., two patients were referred to a dermatologist as well as an allergologist; in case of low back pain the patient was referred to both a neurologist and an orthopedic surgeon; in case of gallstones to an internnist and a surgeon, and in case of anaemia to a hematologist and a gynaecologist. We listed the destinations of referral in the next table.

Table 52

REFERRAL DESTINATIONS FOR THE VARIOUS PATIENT CASES

PATIENT CASE	NUMBER OF REFERRALS	DOUBLE REFERRALS	DESTINATION
1	28	1	Int.medicine (1), Cardiology (27)
2	6	2	Orthopedic Surgery (2), Neurology (4)
3	6	2	Allergology (2), Dermatology (4)
4	29	-	Gynaecology
5	24	2	Intern.Medicine (12), Surgery (12)
6	26	1	Inter.Medicine (25), Gastro-enterology (1)
8	10	2	Intern.Med.(7), Gastro-ent.(1), Gynaecol.(2)

Between brackets the number of referrals to the particular clinical specialty. In the cases of "double referral" the patient was sent to two different consultants.

The differences between the destinations can partly be contributed to the varying diagnoses. Sometimes, however, it is a matter of opinion. In the case of the gallstones half of the participants considered surgical treatment the best solution to the case, while the other half thought of consulting the internist as a first choice. Sometimes one or more of the symptoms displayed by the patient determined the referral destination. In the case of hyperthyreoidism the patient complained about some gastro-intestinal troubles. One physician considered this as a clue to a referral to a gastro-enterologist. The patient with anaemia suffered from menorrhagia (severe blood-loss during menstruation). Two physicians correctly referred the patient to a gynaecologist.

Another question is whether the physician considers the referral as a delegation of responsibility or not. We asked the participant whether he would keep in contact with the patient (asking the patient to return to his office

after the specialist's examination) or leave it to the patient to contact him or not. The next table presents the opinions of the participants per case with regard to 'revisits' in case of referral.

Table 53

EXTENSION OF THE PHYSICIAN'S RESPONSIBILITY IN CASES OF REFERRAL.
(asking a patient to revisit)

PATIENT CASE	REVISIT number	%	NO REVISIT number	%	NR of REFERRAL
1	3	11	24	89	27
2	-	-	4	100	4
3	-	-	4	100	4
4	1	3	28	97	29
5	2	9	20	91	22
6	5	20	20	80	25
7	-	-	-	-	-
8	2	25	6	75	8
TOTALS	13	11	106	89	119

Physician's request for revisit after completion of the referral consultation is limited to approximately 10% of the cases. Apparently medical responsibility is not extended to the consultants' offices.

In general, the participating physicians considered the referral as a delegation of their medical responsibility. Eleven percent of the physicians explicitly asked the patient to contact them after the specialist's consultation. The same was asked with prescriptions.

Table 54

PHYSICIAN'S RESPONSIBILITY FOR TREATMENT EVALUATION AFTER PRESCRIPTION.
(revisits after prescription)

PATIENT CASE	REVISIT number	%	NO REVISIT number	%	NUMBER of cases
1	1	50	1	50	2
2	19	95	1	5	20
3	28	93	2	7	30
4	1	33	2	67	3
5	4	57	3	43	7
6	4	100	-	-	4
7	25	83	5	17	30
8	19	100	-	-	19

Physician's requests to revisit for evaluation of the treatment were noted in 88% of the pertaining cases. This picture is the reverse side of the 'referral picture'. Apparently, responsibility is considered as a unique patient - physician condition, not a general medical professional one.

In this case 88% of all patients having been prescribed drugs were asked to revisit their physician, mainly between 3 and 7 days after their first visit (which was the condition of the simulation)

In total 135 (53%) patients were explicitly asked to revisit, and 118 (47%) were not. Apparently,responsibility is considered as a unique patient-physician deal, not to be delegated beyond this relationship, and not to be seen as a general medical condition.

As mentioned before, not all participants reached the same diagnosis in a particular patient case. It means that comparisons between the precription of the various physicians are hard to obtain. In a minority of cases a sufficient number of equal diagnoses was found to allow comparisons among the prescriptions. We can only present an analysis of the prescription for the cases 3, 7 and 8. For case 3 (atopic eczema & bronchospasms) we found 19 records of approx. the same diagnosis. The 19 physicians prescribed 9 different drugs, mostly in combinations(17 times). In two records a single prescription was provided. The next table shows the frequency of prescriptions for the various medicaments by these 19 physicians. The medicaments are mentioned by their pharmacological group names. In 7 records more than 2 medicaments were prescribed.

Table 55

PRESCRIPTIONS FOR A CASE OF ATOPY

NAME MEDICAMENT-GROUP	NUMBERS	%
CORTICOTROPINES*	15	31
SYMPATHICOMIMETICS	11	22
ANTIHISTAMINICS	8	16
BASIC COMPONENTS FOR CREAMS	7	16
ANTIBIOTICS : TETRACYCLINES	3	6
ANTIPRURITICS	2	4
SULFONAMIDES	1	2
SPASMOLYTICS : XANTHINE-DERIVATIVES	1	2
HOMOEOTHERAPY	1	2

* Corticotropines were prescribed in 11 records. In three records they were prescribed more than once.

For case number 7, asthmatic bronchitis, the same procedure was followed. 21 records of approx. the same diagnosis were found. The 21 physicians prescribed drugs from 9 different classes, for the larger part in combination (16) and 5 in single prescription. In 7 records more than 2 medicaments were prescribed.

Table 56

PRESCRIPTIONS FOR A CASE OF ASTHMATIC BRONCHITIS

NAME MEDICAMENT-GROUP	NUMBER	%
ANTIBIOTICS : TETRACYCLINES	9	19
EXPECTORANS / MUCOLYTICUM	9	19
ANTIBIOTIC : BROAD SPECTRUM PENICILLINES	8	17
SPASMOLYTICS : XANTHINE-DERIVATIVES	5	11
ADRENERGICA: SYMPATHICOMIMETICS	5	11
EMOLLIENTIA	4	9
COUGH DEPRESSING DRUGS	3	6
CORTICOTROPINES	2	4
ANTIHISTAMINICS	2	4

In case 8, iron-deficiency anaemia, a very convincing consensus among the physicians was found. We found 16 records with the same diagnosis. In all these cases a single prescription was provided. All physicians recommended to administer an iron (ferro) derivative orally.

Surveying these three cases, we get the impression that most of the prescriptions are symptom-oriented. This is especially obvious in the case of iron-deficiency anaemia. But in the other cases, one can hardly speak of a cause-related therapy. Although corticotropines are outstanding suppressors of allergic reactions, they are only suppressing a symptom or a cluster of symptoms. The antibiotics prescribed in the case of asthmatic-bronchitis may be excellent medicaments but by their nature-(broad-spectrum)-hardly specific.

This impression is substantiated when surveying the various prescriptions for a variety of diseases as mentioned in the group of participants. The personal appreciation and interpretation of the various diseases and their symptoms seems to lead to the variety of drugs. This, in its turn, is consistent with the philosophical backgrounds of inductivism.

Paragraph 6

EXPERIENCE AND CONSEQUENCES

It is widely believed that experience improves the clinical competence and the inference qualities of physicians. For some authors this belief is reason enough to reject simulated examinination (e.g. by means of Patient Management Problems) for measuring clinical competence as the performance on the PMP's is negatively related to experience at the postgraduate level (28). The notion of expertise rests on the presumption that diagnostic skill is an innate trait of experienced physicians that is generalized across clinical situations. However, Elstein and co-workers (1) failed to demonstrate any difference between 17 judged by their peers to be highly proficient diagnosticians ("criterial group") and seven others who were not so nominated ("non-criterial group"). These findings do not support the widespread belief that designation of expertise by medical peers is well-founded and well-validated (2). It also questions the expertise itself. According to Brehmer (29), improvement of clinical competence and expertise on the basis of clinical experience is founded on an incorrect conception of the nature of experience. Indeed, scrutinizing this conception leads to a more pessimistic view about people's ability to learn from experience (see also 30). Prejudice prevents physicians to learn from the information experience can provide. Some of these biases are based on the personalistic and inductive character of the clinical judgment.

In this paragraph we shall match some of the aforementioned variables to the experience of the participants. The experience is expressed in terms of years in medical practice, in day-to-day service in health care. In paragraph 4 we indicated that any trend in experience-related estimation for prevalences, incidences and disease consultation values could not be detected.

In this paragraph we split the group of participants into three compartments: the younger, expected to be the non-experienced-group of physicians with a practice experience between 0 and 5 years, a middle group with 6 to 20 years in practice, and a group of physicians more than 20 years in health care service. These groups represent respectively 11, 67 and 22 percent of the total of 253 records.

These figures can be more or less compared with the frequency distribution within the research sample. When experience plays a part in gaining insight, knowledge or competence in the medical problem-solving skill we may expect a more-or-less linear trend across the various variables. In the next tables we shall present a number of these variables matched with the three "experience groups".

One of the assumptions made in literature (e.g. 13) is that the more senior the doctor, the more relevant information is asked in the least number of questions. As described in paragraph 2 of this chapter we defined a relevancy factor from the symptoms which could be allocated to a hypothesis and those which could not.

The symptoms which could be attributed to a particular hypothesis were called, Hypothesis Prerelated Symptoms (HPS) and the Hypothesis Related Symptoms (HRS). These symptoms contrast with the ones which could not be attributed to these (or other) hypotheses, Non-Hypothesis-Related Symptoms, (NHRS). The relevancy factor thus could be delineated as

$$\frac{HPS + HRS}{HPS + HRS + NHRS} * 100$$

The more relevant to a particular hypothesis the symptoms were asked the higher the factor. Confirmation of the assumption means that we may expect an upward tendency of the relevancy factor with increasing experience. The second part for the suggestions predicts a downward tendency of the number of questions with increasing experience.

Table 57.

PHYSICIANS' EXPERIENCE VERSUS RELEVANCY FACTOR

EXPERIENCE GROUPS	RELEVANCY FACTOR
0 - 5 years	61.2
6 - 20 years	65.5
> 20 years	68.5

The assumption that more experience leads to the acquisition of more relevant information can scarcely be sustained when matched with the Relevancy Factor.

Table 57a

PHYSICIANS' EXPERIENCE VERSUS MEAN NUMBER OF QUESTIONS

EXPERIENCE GROUPS	NUMBER OF QUESTIONS	ST.DEV.
0 - 5 years	32.04	11.08
6 - 20 years	43.49	21.31
> 20 years	31.66	15.13

The thought that more experienced physicians ask less questions (but more relevant ones) than their younger colleagues cannot be confirmed by these figures.

The first part of the hypothesis seems to be slightly corroborated. The second part, however, does not sustain the assumption. In fact, significant differences with respect to the total number of questions between the three groups could not be traced.

To test the possibility that the "oldest" group took a more leisurely time-schedule, we also matched the groups with the mean total time used for solving a patient case.

Table 58

PHYSICIANS' EXPERIENCE VERSUS MEAN TOTAL TIME

EXPERIENCE GROUPS	TOTAL TIME (mean)	ST.DEV.
0 - 5 years	11.81	3.75
6 - 20 years	11.79	6.05
> 20 years	12.08	4.97

Experience does not influence the time needed to solve a patient's problem.

Again we could not find differences between the groups with regard to time-spending.

Another question is whether more experienced physicians are more elaborate in their generation of hypotheses. In literature different opinions are found. Some authors believe the experts to be more specific in their work-up with correspondingly less hypotheses, while others assume a different attitude. Our results are presented in table 59.

Table 59

PHYSICIANS' EXPERIENCE AND THE GENERATTION OF HYPOTHESES

EXPERIENCE GROUPS	NUMBER OF HYPOTHESES (mean values)	ST.DEV.
0 - 5 years	3.55	1.65
6 - 20 years	2.89	1.55
> 20 years	2.79	1.12

From the mean number of hypotheses and the standard deviations we may conclude a tendency towards the generation of less hypotheses with increasing experience.

Although differences are small,we may assume a tendency to the opinion that expertise parallels specificity of the work-up. We wondered whether this notion is reflected in the subjective estimates of prior and posterior probabilities of the hypotheses. We may remember from paragraph 4 of this chapter that it was assumed that the prior probability reflected the prevalence (or incidence) rates of the supposed disease, and the posterior probabilities as the adjustment of the prior probability considering the acquired information. This view is widely held in primary health care that unexperienced family physicians lack the knowledge and the insight in the frequency distribution of diseases and their symptoms in more or less specified populations. In terms of decision making jargon it can be denoted as the unexperienced physician being unacquainted with the P[D] and the P[S|D], and their counterparts P[d] (=nonD), P[s|d] (=nonS and nonD). We matched the figures of prior probability and posterior probability, as they are presented in paragraph 4, with the groups of experience

Table 60

PHYSICIAN'S EXPERIENCE AND THEIR ESTIMATES OF PRIOR AND POSTERIOR PROBABILITIES

EXPERIENCE GROUPS	FIRST PROBABILITY		LAST PROBABILITY	
	MEAN	ST.DEV.	MEAN	ST.DEV.
0 - 5 years	39.26	24.01	73.26	23.77
6 - 20 years	47.58	20.24	73.85	25.20
> 20 years	36.43	20.31	74.38	22.94

Remarkably, the group of physicians with an experience in medical practice of 6 to 20 years estimated the prior probabilities for their first hypotheses higher than the other groups. On the diagnostic level the posterior probability estimates bring the groups together.

As we have seen in figures 3, 4, 5, the values for prior probabilities of the "youngest" and the "oldest" group astonishingly resemble each other, while

"middle" group takes a different position. With regard to the posterior probabilities, only an optimist will observe a tendency towards increasing experience.

Probability-estimates,as weighing factors for the hypotheses, sometimes reflect the state of (un)certainty of the physician. We matched therefore this variable also with the experience groups.

Table 61

PHYSICIANS' EXPERIENCE AND THEIR FEELINGS OF UNCERTAINTY

EXPERIENCE GROUPS	FIRST CERTAINTY ESTIMATE		LAST CERTAINTY ESTIMATE	
	mean values	st.deviat.	mean values	st.deviat.
0 - 5 years	45.44	27.70	61.33	26.34
6 - 20 years	44.12	29.32	69.59	21.34
> 20 years	35.36	27.63	65.32	23.30

Primary uncertainty is highest for the most experienced physicians, although the values for standard deviations make firm conclusions hazardous. The last certainty estimates (on diagnostic level) are approx. on even level.

The most striking phenomenon of this table is the rather high degree of primary uncertainty among the most experienced group of physicians. It may be an artefact of the investigation. The older physicians are, on the average, the less they are accustomed to being observed - especially with videocameras - in comparison with their younger collegues, who are acquainted with this type of equipment. However, a plausible explanation may also be, that more experienced physicians are more careful and prudent. From table 57 we know that they use slightly more time than their other collegues. Why this is so, has to be investigated as yet.

With regard to management and treatment one might expect a more restraint in the "older" group towards referral of patients to hospitals and specialists. Their knowledge and insight on prognosis, course of illness and the various reactions of drugs are greater than that of their less experienced colleagues. We matched two features of therapeutic action, prescription and referral, with the three experienced groups.

Table 62

PHYSICIANS' EXPERIENCE AND TREATMENT PLANS FOR PRESCRIPTION AND REFERRAL

EXPERIENCE GROUPS	PRESCRIPTION	REFERRAL
0 - 5 years	48 %	52 %
6 - 20 years	49 %	51 %
> 20 years	50 %	50 %

(The percentages are calculated from the number of the particular actions divided by the total number of solved cases within that particular experience group)
Differences of conception with regard to patient management between the experience groups are not found.

Eventually, we wanted to know the differences in strategy across the various experienced groups. We hypothesized that in the more experience group pattern-recognition would play a more distinct rôle than in the less experienced group. Furthermore, we assumed a more elaborate work-up of the cases in the "younger" group because of their rather recent background of medical education.

That means a relatively larger portion of the more "specialistic" way of inference, in our nomenclature the inductive-algorithmic method.

In the next table the row of percentages represent the proportions of that type within the "experience group". The column percentages represent the proportion of the types across the "experience groups". To allow comparisons the three groups were standardized to 100.

Table 63

PHYSICIAN'S EXPERIENCE AND THEIR EMPLOYMENT OF THE STRATEGIES

STRATEGY	PATT.RECOGN.	HES.PATTT.REC.	INDUCT.HEUR.	INDUCT.ALGOR.	TOTAL
EXPERIENCE GROUPS	row % col %	row % col %	row % col %	row % col %	
0 - 5 yrs	7.4 27.5	7.4 27.0	85.1 37.7	- -	100
6 - 20 yrs	12.3 45.7	11.2 40.7	62.0 27.5	14.7 73.5	100
> 20 yrs	7.2 26.8	8.9 32.3	78.7 34.8	5.3 26.5	100
TOTAL	100	100	100	100	

The "6 - 20 years experience group" counts the largest number of physicians using the Pattern Recognition method, and also the largest number of doctors employing the Inductive Algorithmic strategy. The similarities between the "youngest" and the "oldest" group are striking.

From these figures we must conclude, that our hypotheses have to be rejected. For the least experienced group the most employed strategy is the inductive-heuristic one, while the inductive-algorithmic method is largely used in the middle group. This group also counts, proportionally, the largest number of "pattern-recognitioners". There is hardly any difference between the 'oldest' and the 'youngest' group as far as three of the four strategies are concerned. The inductive-algorithmic strategy represents an exception to the rule, which can largely be attributed to the number of general internists constituting this group. For all the experienced groups the inductive heuristic strategy is the most employed method. The other strategies play a minor, although marked, role. Informal information shows that pattern-recognition plays a much larger role in the daily practice of a busy family-physician.

How busy is a physician in his curing and caring that he can allow himself some continuing eduction? Our last question with regard to age and experience was, how much time was spent in formal continuing education i.e. the time spent in official courses. The time is noted in days per two year.

Table 64

PHYSICIANS' EXPERIENCE AND THE CONTINUING EDUCATION

EXPERIENCE GROUP	TIME SPENT IN COURSES* (mean values)	ST.DEV.
0 - 5 years	6.71 days	4.5 days
6 - 20 years	11.52 days	8.0 days
> 20 years	9.2 days	20.2 days

* the figures refer to a period of two years.
Apparently, the physician's need for continuing education is greatest when he is between six and twenty years in practice. Continuing education seems less attractive to 'younger' and 'older' doctors.

These results reflect our presumptions that the younger group experiences less need to continuing education than the older colleagues. However, the extent of the time spent in continuing education surpassed our expectations. Approximately one week a year the physician attends official courses for continuing education. Whether this education increases medical competence or diagnostic or therapeutic consistency among physicians is to be questioned.

As mentioned in chapter 1, the observation of the lack of consistency among family physicians was one of the major incentives to start this investigation. Observing and realizing the personal character of the problem solving activities and their outcomes, the more curious we became about the diagnoses and their contents when strict uniform patient cases were offered. It was our strict intention not to judge the performances of the participants. We were interested in the process, not the product. Besides, according to which standards should we judge a physician knowing that practically all judgments have a personal nature? The patients' stories and qualities which composed the patient scenarios for the simulation were all derived from individuals for whom no conclusive diagnosis was made at that time. In actual health care we meet the same situation.

However, another question is whether the physicians, given a uniform patient case, shall come to equal conclusions; and further, whether the composing elements of this conclusion (= diagnosis) will broadly cover a uniform cluster of symptoms.

The former question was approached in the following way:

Because of the nature of the study we cannot claim one type of diagnosis as correct, and at the same time proclaiming another judgment as incorrect. In this study we only have to observe gradations between agreement on one name of a disease and the lack of any common denominator. For practical reasons we chose four stages of gradation:

DIAGNOSTIC GRADATION

1.) One joint name of a disease;

2.) Various names of diseases belonging to the same organ system;

3.) (various) names of diseases belonging to an adjacent organ system;

4.) (various) names of diseases not belonging to the organ systems mentioned in 2 and 3 and having less than 25% of the symptoms mentioned in 1.) in common.

Returning to the question as mentioned within the continuing education item, we matched these stages with three classes of continuing eduction. The mean number of days spent in official courses is 9.2 days. Remember that among these physicians some of them did not attend any course at all. This leads to the following categorization:

a.) physicians who spent less than 9.2 days per 2 years in official courses;

b.) physicians who spent more than 9.2 days per 2 years in official courses;

c.) physicians who did not attend continuing education.Because of preliminary inclusion criteria this class could only be split off from the preceding ones constituting an 'extra' number of records.

The matching led to the following results.

Table 65

DIAGNOSTIC GRADATION VERSUS CONTINUING EDUCATION

DIAGNOSTIC GRADATION	CONTINUING EDUCATION #					
	LESS THAN 9.2 DAYS		MORE THAN 9.2 DAYS		NO COURSES	
	number	%	number	%	number	%
1	60 *	48	61	48	18	43
2	47	38	42	33	16	38
3	10	8	13	10	4	9.5
4	8	6	12	9	4	9.5
TOTAL	125	100	128	100	42	100

* number of records = number of solved 'patients'
# continuing education refers to official courses within a two-year period
From this table follows that whatever amount of continuing education is attended or not, improved consensus of diagnoses does not automatically result.

Apparently, consensus of diagnostics does not improve by continuing education. We can think of several explanations:

a.) courses do not focus on diagnostic procedures;
b.) courses are more concerned with new developments in medicine;
c.) courses have not been validated towards routine practice situations.

We cannot specify this effect from our material. It may be recommended to encourage further studies in this field.

The problem-orientation of the clinical inference, as met within the preceding paragraph, may easily lead to the prediction that the diagnostic gradation will show marked differences among the outcomes of the various patient cases.

Table 66

DIAGNOSTIC GRADATION AND CASE - DEPENDENCY

PATIENT CASE	1		2		3		4		5		6		7		8	
	nr	%	nr	%	nr	%	nr	%	nr	%	nr	%	nr	%	nr	%
DIAGNOSTIC GRADATION																
1	20	65	10	32	9	27	24	73	21	64	27	84	13	43	12	40
2	9	29	14	45	18	55	1	3	9	27	-	-	12	40	11	37
3	1	3	4	13	5	15	7	21	2	6	-	-	4	13	-	-
4	1	3	3	10	1	3	1	3	1	3	5	16	1	3	7	23
TOTAL	31	100	31	100	33	100	33	100	33	100	32	100	30	100	30	100

Focussing on diagnostic gradation (types of corresponding diagnoses or clusters of symptoms) we can see a great variability of the results among the various cases. It seems that consensus of opinion about diagnoses and symptom-clusters is problem-oriented.

It seems that especially for the cases 2 (ischialgia), 3 (atopic eczema), 7 (asthmatic bronchitis) and 8 (iron-deficiency anaemia) the medical nomenclature is rather confusing and/or the physician's designation of this particular case to a well-defined disease class is difficult. The more clear-defined disease entities like hyperthyreoidism score higher agreement on a common denominator than the aforementioned ones, although the latter ones are much more prevalent in the population. This finding sustains the notion of Croft that there is a lack of standard medical definitions (25).

However, another explanation may be that some problems are more difficult to solve than others. Assuming that experience is a trend towards increasing diagnostic skills, we may hypothesize that the most experienced group, more than 20 years in practical health care, should perform on a more consistent level than the other groups.

Table 67

PHYSICIANS' EXPERIENCE VERSUS DIAGNOSTIC GRADATION

EXPERIENCE GROUPS / Diagn. Gradation	0 - 5 YEARS number	%	6 - 20 YEARS number	%	MORE THAN 20 YEARS number	%
1	13	48	87	51	21	38
2	7	26	59	35	23	41
3	3	11	13	8	7	13
4	4	15	11	6	5	9
TOTAL	27	100	170	100	56	100

Consistency of diagnostic nomenclature among physicians does not necessarily increase with advanced experience.

These findings corroborate Brehmer's statement that experience does not improve clinical judgment (29). An alternative explanation may be the more variable nomenclature among more experienced physicians. However, this leads to the same confusion. Because of the structure of this study, we cannot give a definitive judgment about this matter. It would be interesting to study this phenomenon in greater detail.

In a similar context it is interesting to know whether the opinions about diseases and the nomenclature of family physicians differ from those of general internists. We may assume that the concepts about several clinical diseases of general internists are much more uniform than those of family physicians as they are confronted with a much wider variety of diseases, illnesses and social troubles.

Table 68

DIAGNOSTIC GRADATION FOR FAMILY PHYSICIANS AND GENERAL INTERNISTS

DIAGNOSTIC GRADATION	FAMILY PHYSICIANS number	%	GENERAL INTERNISTS number	%
1	107	46	14	67
2	83	36	6	29
3	23	10	-	-
4	19	8	1	5
TOTAL	232	100	21	100

Consensus of diagnostic nomenclature is greater among the general internists than among the family physicians. This can be explained by the great variety of diseases, illnesses and ailments the family physician is confronted with.

Although a conclusive judgment cannot be given, these results seem to confirm the statement. This phenomenon is especially important in a number of circumstances. E.g. when the outcomes of the health care work of family physicians and/of specialist/hospital based care are compared one must be aware of this deviation in medical definitions; in case of epidemiological studies and medical classifications the results obtained from primary health care and specialist and/or hospital care cannot automatically be matched or mixed.

We tried to create a similar gradation system for the therapy-part of the medical process. But here we lack a more or less generally accepted nomenclature. Also, there is no general agreement about optimal treatment; at most one can find recommendations and personal preferences for most of the diseases described in textbooks. A joint denominator on therapy cannot be expected to be found. To bypass this obstacle we used a book in the possession of all physicians, because it was sent to them free of charge. It is the counterpart of a textbook known in the U.S.A. as "Current Medical Diagnosis and Treatment", and revised, translated and published in Holland in 1983. For each disease a concise treatment recommendation is advised. Our gradation for therapy runs as follows:

a.) the advice of the book is followed completely or at least for three-quarters;

b.) half or more of the recommendation for the particular disease are followed;

c.) less than half of the treatment advices are prescribed by the participant.

d.) neither a,b or c can be applied.

We are aware of the vulnaribility of this system but at least it will give some indication, some awareness of the physician's behaviour.

Within this context we give some figures matching the diagnostic gradation with its therapeutic counterpart.

Table 69

DISTRIBUTION AND CONSISTENCY FOR THE VARIABLES : DIAGNOSTIC AND THERAPEUTIC GRADATION

	DIAGNOSTIC GRADATION	1		2		3		4		TOTAL
THERAPEUTIC GRADATION		REAL	EXP	REAL	EXP	REAL	EXP	REAL	EXP	
a		88	60	26	44	9	11	3	10	126
b		27	48	53	36	12	9	10	8	102
c		6	10	9	7	1	2	5	2	21
d		-	2	1	1	1	-	2	-	4
TOTAL		121		89		23		20		253

The diagnostic consensus and the therapeutic conformity seem to coincide. The less diagnostic consensus tthe less therapeutic conformity is to be found. The question of the consistency between these two variables is answered by the calculation of the expected values for each box of this cross-table. The deviations to the real figures indicate a considerable consistency between the variables.

REAL = real values as were found in the investigation.

* EXP = values as expected from calculation (explanation in text).

The figures in the boxes represent the number of records to which both a diagnostic gradation and a therapeutic gradation pertains.To test the consistency between these two variables,we calculated for each cell the standardized predicted values from the totals of each class of gradation in case the values follow the relative composition of these totals.It is calculated as the total of a certain diagnostic class multiplied by the total of the pertaining therapeutic gradation class divided by the total number of records.E.g. for the upper left hand cell (1a) the calculation is 121*126 /253,and for 2b 89*102/253.

The most striking phenomenon in this table is, in my opinion, that the more uniform the diagnosis, the more consensus we can find on the part of the therapy. The highest gradations from both sides match for 72%, all other gradations are strikingly lower. The second stage of both gradations scores 60% and decreases relatively with (decreasing) diagnostic gradation. The lower gradation on the diagnostic side also scores remarkably low in the therapeutic part. Apparently, diagnostic and therapeutic uniformity and consensus parallel each other. This aspect deserves great attention. As we mentioned before, the action of management and treatment cannot be viewed as separate parts in the medical process. Both are parts of the same process, maybe the same strategy. This latter question cannot be answered from our material, because treatment recommendations were coupled by us to the diagnosis, not to a majority-diagnosis within the same patient case. Therefore, we can only present the results of the matching between the diagnostic gradations and the types of strategy. To allow comparisons we standardized the four groups of diagnostic gradation.

Table 70

DISTRIBUTION OF TYPES OF STRATEGY VERSUS DIAGNOSTIC GRADATION

STRATEGY	PATT.RECOGN.		HES.PAT.REC.		INDUCT.HEUR.		IND.ALGOR.		TOTAL
	ROW%	col%	ROW%	col%	ROW%	col%	ROW%	col%	
DIAGNOSTIC GRADATION									
1	10.7 *	27	8.2	22	68.0	23	13.2	38	100
2	11.3	28	14.7	39	62.7	22	11.3	33	100
3	13.0	32.5	4.4	12	82.6	29	-	-	100
4	5.0	12.5	10.0	27	75.0	26	10.0	29	100
TOTAL		100		100		100		100	

* the number of records is standardized for each of the variables.
From the relative figures of this table we observe a high score for the Inductive Algorithmic method in case of diagnostic gradation 1. Taking the sum of the diagnostic gradation 1 and 2 togetther again the Inducttive Algorithmic method scored highest, and lowest in the Inductive Heuristic. It seems that consensus of opinion about diagnoses among physicians employing the Algorithmic method is markedly greater.

For the group of diagnostic gradation 1 the largest proportion of consensus is found with the inductive-algorithmic strategy.

Because of its specification (the same disease,or one belonging to the same organ system) we are especially interested in the diagnostic gradation 1 and 2.From the relative figures of this table it is seen that the Inductive Algorithm method scores highest for diagnostic gradation as well as gradation 1

and 2 together.At a long distance it is followed by the Hesitating Pattern Recognition method.Remarkably,the Inductive Heuristic strategy is the type with the lowest score for the sum of diagnostic gradation 1 and 2.

We have tried to answer the consensus of opinion among physicians regarding the bigger items in medical nomenclature and classification, the names of diseases and illnesses. Our second question was directed toward the consensus at a lower level of sophistication: the consensus among physicians about the composing elements of these diseases and diagnoses, the symptoms.

To conclude this paragraph and this chapter we shall sketch few elements of this question.

After they had made a diagnosis we asked the participating physicians to delineate the composing elements of the diagnosis they made a moment ago, to denominate the colours of the picture they have in mind. They all pictured the particular disease as they saw it as characteristically in terms of a number of symptoms. This number counted between 4 and 14 symptoms; on the average 6 symptoms. Obviously, one can distinguish two types of consistency regarding these elements: the external and the internal consistency. The former one points to the agreement upon symptoms among physicians concerning the same disease. E.g. when 10 physicians are making the diagnosis myocardial infarction, about how many symptoms do they agree in making that particular diagnosis. The internal consistency refers to the matching of the - by the physician - given picture of the disease and the actual acquired symptoms during the work-up.

First, we shall deal with the external consistency. This attribute will be described in figures, in tables and in graphics.

Considering a particular diagnosis a - large - number of symptoms concerning this diagnosis have been named. Some of these symptoms are mentioned by all physicians who have diagnosed this particular disease (100% external consistency). Other symptoms are only mentioned by half of these physicians (50%) or a quarter (25%) and so on.

When an amount of symptoms is mentioned for a particular disease any of these symptoms has a corresponding percentage related to the number of physicians who had asked this symptom.The sum of these percentages divided by the total number of different symptoms constitute the MEAN EXTERNAL CONSISTENCY FACTOR.

The Mean External Consistency Factor can be defined as:the average with which any symptom within a cluster of symptoms,corresponding to a particular disease,is mentioned within a group of physicians who asked these symptoms.

In the tables the symptoms are denoted by a letter code. For the leading symptoms we shall present a decoding. The numbers refer to the times this particular symptom is mentioned within the group of physicians who diagnosed the same disease. The percentages in the third row express to be the external consistency for that particular symptom.

Table 71

EXTERNAL CONSISTENCY

DECOMPOSITION OF THE DIAGNOSES INTO THEIR SYMPTTOMS ; CONSENSUS OF OPINION ABOUT THESE SYMPTOMS AMONG THE PHYSICIANS

I

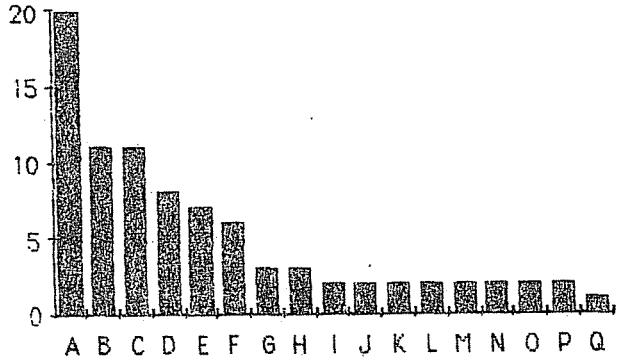
MYOCARDIAL INFARCTION

NUMBER OF PHYSICIANS
WITH THE SAME DIAGNOSIS: 20 TOTAL NUMBER OF SYMPTOMS: 94

SYMPTOMS	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
NUMBER	20	11	11	8	7	6	3	3	2	2	2	2	2	2	2	2	1 (11X)
%	100	55	55	40	35	30	15	15	10	10	10	10	10	10	10	10	

MEAN EXTERNAL CONSISTENCY FACTOR : 16.0

Leading symptoms:
A: pain
B: blood pressure
C: perspiration



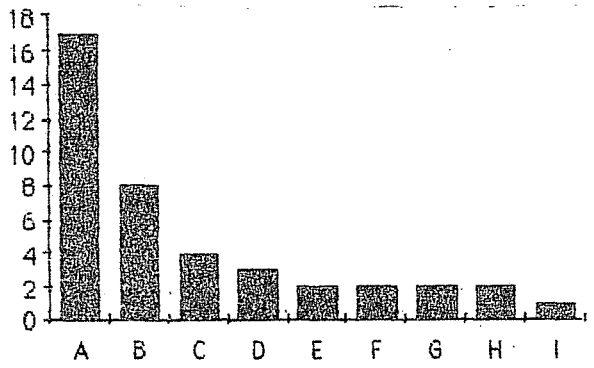
ISCHIALGIA

NUMBER OF PHYSICIANS
WITH THE SAME DIAGNOSIS: 17 TOTAL NUMBER OF SYMPTOMS: 56

SYMPTOMS	A	B	C	D	E	F	G	H	I
NUMBER	17	8	4	3	2	2	2	2	1 (16X)
%	100	47	24	18	12	12	12	12	-

MEAN EXTERNAL CONSISTENCY FACTOR: 9.9

Leading symptoms:
A: pain
B: Achilles tendon reflex
C: lower extremity sensory loss



III

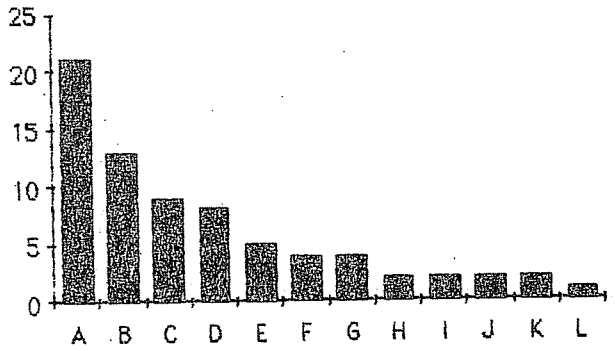
CHOLELITHIASIS (GALL-STONES)

NUMBER OF PHYSICIANS
WITH THE SAME DIAGNOSIS: 21 TOTAL NUMBER OF SYMPTOMS: 87

SYMPTOMS	A	B	C	D	E	F	G	H	I	J	K	L
NUMBER	21	13	9	8	5	4	4	2	2	2	2	1 (15X)
%	100	62	43	38	24	19	19	9	9	9	9	-

MEAN EXTERNAL CONSISTENCY FACTOR: 13.1

Leading symptoms:
A: pain
B: food intolerance
C: discoloration of stools



IV

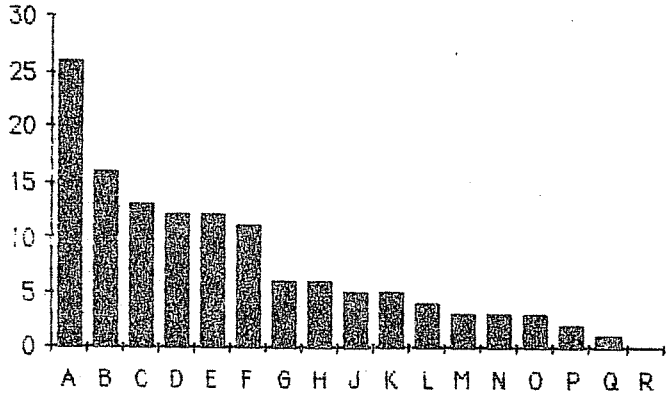
HYPERTHYREOIDISM

NUMBER OF PHYSICIANS
WITH THE SAME DIAGNOSIS: 27 TOTAL NUMBER OF SYMPTOMS: 155

SYMPTOMS	A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	Q	R
NUMBER	26	16	13	12	12	11	6	6	5	5	4	3	3	3	2	1	(25X)
%	96	59	48	44	44	41	22	22	18	18	15	11	11	11	7	-	

MEAN EXTERNAL CONSISTENCY FACTOR: 11.7

Leading symptoms:
A: weight loss
B: (normal) appetite
C: palpitations (heart)
D: perspiration
E: stool - frequency



V

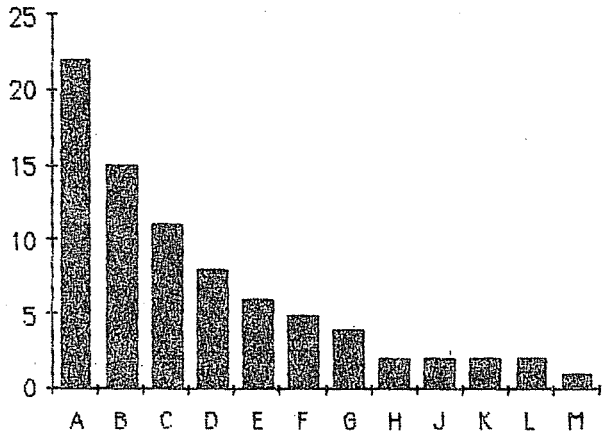
EXTRA - UTERINE PREGNANCY

NUMBER OF PHYSICIANS
WITH THE SAME DIAGNOSIS: 24 TOTAL NUMBER OF SYMPTOMS: 100

SYMPTOMS	A	B	C	D	E	F	G	H	J	K	L	M
NUMBER	22	15	11	8	6	5	4	2	2	2	2	1 (21X)
%	92	63	46	33	25	21	17	8	8	8	8	-

MEAN EXTERNAL CONSISTENCY FACTOR: 10.0

;eading symptoms
A: pain
B: menstrual cycle alteration
C: perspiration



VI

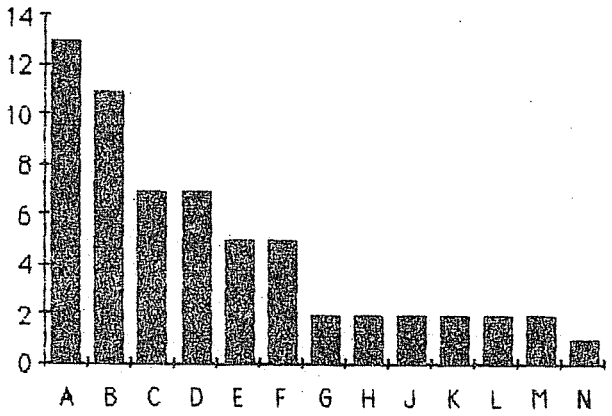
ASTHMATIC BRONCHITIS

NUMBER OF PHYSICIANS
WITH THE SAME DIAGNOSIS: 13 TOTAL NUMBER OF SYMPTOMS: 73

SYMPTOMS	A	B	C	D	E	F	G	H	J	K	L	M	N
NUMBER	13	11	7	7	5	5	2	2	2	2	2	2	1 (13X)
%	100	85	54	54	38	38	15	15	15	15	15	15	-

MEAN EXTERNAL CONSISTENCY FACTOR: 18.3

Leading symptoms:
A: dyspnea
B: cough
C: rales
D: sputum



VII

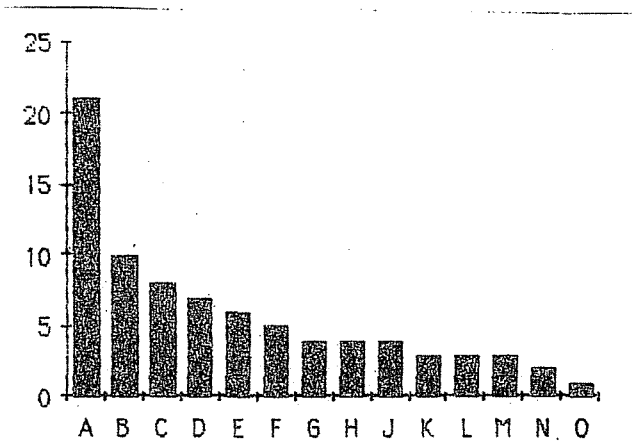
IRON - DEFICIENCY ANAEMIA

NUMBER OF PHYSICIANS
WITH THE SAME DIAGNOSIS: 13 TOTAL NUMBER OF SYMPTOMS: 73

SYMPTOMS	A	B	C	D	E	F	G	H	J	K	L	M	N	O
NUMBER	21	10	8	7	6	5	4	4	4	3	3	3	2	1 (19X)
%	100	48	38	33	29	24	19	19	19	14	14	14	5	-

MEAN EXTERNAL CONSISTENCY FACTOR: 11.75

Leading symptoms:
A: tiredness
B: dizziness
C: nervousness



The mean external consistency factor ranges between 9.9 and 18.3. This means that on the average physicians agree for approximately 10 - 15% upon a symptom within the symptom-cluster of a disease in case of a description of a "textbook picture" of this particular disease. For only one symptom per disease the participating physicians fully agree. (The agreement on symptoms with regard to the presented simulation is discussed in paragraph 2 of this chapter).

This striking phenomenon can be considered as a even more serious obstacle to the idenfication and classification of diseases than the preceding inconsistency of disease nomenclature. It means that there are grave limitations to the construction of data-bases, unless the items are coded in basic elements like symptoms. However, this requires a new registration system,which has to be developed and applied in close co-operation with physicians who can handle and use it.

The internal consistency factor compares the symptoms of the description of the disease by the physician with the actual acquisition of these symptoms in the problem-solving of that particular case.This factor focuses on two aspects:

1) the minimum number of symptoms proportional to the total number of symptoms named in the personal description,which is required to make a diagnosis;

2) the number of symptoms acquired which are irrelevant to the ultimate diagnosis,the redundancy.

We shall picture these aspects separately.

We may remember from our sketch of the 'paper patient' (chapter VII) that each symptom was accompanied by a rather fixed number of aspects. In medical

history the number of aspects for each symptom ranged between 20 and 25, and for physical examination between 8 and 15. For each symptom a number of questions can be asked. For some symptoms e.g. pain, practically all various aspects were asked by the joint participants, while for other symptoms one or two aspects sufficed.

It must be seen as an artefact therefore to depict the total number of questions in relation to redundancy. Another question is whether the physicians are fully aware of the deepening of their insight into a symptom by asking the several aspects adherent to this symptom. The next table shows the relation between questions and symptoms, for a number of diseases:

Table 72

THE RELATION OF QUESTIONS AND SYMPTOMS VISITED MORE THAN ONCE

DIAGNOSIS	TOTAL NUMBER of SYMPTOMS mean values*	TOTAL NUMBER of QUESTIONS mean values#	NUMBER of QUESTIONS per SYMPTOM mean values
Myocard.infarct.	12.2	24.5	1.8
Ischialgia	12.0	26.1	2.1
Cholelithiasis	13.0	23.7	1.9
Hyperthyreoidism	20.8	33.2	1.6
Asthm. Bronchitis	15.4	25.0	1.6
Extra-ut.pregnancy	19.6	34.2	1.7
Iron-def.Anaemia	21.0	35.0	1.7
Salpingitis	11.0	20.5	1.9
Colitis	37.5	58.5	1.6

* per physician
# in relation to the symptoms mentioned in the first column.

The table indicates the interviewing of a symptom and its substructure. On the average a symptom is visited twice during a work-up. Significant differences between the cases for this "visiting effect" could not be found.

These results confirm that each symptom is touched upon more than once by the questioning physician. Differences between general internists and family physicians were not found. On the average two questions are used to acquire insight in the meaning of a particular symptom. In our opinion, this figure is rather low and could be improved.

Relating the total number of symptoms to symptoms NOT depicted into the diagnostic sketch led to the redundancy factor. As was argued before, we have to realize that redundancy can only be established in hindsight. It does not and it cannot judge the quality of the problem solving or decision making process. It just pictures the groping and the searching of a physician in order to come to a conclusion. The next table gives the results for a number of diseases.

Table 73

REDUNDANCY OF SYMPTOMS

DIAGNOSIS	TOTAL NUMBER of SYMPTOMS mean values*	TOTAL NUMBER of QUESTIONS mean values#	REDUNDANCY FACTOR %
Myocard.Infarct.	12.2	9.4	77
Ischialgia	12.0	10.1	84
Cholelithiasis	13.0	10.8	83
Hyperthyreoidism	20.8	16.9	81
Asthm.Bronchitis	15.4	12.4	80
Extra-ut.pregnancy	19.6	16.5	84
Iron-def.anaemia	21.0	17.8	85
Salpingitis	11.0	8.5	77
Colitis	37.5	34.5	92

* per physician
\# in relation to the symptoms mentioned in the first column

Symptoms not related to the diagnosis are assumed to be redundant. This type of redundancy constitutes approximately 80% of the total number of collected symptoms. This type of redundancy can only be established in hindsight.

Approx. 80% of the collected symptoms are redundant in relation to the ultimate diagnosis. It also means that 80% of the time spent in decision-making, the physician follows the wrong tracks. In the light of this phenomenon the attempts of medical decision making to formalize the process and making it more efficient, seem very convincing. However, not only is this knowledge hindsight knowledge, but the methods of clinical decision making lack, in our opinion, also several elements of reality. Nevertheless, any strategy which could effectuate a more efficient way towards a decision should be welcomed.

This is especially true when we realize that the minimum number of symptoms regarded by the participating physicians to be necessary to make a diagnosis, is approx. 4.5. This means that only a restricted number of symptoms suffices to come to a diagnosis. This restricted number proportional to the number of symptoms depicted in the diagnostic sketch is the internal consistency factor. In our last table the results are shown, again, for the diseases mentioned in the preceding tables.

Table 74

INTERNAL CONSISTENCY

"HOW CONSISTENT IS THE PHYSICIAN'S PICTURE ?"

DIAGNOSIS	NUMBER of DEPICTED SYMPTOMS mean values	NUMBER of SYMPTOMS COLLECTED mean values	INTERNAL CONSISTENCY FACTOR %
Myocard.Infarct.	4.7	2.8	60
Ischialgia	3.3	1.9	58
Cholelithiasis	4.1	2.7	66
Hyperthyreoidism	5.7	3.9	68
Asthm.Bronchitis	5.2	3.0	58
Extra-Ut.Pregn.	4.2	3.1	74
Iron-def.Anaemia	4.6	3.2	70
Salpingitis	4.0	2.5	63
Colitis	5.0	3.0	60

The picture of the diagnosis as painted by the physicians in symptoms as colours consists on the average of 4.5 symptoms. The average number of collected equal symptoms is 2.9. The mean Internal Consistency Factor across these 9 different diseases is 64 %.

The physician's picture of the diagnosis consists on the average of 4 - 5 symptoms. In his workup he gathers 2 - 3 of these symptoms. Apparently, two-thirds of the (diagnostic cluster of) symptoms is necessary to make the doctor decide upon a particular diagnosis. The Internal Consistency Factor delineates the discrepancy.

Presumably, this factor is empirically founded, and based on the - often extreme - variability known to the physicians to exist in disease presentation. Maybe it is this variation of disease presentation together with the capriciousness of nature which leads physicians to the use of inductive strategies and a personal based medical knowledge.

REFERENCES CHAPTER VIII

1. Elstein A.S., Shulman L.S., Sprafka S., Medical Problem Solving, Harvard University Press. Cambridge (Mass). 1978.
2. McGachie W.C., Medical problem solving a reanalysis. J. Med. Educ., 1980, 55, 912 - 921.
3. Bolinger R.E., Ahlers P.,: The science of "pattern recognition" JAMA, 1975, 233, 1289 - 1290.
4. Feinstein A.R., Clinical judgment. Baltimore, Williams & Wilkins, 1967.
5. Knill-Jones R.P., The diagnosis of jaundice by the computation of probabilities J Roy. Coll. Phycns. Lond., 1975, 205 - 210.
6. Lindley D.V., Probability and medical diagnosis. J. Roy. Coll. Phycns. Lond. 1975, 9, 197 - 204.
7. Pipberger H.V., Klingeman J.D., Cosma J.,: Computer evaluation of statistical properties of clinical information in the differential diagnosis of chest pain. Meth. Inf. Med., 1968, 7, 79 - 92.
8. Gill P.W., Leaper D.J., Guillou P.J., Staniland J.R., Horrocks J.C., Dombal de F.T., Observer variation in clinical diagnosis. A computer aided assessment of its magnitude and importance in 552 patients with abdominal pain. Meth. Inf. Med., 1973, 12, 108 - 113.
9. Doroszewski J., Unity and Diversity of the medical action. A review of its components and their interconnections. Metamedicine. 1981, 2, 155 - 68.
10. Miller G.A., The magical number seven, plus or minus two: some limits on our capacity for processing information. Psych. Review, 1956, 63, 81 97.
11. Biorck G., The essence of the clinician's art. Acta Med. Scand. 1977, 201, 145 - 147.
12. Gale J., Marsden P., Clinical problem solving: the beginning of the process. Med. Educ., 1982, 166, 22 - 26.
13. Dombal de F.T., Medical diagnosis from a clinician's point of view. Meth. Inf. Med. 1978, 17, 28 - 35.
14. Smith D.H., McWhinney I.R., Comparison of the diagnostic methods of family physicians and internists. J. Med. Educ., 1975, 50, 264 - 269.
15. Gerrritsma J.G.M., Smal J.A.,: De werkwijze van huisarts en internist. Een vergelijkend onderzoek m.b.v. een interactieve patienten simulatie. Utrecht, W.U. Bunge, 1982.
16. Raynes N., A preliminary study of search procedures and patient management techniques in general practice. J. Roy. Coll. Gen. Pract. 1980, 30, 166 - 172.
17. Hull F.M., Diagnostic Pathways in General Practice, J Roy, Coll. Gen. Pract., 1972, 22, 241 - 257.
18. Slovic P., Fischhoff B., Lichtenstein S., Behavioral decision theory. Ann. Rev. Psych., 1977, 28, 1 - 28.
19. McWhinney I.R., Problem-solving and decision-making in Family Practice. Can. Fam. Physician, 1979, 25, 1473 - 1477.
20. Cromble D.L, Pinsert R.J.F.H., : The nature of information used in making clinical decisions in general practice. J. Roy. Coll. Gen. Pract., 1976, 26, 502 - 506.
21. Morrel D.C, Symptom interpretation in general practice. J. Roy. Coll. Gen. Pract., 1972, 22, 297 - 309.
22. Morrel D.C., Expressions of morbidity in general practice. Br. Med. J., 1971, 2, 454 - 458.
23. Tversky A., Kahnemann D., Judgement under uncertainty: Heuristics and biases. Biases in judgement reveal some heuristics of thinking under uncertainty. Science, 1974, 185, 1124 - 1131.
24. Gustafson D.H., Kestly J.J., Greist J.H., Jensen N.M., Initial

evaluation of a subjective Bayesian diagnostic system. Fall, 1971, 204 - 213.
25. Croft D.J.,: Is computerized diagnosis possible? Comp. and Biomed. Research, 1972, 5, 351 - 367.
26. Pauker S.G., Kassirer J.P., : The threshold approach to clinical decision making. N. Eng. J. Med., 1980, 302, 1109 - 1113.
27. Greenes R.A., McNeil B.J.,: Modelling of diagnostic test order decisions: a computer technique for prospectively aiding the clinician. Proceedings Medinfo 1980, Tokyo, 815 - 818.
28. Goldberg L.R., : Man versus model of man: A rational, plus some evidence, for a method of improving on clinical inferences. Psych. Bull, 1970. 73, 422 - 432.
29. Brehmer B.,: In one word: not from experience. Acta Psychologica, 1980, 45, 223 - 241.
30. Goldberg L.R.,: Simple Models or simple processes? American Psychologist, 1968, 23, 483 - 496.

CHAPTER IX

CONCLUSIONS, CONSEQUENCES, PROSPECTS

We return to one of our first questions: "Do physicians know what they are doing?" According to McLuhan thinking is substituted by coupling ideas to ideas, concepts to concepts, patterns to patterns. In his vision, thinking was prostituted by easy message wrapped in tinsel. In this version the answer to the aforementioned question must be negative. For thinking is often assumed to be logical thinking, an explicit process of reasoning which can be traced in reverse order. It is this - so-called scientific - concept that largely determines the way people view at medical problem solving processes. It led to emotional discussions and arguments between science-oriented individuals and people considering medicine as an art It opposed the logical to the illogical. It created two worlds in the one people say "we do not understand you", and in the other, people answer "we do not want you to understand".

The two worlds may be considered as east and west, and 'never the twain shall meet'.

But these attitudes referring to McLuhan's opinion about coupling concepts to concepts, disregard objective thinking and the possibilities to understand the actual processes. According to Braithwaite (1) "the function of a science (..) is to establish general laws covering the behaviour of empirical events, or objects with which the science in question is concerned". This characterization emphasizes its concern with empirical phenomena, together with its function in expanding knowledge of the laws concerning these phenomena (2). This conception includes the study and knowledge of disease processes and the methodology to diagnose, treat and prevent these diseases. Trying to understand the actual processes as they are taking place in day-to-day medical practice was the main purpose of this study.

Avoiding the two-world scenery we took a different position. We returned to an old, almost neglected domain of scientific inference, the inductive way of reasoning. Although the disadvantages and drawbacks of this method were broadly advocated by philosophers like Popper, (3), Braithwaite (1) and Medawar(4),its existence cannot be negated. It is practiced as an ordinary way of reasoning in daily life. People regularly infer general statements from particular elements or single observations. They often reduce complicated questions to comprehensible sizes and then jump to conclusions. They often make general conclusions from special memories or specific experiences and tell them as precious advice to their friends and relatives. They estimate chances from single perceptions and when the outcome flaws they invent new and different arguments to support a general statement sometimes opposite the former one. Inductive reasoning can be observed every day in daily life, on television, in newspapers, in politics etc. We are so accustomed to this type of reasoning that it passes without noticing. It can be viewed as an innate trait of the human being.

But why is inductive reasoning practiced on such a world wide scale?

In our opinion, because inductive inference (and judgment) bears the conviction of prediction and truth. To generalize the particular it is believed that this particular event will present itself in future for many more times and in approximately equal shape. We believe this prediction to be true because we believe the inference and the evidence upon which the judgment rests to be true.

The evidence for the inductive statement stems from successful past applications of induction. The theory of induction holds that we experience many things of a similar kind. It creates a "habit of mind" which makes us believe that we will see more things of this kind. Then, when we have seen five white swans, we believe the sixth to be white too. This means that our belief in the inductive reasoning is strengthened when we see many similar things. But how do we know that the things that we encounter are really similar and instances of

the same phenomenon? If we still continue to feel that there is some consistent relation between our beliefs and the factual evidence presented to us, we must regard this reasoning (with Hume) as a mere habit without acknowledging any justification of the convictions expressed by the habit.

The 'real world of experience' becomes 'our world'; genuine statements become subjective statements; 'facts' are only perceived by one's subjective interpretation. When we collect data, we do so with an implicit or explicit question in mind. Popper makes this point for his pupils: "My experiment consists in asking you to observe, here and now. I hope you are all co-operating and observing! However, I fear that at least some of you, instead of observing, will feel a strong urge to ask: "What do you want me to observe?" (5).

In inductive reasoning not the evidence is compared to the hypotheses, but hypotheses are tested by other hypotheses, weighed against each other on basis of subjective (probability) estimates. This way of reasoning treats the degree of probability of a certain statement (hypothesis, conclusion from the given evidence) as a measure of the feelings of certainty or uncertainty, of beliefs or doubts, which may be aroused in us by certain assertions or conjectures. Hypothesis testing in inductive reasoning is comparing the sequence of hypotheses for its most believable, most convincing one.

It is true that one can obtain classes of hypotheses, diagnoses, individuals in this way, but these classes will still be individual concepts - concepts defined with the help of personal oriented names. Individual, because the validity of the inductive inference is based upon the proportion of former experienced inferences which led to conclusions which were true to the subject. He has merely drawn a portrait of his conceptually prefigured conclusion (Hadamard, in 6)

A diagnosis can be regarded as a very special personal statement bearing in it a personal opinion and prediction. Questioning this statement is questioning the person. The physician has committed himself to his judgment. The reflecting physician is then caught in an insoluble conflict between a demand for an impersonality which would discredit all the commitments (to the judgment) and an urge to make up his mind which drives him to recommit himself (6).

Against this background we have to face the results of this study. Not the conclusion (inductive strategy) but its consequences give rise to much concern. It questions the generality of medical science. It questions the generality of medical knowledge and its methods; it questions its classification and its concepts. It demarcates, in Popperian sense the physics from the metaphysics, science from art. In our study several examples of this phenomenon have been given.

Although not 'strictly valid' the investigation attained some degree of 'reliability' or of 'probability' by its participants' judgments. In this sense the conclusion can only be an inductive one: a theory temporarily accepted until a better, more valid one comes along. It is the fate of a descriptive study: it observes predicted features and elements; it proves nothing.The study has created a theory of problem solving to which Newell et al. (7) formulated a number of conditions:

1.) it should predict the performance of a problem solver handling specific tasks;
2.) it should explain how human problem solving takes place;
3.) it should indicate what processes are used;
4.) it should indicate what mechanisms perform these processes;
5.) it should predict the incidental phenomena that accompany problem solving, and the relation of these to the problem solving task;
6.) it should show how changes in the attendant conditions - both changes "inside" the problem solver and changes in the task confronting him - alter the problem-solving behaviour;
7.) it should explain how specific and general problem-solving skills are learned, and what it is that the problem-solver "got" when he has learned them.

Briefly, it can be stated that validity may be assumed when the model predicts the operations actually employed in the task performance.

This investigation produced failure and success. Failure in that we could not trace a deductive strategy, as it was designed in our model. In the first phase of classification, we selected 10 records with a rather abundant primary data collection, (10 questions or more ahead of the generation of the first hypothesis). However, they nearly all failed to fulfil the second criterion of the progression down the levels. In only one record a trace of such a progression could be signalized. But this last record failed in the criterion of rejection: two hypotheses remained apart from the diagnosis. Thoroughly looking into the record and applying the HRS/ HPS criteria, we learned that this case fitted into the Inductive-Heuristic strategy.

We seriously regret this absence of a deductive strategy. Although Harre (8) supposes the hypothetico-deductive method being actually an inductive mode, we think it possible a deductive strategy could exist. The failure to detect any deductive strategy in our material can be attributed to several causes.

1.) the method of investigation was not suitable to detect the deductive strategy;

2.) the number of participants or the composition (mainly family physicians, a small number of general internist) of the group have made the detection of a deductive strategy less probable;

3.) physicians do not practice in a deductive way;

4.) medical science is not structured in a way which permits a deductive way of reasoning.

From our material and literature we shall conclude that the latter considerations are the more valid ones.

The absence of a deductive strategy in our material deprives us of the special and necessary insight into this method on the basis of empirical evidence. Comparisons between the deductive and inductive inference processes could have provided several questions and solutions thusfar unknown or unthought of.

We gained success sofar, that we ascertained a mode of problem solving consistent across physicians, problems and situations. We realize that generalizations from this study may be hazardous.However,the methods and the procedures physicians employ in solving patients' problems are so strikingly similar and uniform across the various physicians and problems,that as a result we cannot see why the 68 physicians who participated in this study would essentially differ from their colleagues elsewhere.As we argued before (chapter IV) a notable characteristic of the medical profession is that its practitioners almost perfectly recognize each other's troubles,behaviour,and working patterns.This study elucidated these working patterns as explicit methods as they are employed in day-to-day practice.

The validity of the strategy discriminating criteria is in the predictive force of the model, and the experience of reality of the simulation by the participants. (chapter VII).

Recapitulating we found::

1. - a cycling process as could be pictured by (re)naming the adjacent symptoms towards the hypothesis to which they belonged (HRS, HPS).The coding system for symptoms and hypotheses enabled us to elicit the special circular arrangement of the hypotheses (see chapter VII).

This procedure follows from the basic principle of inductive inference: evidence (or premise) stems from hypothesis rather than from the presented facts. As argued before, inductive reasoning is a weighing of induced hypotheses rather than a processing procedure of data.

As a consequence we found:

2. - the creation of a "hypotheses-base" which is visited and revisited during the course of the work-up. The subjective (probability) estimates attached to each hypothesis eventually lead to the weighing procedure;

3. - the 'inductive' hypothesis cannot be viewed as a member of a hierarchical

system but must be conceived as a pattern which has been evoked by a more or less striking resemblance with a memorized pattern. It does not follow from the evidence, nor does the evidence cover all attributes of the pattern. The 'inductive' hypothesis cannot be generalized across physicians and patient problems because of its personal character and commitment;

4. - As a consequence of the former item the levels of the generated hypotheses do not follow a particular pattern. Every sequence of hypotheses (hypothesis base) which follows a more or less hierarchical nesting is quite accidental;

5. - For reasons of personal commitment rejection of hypotheses plays only a minor role. Approximately 20% of the hypotheses in the relevant strategies i.e. excluding the pattern-recognition strategies, were rejected. Rejection of hypotheses can be assumed to be quite accidental;

6. - We also found a conspicuous variability of hypotheses across the various organ systems. In our opinion, it pictures a search behaviour in an uncertain task environment.

On the other hand, the fuzziness of medical symptomatology forces the scrupulous practitioner to visit various organ systems, especially the ones adjacent to the one in consideration. It raises the question whether medical science itself forces its practitioners to particular methods and behaviour;

7. - In a sequence of hypotheses, the diagnosis represents neither the definitive one nor the one with the lowest-level refinement.

Besides, diagnoses seem to be very personal statements uncomparable to each other, even in similar cases.

This finding questions medical taxonomy when based on the classification of physician's diagnoses and/or other reference features (e.g. reasons for encounter).

8.) We found corroboration for the hypothesis concerning the endpoint of the diagnostic process. the endpoint being marked by a maximum subjective (posterior) probability and a minimum of (subjectively experienced) uncertainty. Not only the weighing of the hypotheses plays a role, but the personal confidence of the physician as well.

(We found some indications that in a minority of cases declining certainty estimates may contribute to precocious action, mostly the referring of the patient.)

The confidence in the conclusion of the diagnostic process is reflected in the - practical - absence of any testing of the diagnosis. Even in the reviewing and reconsidering of the process by videotape after 3 - 4 days the confidence in one's own judgment was unshakable. Nobody changed or even questioned his judgment.

9.) The personal and subjective characteristics of the diagnosis and the treatment led to the confusing arrangement of the attributes. Supposing the majority of equal diagnoses (= equal names of diseases) to be right less than 50% of the diagnoses were accurate. (lowest 27%, highest 84% of the total number of diagnoses mentioned per case.

10.) As the model predicted, action and the decision to treatment automatically follow the conclusion of the diagnostic process. This important decision was carried out within a minimum of time, between 1/2 and 2 minutes later. It gives the impression that the decision about management and treatment is concealed in the diagnostic reasoning process itself; it just has to be triggered to pop out.

11.) The physicians' behaviour does not reflect the continuing patient care as advocated in the domain of family practice. We found that continuation of care by the family physician was mainly restricted to cases which he himself attended to. In cases of referral to colleagues (consultant specialists) the responsibility of the

care was largely transmitted to the other person. It may partly reflect the personal commitment of the physician. Each health care attendant claims to have his own personal commitment and ethics. These moral values are not transferable, and cannot be questioned. This attitude may hamper the effectivity of medical audit and other forms of interphysician testing.

12.) The effectivity of the physician's data collection is often questioned in many ways. As data acquisition in the inductive strategy is conceived as a search for positive symptoms in order to confirm a hypothesis, the rather steep decline of the 'ratio-effectiveness' (for description see chapter VIII) gives some support to this thought. However, one has to make up one's mind whether to choose for efficiency or completeness. In our material these matters seem rather at variance. The contrast may picture the different task environments between the two groups: family physicians and general internists. A family physician has to 'see' a large number of patients with a large variety of illnesses and complaints. His main task can be pictured as a primary selection among the patients visiting him. His problem solving process, therefore, can more or less be characterized by skimming the features and problems the patient presents to him. An internist is urged to take a thorough look at the diseased state of the patient. He is called to a more precise task performance for a smaller number of people. He strives for perfection.

In other words, a general statement about physicians' effectivity in data acquisition cannot be given. It has to be specified for each group of physicians according to their speciality.

The closer look at the details, like in the preceding example, creates a distinction between the groups, and, while progressing downwards, eventually between individuals. It was our purpose to create <u>that</u> particular distance which would allow us to observe the general without blurring the details. To this end we originated three subtypes within the inductive strategy. This subdivision allowed to observe and study a number of specified items within a particular group or between the groups of participants. It might be clear that we did not reach beyond the dimension of a group of participants. Statements about individuals cannot be made. This subdivision consisted of three types:

1) <u>Pattern-Recognition</u>

It is the strategy of "the happy thought". It is not one among a great many ideas to be pondered upon at leisure, but one which carries conviction from the start. This strategy can be viewed as the most fundamental one among the inductive strategies. It underlines Hume's theory of induction that we believe to experience many things of a similar kind. It is the belief in orderliness, the belief that things will happen in the same way in the future. "When we have seen five white swans, we believe all swans to be white". It is the "habit of mind" which makes us believe that we will see more things of this kind, without ackowledging any justification of the convictions expressed by this habit.

Although Pattern-Recognition as a way of observing and thinking is a major strategy in the zoological and human world, its functioning is hardly understood. It operates through recognition of superficial resemblances. It operates in flashes, enabling animals to survive by fast recognition of their enemies.

Although patients are, generally, not their physicians' enemies, the same strategy enables physicians to recognize and to react fast. It enables family physicians to perform their main task: making quick selections among a large population. The results of this study - approximately 100% of the pattern recognition strategy was employed by family physicians - confirm this thesis. In modeling this strategy, we originally delineated the process as a quick hypothesis generation after the presentation of the complaint (2 items). In our material, however, we found a group of 26 family physicians cautiously asking a varying number of questions ahead of the first and only hypothesis. They exhibited a slightly different behaviour: a hesitating start and a - very - confident conclusion. We

named this group after the hesitating start. We do not know whether it is either a product of the simulation, the environment, or in general, the research, in which case it has to be referred to as an artefact; or a mode of action actually to be found in practice. Our opinion tends to the former explanation.

2.) Inductive Algorithmic method . It consists of the induction of a sequence of "patterns" (hypotheses) through the application of systematic methods (algorithms). Systematic methods are employed in many cases and in many problems. Usually they are far too laborious to be carried out in practice. It cannot be a coincidence that this strategy falls largely within the group of internists.

Regrettably, we must ascertain that 'algorithmic' or 'systematic' mainly refers to a personal conception of these terms. Any generalization of this 'systematic method' appears to be illusory.

3.) Inductive Heuristic method .

"The difference between the two kinds of problem solving, the systematic and the heuristic, reappears in the fact, that while a systematic operation is a wholly deliberate act, a heuristic process is a combination of active and passive stages" (6).

Heuristic can be described as a trial and error search program based on series of successful (past) random trials.

The difference in behaviour between these two latter strategies can be elucidated by two features: the pace of questioning and the hypothesis-relevancy of the questions.

As mentioned the I.-A. mode is marked by its more or less fixed runs or questions, only scarcely influenced by the patient's answers. Because of its rather traditional sequence the questions can be asked at a high pace. The purpose of this type of questioning is the screening of the health status of the patient regardless of the particular hypotheses for this case.

In the heuristic fashion the questioning is more or less linked with the answers of the patient. We observed an acceleration of the questioning when answers were affirmative (completion of the pattern in mind), and a slackening when answers were repeatedly negative. In real encounters we could observe that vague questions led to uncertain answers, which in turn led to fuzzy questions from the part of the physician, and so on. A deadlock could be overcome by a prescription.

In our material we found a difference of 1.3 questions per minute: average figures for family physicians were 2.9 questions/minute, for general internists 4.2 questions/minute.

The Non-Hypothesis-Related-symptoms (NHRS) took up 41% in the I-A strategy, in the I-H strategy 26 %.

The prediction for both 'strategy-behaviours' has been confirmed by the figures.

After this brief sketch of some general conclusions, we now again face the question of the validity and reliability of the model. We may, therefore, return to the Newell et al. (7) condition scheme (see page 2 of this chapter).

We shall discuss these conditions in the order as they have been posed.

1.) The model should predict the performance of a problem solver handling specific tasks. As we have argued before, it is our opinion that the detailed approach of the medical problem solving blurs the recognition of the general traits and characteristics of this process. Every culture, every nation has its particular characteristics recognizable and known to most of us. However, this does not mean that each individual as a part of this culture or nation, therefore, represents these characteristics. The French, the English, or the German national characteristics are distinguishable from each other, but it does not mean that while observing Monsieur Dupont, or Mr. Brown, or Herr Mueller, we are able to recognise these particular characteristics. In our situation, the uniqueness of the medical problem solver as a person prevents him from observing the elements he has in common with his colleagues. We focused on groups of physicians and found a general strategy (inductive

inference). Both groups, family physicians and general internists could be typified according to the subtypes within the inductive strategy. In the preceding section we gave various examples of differences in handling specific tasks regarding the participants.

2.) The model should explain how human problem solving takes place.The problem solving process mainly takes place in an inductive way of reasoning. It gives the process that peculiar personal touch which is often referred to as "the physician's art", or "the clinical view", or the "flair clinique". It is based upon (a sequence of) patterns as analogies of memorizations, as the expression of the physician's recollection of similar combinations of features or circumstances. The details of this type of problem solving have been sketched in previous sections and chapters.

Unfortunately, we were unable to demonstrate a (hypothetico-) deductive way of reasoning. We sincerely deplore this circumstance, because the controversy between the opinions thus still remains. For we cannot prove whether the research method created the inductive method and/or prohibited a deductive method. Our material provides a reasonable belief, nothing more, nothing less.

3.) The model should indicate what processes are used. A number of typical elements of the inductive inference has been explained. To repeat briefly:

a.) a circular way of reasoning, as descended from the hypothesis related-symptoms (see chapters VI and VIII);

b.) construction of a hypothesis-base;

c.) hypotheses are tested against each other;

d.) hypotheses are weighed according to subjectively estimated probabilities and the physician's feeling of (un)certainty;

e.) the acquired evidence need not cover the diagnosis (viewed as a set of symptoms), or vice versa;

f.) the therapeutic management plan is already implicitly incorporated in the diagnostic process.

4. The model should indicate which mechanisms perform these processes.

After the problem has been stated and an internal representation has taken place (the physician has translated the patient's problem into his own words) a suitable method is selected and applied to proceed towards the solution of the problem.

"A method, then, is tied closely to its associated problem formulation. The method resides in the problem solver prior to its evocation to deal with any individual task. Thus the method must be fashioned without specific knowledge of the particular problem situations it will be called upon to handle. Consequently, mechanisms must exist in the problem solver for bringing the problem solving method into effective correspondence with each individual problem situation, so that the method can actually discover something about the situation and act on it" (9).

Because of the character of this study, we can only guess for the mental mechanisms which may take place.

We have already described the cycling process as based on a series of patterns elicited by only a small number of 'trigger-symptoms'. When these "trigger-symptoms" meet tentacles (daemons) from some frames in long-term memory matching can take place, which causes the full frame (a complete disease description) to enter into short-term memory (see also Pauker et al.(10).

The first pattern which rises to consciousness is apparently based on only a few (mainly 2) symptoms (the presented complaint). However, to make up one's mind in order to generate an alternative pattern (and/or to overthrow the previous one) asks for a more substantial number of symptoms. We found a more or less fixed number of questions (16 plus or minus 1) in each cycle, regardless of the presented patient case. This finding applies for all six full cycles. Its meaning cannot be ascertained within the context of this study. It would be intriguing to link this finding to Miller's theory on Channel Capacity and "chunking" (11), combining bits of information to chunks, or as a variation to this theme, symptoms (aspects) into patterns.

The quality and the difficulty of the presented problem evidently influence the choice for one of the three (four) subtypes of the inductive strategy. Apparently, this feature does not fundamentally alter the strategy, i.e. from inductive to deductive.

We predicted the 'even-numbered' patient-scenarios as the more difficult ones (and consequently the 'odd-numbered as the easy ones).This prediction could be corroborated for most of the variables as we have defined for the medical process.However, complete problem-dependency could not be assigned to the various case work-ups. Evidently,the procedure is more or less superimposed by personal characteristics of the physician. We could observe a personal commitment to the process and its outcome as can be typified by the absence of any reconsideration or readjustment of the work-up and its diagnosis during the revision of the videotape.

The real mental mechanisms still remain obscure. The study may only be of help to lift a corner of the veil.

5) It should predict the incidental phenomena that accompany problem solving,and the relation of these to the problem solving task.

A number of the incidental phenomena accompanying the problem soving has been described as being part of inductivism. We shall repeat briefly some of these phenomena.

a.) the heuristic search for hypotheses is typified by the high degree of variability of these hypotheses over the levels;

b.) a detailed description of the diseases as diagnosed by the physicians, and, consequently, the possibility for a reliable and valid classification of diseases,is contrasted by the level variation of the diagnoses. (see table 13, chapter VIII);

c.) there is a lack of agreement among the participants for the symptoms to base a diagnosis upon.

Generally, only half of them agreed on one or two symptoms. The basis for a diagnosis, therefore, lacks consistency, as viewed from a descriptive standpoint;

d.) the large variation of subjective estimates for probabilities, prevalences and incidences lends a very personal character to the problem solving;

e) the variation among the composing elements of a disease description as the physicians had memorized (external consistency), places serious restrictions to diagnostic comparisons, medical audits, and the construction of medical data bases. This finding can be examplified by the variation of disease descriptions as can be found in most medical textbooks;

f) when the diagnosis (the set of composing elements constructing a disease description as memorized by the physician) is related to the acquired symptoms,these symptoms cover only two-thirds of the totally memorized set of symptoms.In our opinion,it indicates the physician's alertness for the variability in patients,in the presentation of complaints,in co-operation,in situations,reactions and so on.Because of its implication its extent is difficult to ascertain.

6) It should show how changes in the attendant conditions - both changes "inside" the problem solver and changes in the task confronting him - alter the problem solving behaviour.

As we did not change the attendant conditions,this requirement cannot formally be fulfilled.We can only elucidate this condition by some casual observations.

Some physicians started quite determinedly in a kind of pattern-recognition mode and ended up in a lengthy inductive-algorithmic type. Other physicians started with a hesitating pattern recognition method and ended up in an inductive-heuristic strategy.

Some physicians who were unacquainted with a particular patient presentation reacted by restriction of the interviewing and referred the patient

to a colleague-consultant. (One physician referred the patient immediately after the presentation of the complaint without asking a single question.) Other physicians reacted in an opposite direction in similar cases: they went on and on interviewing the patient desperately grasping for a foothold. But all of these "inside" or "outside" changes took place at an individualistic level and cannot be generalized. As argued before, too detailed an observation blurs the vision.

7.)It should explain how specific and general problem-solving skills are learned,and what it is that the problem solver "got" when he has learned them.

Thusfar, the student was supposed to learn the (hypothetico-)deductive method, the "universal method in problem-solving (12)". But when neither older or younger physicians, especially the ones who had recently left medical school, nor the teachers (a number of the family physicians-participants and most general internists are assigned to a formal teaching status) employed this "universal method", one starts to question the aforementioned assertion of student learning. Most student-teaching starts from "the presupposition that it is possible to come to a faultless diagnosis (according to the medical classification system) with the rules of logic when all diseases (of that classification) are known and the physician is able to collect all information from the sick patient" (13). But, as Gerritsma & Small continue, when medical training is attuned to only one technique of problem solving, one has to realize that this technique will be applied in practice by only a small proportion of the physicians".

Our findings question the presumption of the teaching of a (hypothetico) deductive mode of medical problem solving. We suggest that medical problem solving teaching, when taught, implicitly more than explicitly, transmits an inductive inference mode. The sooner this notion is acknowledged the sooner teaching programs can be reconstructed in an explicit way and aimed at methods which are really practiced in day-to-day health care service.

Most conditions of the Newell et all (7) axioms to model validity have been fulfilled. Our findings largely cover the predictions which can be made from the model. We shall conclude the inductive model to be a reliable and valid theory for studying physicians' problem solving methods and behaviour. Especially because we can mention the general acceptance of the procedure by the participants, having enjoyed this particular type of investigation.

On several occasions scattered through this book we mentioned a number of consequences resulting from inductivism. We shall now discuss in close-up and more coherently. The main consequences descend from the typical characteristics of inductivism.

I : it is based on personal - or in the words of Polanyi - tacit knowledge;
II: the process' steps are irreversable.

The main consequence of the personal character is that the knowledge and the acts can only partly be formalized. By reflecting on the way in which we are performing a personal act, we may seem to establish rules for our own guidance. But such formalization must remain within the framework of the personal judgment. Any attempt to generalize this formalization is likely to go too far. All attempts to formalize the process of inductive inference go astray precisely in this respect "(6). In this sense, medicine resembles art. The rules of art do not determine its practice but they function as maxims, which can serve as a guide only if they can be integrated into the practical knowledge of the art. Whether the rules are efficiently integrated can only be established by the outcome. Therefore, the judge's action is considered more authentic than what

he said what he is doing. The predominant emphasis physicians put on their diagnoses and treatment plans examplifies these thoughts.

Skills, like connoisseurship of art, can be communicated only by example, not by precept. To be trained as a medical diagnostician one must go a long way of experience under the guidance of a master. Each new patient case, each new example is assumed to augment the - personal - knowledge of the student. This is the usual process of trial and error by which we feel our way to success and may continue to improve on our success without specifically knowing how we do it, for we never meet the causes of our success as identifiable things which can be described in terms of classes of which such things are members. We are faced with the general principle by which our beliefs are anchored in ourselves. A rule, a pattern, a treatment suggestion is accepted if we believe it to be actually useful to our purpose. The act of personal knowing can sustain these relations only because the physician (or the student) believes that they are apposite: that he has not made them but discovered them. The discovery makes him believe that the element is a "real fact" existing outside himself. "Any presumed contact with reality inevitably claims universality" (6).

The actual facts, on the other hand, are accredited facts as seen within the personal commitment situation. People are unable to observe without prejudice as Popper demonstrated for his pupils.

But the subject believes the fact to be true. Bertrand Russel defines truth as a coincidence between one's subjective belief and the actual fact; yet it is impossible to say how the two could coincide.

The basic principles of personal knowing are based upon rules of rightness, which account only for the successful working of the procedure, but their failures remain entirely unexplained. In a system of beliefs, e.g. a hypotheses-base, we will admit positive but no negative, evidence. These thoughts may explain the observations of various investigators in this field who describe physicians as being fervently looking for positive evidence and ignoring or neglecting negative evidence.

Only if we believe that a solution exists we can passionately search for it and tempt ourselves to heuristic steps towards its discovery. When physicians stop to believe to be able to offer patients (a number of) solutions to their problems, they will lose ambition and, eventually, interest in their function. "Therefore, as the solution emerges in response to our search for something we believe to be there, it will always come to us with the conviction of its being true" (6).

We can see how the whole process of discovery and confirmation ultimately relies on our own accrediting of our own vision of reality. In this sense, our confidence is an act of social allegiance. All these tacit commitments appear self-satisfying, irreversible and hence unspecifiable. These thoughts will be examplified by some of the findings.

1.) Personal commitment.

The conception as laid down in the preceding lines was that physicians consider their diagnoses and treatment plans as facts, as truths. They feel a strong commitment to their predictions and stick to them even when proven untenable. Questioning the diagnosis or the treatment plan means questioning the person, questioning his integrity. Medical law suits are often considered by the physician as a grave, personal attack. The emotional repercussions are sometimes far beyond the seriousness of the case because of personal involvement.

A great deal of cognitive work and emotional investment required to reach a decision makes a person reluctant to admit to himself that his brainchild is defective. Commitment stabilizes a decision. One's commitment provides an anchor by which one's beliefs, attitudes, and behaviour are marshaled or organized, which influences the way one evaluates and responds to subsequent appeals or demands (14)

A second aspect of the stability of the physician's statements arises from

the automatic expansion of the circle in which an interpretative system operates. It readily supplies elaborations of the system which will cover almost any conceivable eventuality. Scientific theories which possess this self-expanding capacity are sometimes described as epicyclical (6). This epicyclical structure is a ready reserve of possible hypotheses available to explain any conceivable event. Anybody who has visited a physician knows the doctor's unfailing source of explanations to unexplained events.

The study enabled the participating physicians to comment, to reconsider, and to revise their processes. To this end every participant reviewed the videotape of these processes. Comments were audiotaped for elaboration afterwards. However, we found that none of the 68 physicians participating in the study, all together solving 253 cases, reconsidered or revised their original process although several suggestions from the part of the investigator have been made. Many of them lost interest in the review after one or two cases (out of four).

Personal commitment is also extended to the hypotheses. Hypotheses may be regarded as the physician's children of memory. They reflect or simulate past personal experiences, good or bad, but always precious to the owner. Rejection of hypotheses not only means a changing of a decisional direction, but also something like throwing away your brainchild.

On the average, less than 1 hypothesis per case has been rejected, representing approximately 20% of all hypotheses generated.

The more the physician is involved, the higher the degree of commitment, the less likely hypotheses are rejected. The more a father knows his children the more he loves them. This is examplified by the rapidly decreasing rate of rejection progressing down the cycles. In the first cycle about half of the hypotheses is rejected, while in the third cycle the percentage decreased to 11, and in the fourth cycle to only 0.5. As we may remember, this finding can only slightly be attributed to the levels of refinement.

Although the possibility was offered during the videoreviewing, rejection of hypotheses was not foremost in the physician's mind.

Effects and consequences of commitment have more extensively been discussed by Janis & Mann (14).

2.) External Inconsistency ,

stands for the lack of reliability across judges of the same data and the same occasion. The physicians (judges) were asked to break down the picture of the diagnosis, they just arrived at, into its composing elements (symptoms, signs, tests, etc.) These sets of components were compared to each other regarding the same diagnosis.

We found a large variation of diagnoses (see table 66 chapter VIII, diagnostic gradation). Among these diagnoses only a few were mentioned sufficient times to allow reliable comparisons. Seven of these homonymous diagnoses were found. In five of these seven diagnoses interjudge agreement could be reached for only one (!) symptom. The external consistency factor (for explanation see chapter VIII) ranged from 9.9 to 18.3 (maximum 100).

3.) Internal Inconsistency ,

can be defined as the reliability across data sources administered on the same occasion and interpreted by the same judge.

The breakdown of the "diagnostic picture" was compared with the symptoms actually collected during the work-up of the case. The values represent the intrajudge agreement and are given in percentages (compare with intra-observer agreement, chapter IV).

On average, two thirds (64%) of the collected symptoms covered the set of composing (diagnostic) symptoms. For various cases the percentages ranged between 58 to 75, and remained within rather stable limits.

This factor's relation to inductivism is in the personal character of each diagnostic process and in its irreversability. As the evidence does not cover the diagnosis, as we have already sketched as a characteristic of inductive inference, retracing of the process by means of its elements seems impossible.

Either the memorized pictures of diseases vary in time, or the physician's verbal report on the diagnosis is "telling more than he can know".

The lack of internal consistency is a serious drawback to comparability and judgment of problem solving processes.

4.) Estimations .

The personal character of the inductive inference process is also reflected by the subjective estimates of hypotheses-probabilities, prevalences, incidences and disease-consultation rates.

The estimates for diseases in a population (prevalence, incidence and disease-consultation) stand out by their excessive variability. Prevalence rates for iron-deficiency anaemia vary from practically zero to over 1000 cases per 10.000 inhabitants. Incidence rates for the same disease vary from 6 to 1000 (per 10.000). Only the disease consultation rate (the number of times the patient visits the doctor for the same disease) shows some stability : between 50 and 300. For myocardial infarction these figures range from 1 to 30 (prevalence) 10 to 110 (incidence), and 10 to 500 (disease-consultation). These figures are across the years of experience of the participating physicians. (see figures 3, 4, 5, chapter VIII).

The values for the probability estimates follow a similar variability-pattern across the physicians, for the prior probabilities as well as the posterior probabilities. (see figure 6, chapter VIII).

These - often extreme - variabilities may pose serious limitations to projected reliable medical decision making.

5.) Medical Decision Making

Inductivism hampers the application of medical decision making in the general practice. By its character medical decision making contrasts with the inductivistic problem solving behaviour. Medical decision making treats the medical problem solving process as a scientific system. Any scientific system could be treated as a simple branching system (1). This branching system together with a system of weighings at each node and at each outcome of a branch is the basis-scheme of decision theory, in which processing steps are reversible along the branches. The system of weighing is based on the prior probability, generally represented by the prevalence of the disease in question, and the adjustments of this probability in the light of the acquired evidence towards a certain outcome. With the help of a value given to a predicted or expected state-of-health, as outcome of the disease (expected utility), to each node in the branching system a value (probability) can be attached.

Within a system feedback plays a dominant role. As a matter of fact, negative feedback is the fundamental strategem of all control systems. However, negative feedback is avoided in inductive inference. As we have pictured before, neither are we able to construct reliable weighing systems, nor do we have the possibilities to retrace the process as it actually occurs in practice. The basic structure of decision making is a deductive one and does not coincide with the inductivistic processes in daily health care.

In our opinion, in medical decision making as a discipline a choice has to be made: it has to build its own medical theory and structure, or it has to subsume to the inductivistic way of reasoning by rebuilding its system in close co-operation with medical practitioners.

6.) Experience

There are two main elements that hamper learning from experience: the inductivistic character of the medical process and the medical philosophy. Medicine has two qualities distinctive from science: it deals with irreplaceable individuals and it has to act even when wise (or scientific) policy would be to restrain.

Nowadays physicians may not stand in the wings and wait and see what happens to the patient. As we have pictured in chapter IV, current medical science can bear no knowledge about the natural course of diseases. Consequently, the prediction for a disease can no longer be based on natural outcomes, but the outcomes are nearly always drug-induced. These outcomes are no longer based on pathophysiological consequences of diseases but are iatrogenic products. Needless to say that the latter ones bear a very personal character.

To realize what one is doing and to determine the effectivity and efficiency of one's process, the prediction to a case has to be compared with what actually happens. But when the prediction is based on varying and personal standards, how can anybody establish the effectivity and the quality of the actions taken? And even when the outcome satisfies patient and physician, this simple fact does not explain how one has arrived at this result.

In medicine acting means the exclusion of non-acting. The individualistic character of the medical process excludes by its very existence a number of options. When a physician prescribes drug X it excludes the prescription of drug Y for the same patient, otherwise than as a combination. The prescription also debars the non-prescription. The advantages or disadvantages of drug X for the patient cannot be compared with the advantages or disadvantages for drug Y for the same patient at the same moment. Other patients and other moments create different circumstances.

The observation of the results of the treatment poses another possibility

of error, an error extension of which we are unaware.

Sometimes a physician has to act in cases of life-threathening diseases or cases with the risk of serious disablement. His life-saving acts, however, cannot be proved to be the better ones because the acting is irreversible. The physician acts, and has to act, on basis of reasonable belief, from the inductivistic character of his problem solving as well as from vocational duty.

But, as we argued, inductive reasoning includes irreversibility of the process. It cannot explain the causes of failures, because the operational principle bears in it the conditions of the success. In this sense, the medical process is a one-way process. But when feedback mechanisms do not exist, how can a physician learn from these instances? How can he learn by trial and error when he cannot detect the causes of the errors?

To test the hypothesis that experience does not enhance medical capabilities, we matched a number of the variables with groups of physicians.

These groups were differentiated by the years in practice: a group of physicians 0 - 5 years in practice, one group from 6 - 20 years, and a last one with physicians who practiced for over 20 years. These distinctions were empirically based.

Matching with the relevancy factor, total time per case work-up, the number of hypotheses, posterior probabilities, certainty estimates, and therapeutic management revealed no relevant differences, thus confirming the hypothesis.

For the number of questions and prior probability estimates the middle group scored slightly different than the other groups. For these items experience seems beneficiary, but declines for the elder group. The middle group employed the pattern-recognition strategies in a significantly higher percentage. Probably, the confidence in their judgments scores higher for this group.

On the basis of available literature (e.g.21,69,100), the concepts of the inductive theory, and our findings, we conclude that learning from experience is at least questionable.

The consequences of the employment of the inductive model in the medical processes are far-reaching. We come to realize that medical data bases as they are mainly based on subjective estimates and data like diagnoses, treatments, and probabilities, represent unreliable sources of general information. They contain information strongly colored by personal opinions, interests and concepts. Diagnoses of one physician cannot be compared with the same diagnoses of another physician, and so on.

Unless the medical nomenclature can be clearly defined, these types of information sources can only have a restricted value, and may not contribute to the improvement of the health care.

For the same reason medical classification systems need revision. Even if diseases or diagnoses are inclosed on basis of the highest level of refinement (sharp circumscription), we cannot be sure that one and the same element is inclosed. The current systems reveal the ambiguity of the supposedly impersonal terms. Nevertheless, without the knowledge of the health or diseased state of a population medical science and health care policy run the risk of going astray.

It is our opinion that data-base constructions and medical classification, especially in the world of general practice, must be based on the 'atomic' element in medicine, the symptom. It needs an innovative approach to traditionalistic medicine. It can help to improve clinical judgments and student teaching and learning.

From the latter statement we shall cast a view to the future.

We think that three main routes for future research lay ahead: medical decision making, medical classification, and medical education. These subjects are interdependent and closely interwoven. In the light of the results of this study medical decision making as a discipline has to make up it mind, which of the two options to choose:

1.) - proceeding along the line of the optimizing strategy according to

Decision Theory; or

2.) - adopting the lines of the satisficing strategy which seems much more attuned to the inductivistic and heuristic way of functioning of the medical practitioners.

As already mentioned, the optimizing variant of the decision-theoretical approach appears to contrast with the actual processes in health care. This means, in our opinion, that either the optimizing theory has to build its own medical structure and system to accomodate its theory and its various applications, or it has to drop its claim to be or to become the practical assistance to the routine health care of physicians.

Let us make clear that we do not oppose a reconstruction of medical knowledge, as mentioned on several occasions in this book. It would be the first reconstruction after the eminent works of Thomas Sydenham four centuries ago. However, the challenge is high, the resistance to change too. In a society in which eternity no longer exists, and short term successes are preferred to long term items, this task appears to bear little reward.

Nevertheless, we plan to construct a small scale medical data-base, predominantly based on - with the words of Blois - the 'atomic pieces ', or 'bits' of medical information; on symptoms and their aspects, including signs, tests, etc. For each patient the 'bits' will be clustered into 'chunks' which will primarily be nominated after the physician's original diagnosis. This nomenclature, may be changed when more and more homonymous 'chunks' can be developed. The contents of the 'chunks', the cluster of composing 'bits', primarily can change within limiting ranges, progressively restricting when the 'chunks' get their definitive forms.The scatterdiagrams as presented in Chapter VIII examplify these thoughts.It is our intention to strive for a more complete covering of the "symptom - space" for a particular disease entity.

In this phase the physician is able to call the 'chunks' by means of a (number of) 'bit(s)' in order to compare the contents of the 'chunk' with the symptoms collected from the patients enabling recognition of patterns.

The feedback mechanisms of the system to the physician and the physician to the system should control this medical data-base system. We think it possible that changing societies, changing environments, changing hygiene, changing health states have outdated the current system of diseases and syndromes. The difficulties the - especially family - physicians experience in classifying (i.e. diagnosing) the presented illness picture of the patient underlines these thoughts.

We constructed a model (by lack of a suitable name, we shall call it the R-model containing approximately 6000 items, covering the main bulk of symptoms met in general health care (see chapter VII). With the help of this R-model, we shall try to originate and construct a medical data base which really allows and enables physicians to communicate in a two-way version. In our opinion, the two way communication is an essential condition to persuade medical practitioners to participate and to contribute to an alternative medical knowledge system.

This data base system and a - possibly - alternative medical classification may provide more substantial values for some kind of weighing system, presumably essential to whatever type of decision making system. Possibly, these values can also be incorporated in some kind of Artificial Intelligence system.

It will be clear to every physician that to cope with the overwhelming amount of information in medicine and adjacent domains, and still more to come, some kind of assistence is needed to ascertain a reliable, effective and efficient health care. As we discussed in chapter I, these qualities are more and more questioned. We cannot stay aside and watch developments. We, physicians, must actively participate in these developments.

One of these actions might be the education of our future colleagues. They have to learn substantial methods as they are actually performed and will be carried out in actual practice.

"Training programs aimed at the education of problem-solving techniques

with the principles of formal logic (deductive reasoning), Bayesian statistics and problem oriented medical records incur a risk of transmitting a way of thinking that hardly exists in practice and is probably quite useless"(13).

We are planning to develop a training program for students (and physicians) based on the design of this study. The R-model allows a very good storage capacity and retrieval rate. It can be approached in a written, verbal and a computerized way. Further advantages are:

- it challenges student's problem-solving skills;
- it allows the student to follow his own reasoning pathways;
- it allows good scoring possibilities of the separate data;
- it allows students for an extensive interaction with the 'patient';
- all types of clinical patients can be modelled;
- ambiguities can be prevented by careful designing;
- it allows scoring possibilities for the problem solving methods and pathways;
- it allows revisiting special items;
- it allows retracing of the process' steps;
- it covers the entire medical process, from presenting complaint till establishment of the treatment;
- in case of the verbal mode the cueing problem is avoided (cueing: suggesting questions, options or categories to the testee that otherwise did not cross his mind);
- the construction of a 'paper patient' is easy, fast, and at low costs.

Disadvantages are:

- in case of the verbal mode the use of a competent operator is needed;
- reliable and extensive data collection from real patients, their relatives, and their medical caretakers is needed.

It is our intention to set up training programs for students and physicians (the latter category for continuing education). Whether the R-model can serve as an intermediary for medical audit is still a matter of discussion.

Our aim is to make explicit what thusfar is implicit. To show, to make observable the steps and the actions, the thoughts and the deliberations of physicians solving problems patients present to them. To assist to recognize what is going on for the purpose of understanding and improving the processes of caring and curing people. Perhaps it is this caring aspect of our profession why physicians have adapted and practice an inductive strategy. Because inductive reasoning permits people to recognize individuals.

REFERENCES CHAPTER IX

1. Braithwaite R.B., Scientific explanation. A study of the function of theory, probability and law in science. Cambridge University Press, 1968.
2. Forstrom L.A.,: The scientific Autonomy of clinical medicine. J. Med. Philos., 1977, 2, 8 - 19.
3. Popper K.R.,: The logic of scientific discovery. London, Hutchison, 1959.
4. Medawar P.,: Scientific method in science and medicine. Persp. Biol. Med., 1975, 345 - 352.
5. Johnson-Laird P.N., Wason P.C.,: Thinking, readings in cognitive science. Cambridge University Press, 1977.
6. Polanyi M.,: Personal knowledge. Towards a post-critical philosophy. London, Routledge & Kegan Paul, 1958.
7. Newell A., Shaw J.C., Simon H.A.,: Elements of a theory of human problem solving. Psycholog. Review, 1958, 65, 151 - 166.
8. Harre R.,: The philosophies of science. An introductory survey. Oxford University Press, London, Oxford, New York, 1972.
9. Newell A., Simon H.A., : Human problem solving. Engelwood Cliffs, New Yersey, Prentice Hall, 1972.
10. Pauker S.G., Gorry G.A., Kassirer J.P., Schwartz W.B.,: Towards the simulation of clinical cognition. Taking a present illness by computer. Am. J. Med., 1976, 60, 981 - 996.
11. Miller G.A.,: The magical number seven, plus of minus two : Some limits on our capacity for processing information. Psych. Review, 1956, 63, 81 - 97.
12. Elstein A.S., Shulman L.S., Sprafka S.,: Medical Problem Solving, Harvard University Press. Cambridge (Mass). 1978.
13. Gerritsma J.G.M., Smal J.A. : De werkwijze van huisarts en internist. Een vergelijkend onderzoek m.b.v. een interactieve patienten simulatie. Utrecht, W.U. Bunge, 1982.
14. Janis I.L., Mann L.,: Decision making. New York, Free Press, 1977.
15. Brehmer B.: In one word: not from experience. Acta Psychologica, 1980, 45, 223 - 241.
16. Einhhorn H.J., Hogarth R.M.,: Confidence in Judgment: Persistance of the illusion of validity. Psychological Review, 1978,, 85, 395 - 416.
17. Goldberg L.R.,: Simple Models or simple processes? American Psychologist, 1968, 23, 483 - 496.

SUMMARY

With regard to the question what is to be transmitted in teaching two aspects have to be considered:

a) the contents of the specialty; and

b) the problem-solving methods with regard to the specialty.

On the first aspect an overwhelming amount of books and articles has been written in medicine.

The second aspect is usually only mentioned in passing. In my opinion, this subject has thusfar been greatly undervalued. When one is unable to trace the - problem-solving - processes how can anyone determine the efficacy, the effectivity, and the efficiency of this process, or values the outcome. To state it in Magerian terms: "If you do not know where to go, you may very well end up somewhere else - and not even know it."

How physicians solve clinical problems is the main object of this research. The investigator studied and modeled two of the eldest and famous ways of problem-solving: the deductive and the inductive strategy, with the modern probability reasoning viewed as an extension of the latter strategy.

All 68 physicians who participated in this investigation used the inductive strategy for the -usually four - presented patient-problems. Within the inductive strategy three variants could be distinguished.

The consequences of this finding are far-reaching. As the inductive strategy does not include a logical hierarchy of argumentationsteps, retracing of the process is impossible. (This aspect is also related to our opinions about experience-knowledge and teaching).

As the hypothesis generation is prior to the acquisition of information, this latter aspect can only be viewed in the light of the former, and thus limited to a small number of domains.

As the hypothesis generation is - partly - unrelated to the total available amount of information, the decision making (choosing the ultimate diagnostic hypothesis) will usually follow implicit, personal trends and standards, e.g. satisfying minimal expectations (Satisficing Theory, Simon) or risk-avoiding prospects (Prospect Theory, Kahnemann & Tversky). It suggests a highly personal character of diagnostics and/or the therapeutic management, which is contradictory to general accessibility of medical knowledge and medical teaching.

This feature of personal concepts easily links up with Polanyi's theory of "Personal" or "Tacit Knowledge" as contrasted to "Objective Knowledge" (Popper), which has general accessibility and validity. During the investigation this aspect came forth. The framework of the investigation (patient simulation) and the special definitions and coding of illness elements (symptoms, signs, tests) allowed for comparing similar conceptions (diagnoses, treatments) among the participants. These comparisons confirm Polanyi's theory and the concepts of inductive reasoning. Mutual comparibility of diagnoses seems hardly possible when analysing these conceptions into their basic elements (symptoms etc.).

This aspect touches upon one of the main cornerstones of clinical medicine. When the starting positions have not been unequivocally defined treatment, viewed as the intervention in the natural course of a disease, can only lead to uncertain outcomes.

The lack of standardized medical definitions and a uniform, unambiguous taxonomy inhibits the application of a formalised, normative decision theory for clinical medicine.

Future planning aims at a reconsideration of medical terminology, medical taxonomy and medical problem-solving methods by means of clustering the basic elements of clinical medicine.

Samenvatting

Indien men zich de vraag stelt wat men in het onderwijs moet overdragen, dan richt zich dat op twee facetten:

a) de inhoud van het vakgebied; en

b) de methoden waarmee men voorkomende problemen (in dat vakgebied) kan oplossen.

In de geneeskunde zijn met betrekking tot het eerste facet bibliotheken vol geschreven. Het tweede facet krijgt veel minder aandacht en wordt meestal slechts terloops aangeduid. Naar mijn mening is dit een onderwaardering van dit aspect. Indien men niet in staat is de - probleem-oplossende - processen te traceren, hoe kan men dan de doeltreffendheid (en de doelmatigheid) van het proces te weten komen, of de waarde die men toe moet kennen aan de uitkomst daarvan. Of zoals Mager het zei: " If you do not know where to go, you may very well end up somewhere else - and not even know it".

Hoe artsen patient-problemen oplossen is de vraagstelling in dit onderzoek. De onderzoeker richt zich daarbij op de bestudering en modellering van twee van de oudst bekende wijzen van probleem-oplossen: de deductieve en de inductieve strategie, waarbij de moderne waarschijnlijkheidsleer gezien wordt in het verlengde van de inductieve methode.

Alle 68 deelnemende artsen gebruikten voor het oplossen van de - in het algemeen vier - voorgelegde patient-problemen de inductieve strategie, waarin een drietal varianten voor enige variatie zorgden.

De consequenties van deze bevinding zijn ver-reikend. Omdat de inductieve strategie geen logische hierarchie van (argumentatie)-stappen kent, is het (re)traceren van het proces niet (meer) mogelijk. Dit aspect heeft tevens betekenis voor de opvattingen over ervarings-kennis en ervarings-overdracht.

Omdat in de inductieve strategie de hypothese-vorming vooraf gaat aan de informatie-inwinning, is dit laatste uitsluitend te zien in het licht van het voorafgaande en is derhalve gelimiteerd tot slechts enkele aandachtsvelden.

Omdat de hypothese-vorming gedeeltelijk los staat van de in totaal te verkrijgen informatie, zal de besluitvorming (het kiezen van de diagnostische hypothese) in het algemeen verlopen volgens impliciete, persoonlijke normen, bijvoorbeeld het voldoen aan minimale verwachtingen (Satisficing Theory, Simon), of risico-vermijdende vooruitzichten (Prospect Theory, Kahnemann & Tversky) enz. Dit geeft de diagnose-vorming en therapie-bepaling een eigen en persoonlijk karakter waardoor deze aspecten niet of slecht overdraagbaar zijn.

Dit karakter van persoonlijke concepten sluit aan bij de theorie van Polanyi, die spreekt van "Personal" of "Tacit Knowledge"; dit in tegenstelling tot "Objective Knowledge" (Popper), die algemeen toegankelijk en geldig is.

Tijdens het onderzoek kwam dit vraagstuk eveneens aan de orde. Door de opzet van het onderzoek (simulatie van patienten) en een speciale indeling en codering van ziekte-elementen (symptomen, verschijnselen, testen enz.) waren wij in staat om gelijksoortige concepties (diagnosen, therapieen enz.) onderling (tussen de artsen) te vergelijken. Deze vergelijkingen bevestigen de opvattingen van Polanyi en die van de inductieve inferentie methoden.

Indien men diagnosen als essentiele eenheden van de geneeskunde ontleedt in hun basale elementen (symptomen, verschijnselen enz.) lijkt uniformiteit van diagnostiek, en - dus - vergelijkbaarheid van diagnosen, een ver verwijderd ideaal.

Hiermede komen enkele pijlers in de geneeskunde (preventie, therapie enz.) ter discussie. Indien de uitgangssituaties niet eenduidig gedefinieerd zijn c.q. kunnen worden, dan leveren interventies slechts onzekere uitkomsten op.

Het ontbreken van gestandaardiseerde medische definities en een uniforme taxonomie maakt tevens het eventuele toepassen van een geformaliseerd, normatief besluitvormingsmodel (Medische Besliskunde) illusoir.

De onderzoeker stelt voor om op basis van clustering van de basale elementen der geneeskunde (symptomen, verschijnselen) te komen tot een herbezinning op de medische terminologie, de medische taxonomie en de medische besluitvormingsmethoden.

CURRICULUM VITAE

JACOBUS RIDDERIKHOFF

geboren te HOORN op 2 april 1932

Artsexamen: 1962

Officier van Gezondheid: 1962 - 1963

Huisarts te ROTTERDAM: 1963 - 1975

Medewerker Erasmus Universiteit Rotterdam

Afdeling Huisartsgeneeskunde: 1975 - heden.