A diagnostic decision support system for general practice

drukwerk: Universiteitsdrukkerij Rotterdam.

A diagnostic decision support system for general practice

Een diagnostisch besluitvormings-ondersteunend systeem
voor de huisartsenpraktijk

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR
AAN DE ERASMUSUNIVERSITEIT ROTTERDAM
OP GEZAG VAN DE RECTOR MAGNIFICUS
PROF. DR. P.W.C. AKKERMANS M. Lit.
EN VOLGENS BESLUIT VAN HET COLLEGE VAN DEKANEN.

DE OPENBARE VERDEDIGING ZAL PLAATSVINDEN

OP WOENSDAG 19 JANUARI 1994 OM 11.45 UUR

door

EGBERT VAN HERK

geboren te Amsterdam

Promotiecommissie

Promotores:	Prof. Dr. H.J. Dokter
	Prof. Dr. E.S. Gelsema
Overige leden:	Prof. Dr. Ir. J.D.F. Habbema
	Prof. J.H.P. Wilson

Contents

Introduction

"How do we think?" is a question which has fascinated mankind for many centuries. This book is not an attempt to answer that question, but it does look at one aspect of thinking. An aspect which is just a tiny fraction of the whole field of research on thinking, yet an enormous field in itself, namely medical diagnosis: the way in which doctors try to find out what is wrong with their patients.
This task is not performed logically, although logical, scientific methods have been worked out to mimic, improve or even replace it. Yet although humans do not perform this task in a very logical way, they do perform it in some way, following some rules, and being prone to certain types of error.

It is very hard to know what goes on in the human mind. The invention of the computer, the "thinking machine", has greatly contributed to the interest in the way we think and to research into mimicking thought processes. This research has yielded many successes and surprising results, yet the "artificial intelligence" displayed by current machines is not very impressive from a human point of view. The developing field of neural network research, by attempting to copy to some extent the hardware in our own heads, may offer additional insights into thinking.

This book is the description of an attempt to improve the diagnostic process, by combining human and machine capabilities into a close partnership. It is a report of the underlying philosophy, design, implementation, and a first test of a diagnostic decision support system for general practitioners. The system was designed to assist and improve human reasoning rather than replace it.

Chapter 1. The doctor and the diagnosis

1.1 The concepts of disease and diagnosis

It is curious that most medical curricula do not explicitly teach the student of medicine the meanings of the words 'diagnosis' and 'disease'. The student is assumed to form his or her own mental picture of what is understood by these terms. A result of this is that most doctors have a largely intuitive idea of the meaning of these concepts, as they find out when they start thinking about them. In the next sections we will discuss some of the points which have been brought up with regard to this problem.
Stefan and McManus showed that the concept of 'disease' changes with time for students working their way through medical school, and that the interpretation of the concept gradually gets broader. Fifth-year students regarded significantly more conditions as diseases than third- or first-year students [Ste89].
The reader should take note that the word 'disease' may stand for 'a disease' or for the totality of all diseases, and keep this difference in mind when reading the next paragraphs.

1.1.1 The concept of disease

Having a disease, in other words, being a patient, confers a number of important privileges on a person in our culture. He or she is excused from work, and may be remunerated for treatment and lost time by his/her insurance companies; and he or she will probably be cared for more than would otherwise be the case. Having a disease has far-reaching economic and social consequences. This is why society is so concerned that the decision of whether a person is ill or not should be made strictly and correctly by a doctor. The patient who feels ill is ill until proved otherwise - by the doctor. But the doctor does not relate to this decision in the same way as society does; he does not have the sense of strict demarcation between health and disease which society would like to have him endowed with. His concern is rather with the patient and with influencing the patient's symptoms and complaints in a favourable direction. The doctor's concept of disease concerns prognosis and outcome, rather than the legitimation or otherwise of the sick-role which the patient has assumed or which has been conferred upon him.

There have been attempts to define disease in terms of (the absence of) health, and attempts to define disease directly. An exponent of the first school is Boorse [Boo77] who argues that health can be defined as a value-free theoretical notion, based on biological function and statistical normality. In his view disease is the complement, or absence of health, and likewise definable with value-free definitions. Many others have taken up this challenge and argued that the notion of normality is always dependent on the viewpoint of the observer [Whi81] [Whi81a] [Mer86] [Mil85].

If we attempt to define disease itself rather than trying to delineate it by describing its opposite, we can distinguish several schools of thought.
On the one hand there is a group of somewhat sociologically-oriented writers who stress the arbitrariness of any definition of disease, or a disease, and the extent to which it depends upon cultural values. In their view a disease is composed of a number of defining characteristics, added to which is the notion of the defining body or agent that this is an undesirable situation, and that we would like to influence the situation of the person who is then said to be 'suffering' from these characteristics.
Many authors have attempted to formulate strict definitions of disease that would tally completely with such intuitive opinions as are generally held by doctors. None of these efforts can be said to have been completely successful, although the discussion has brought a number of interesting tacit assumptions and differences into the open. Most authors agree that in any definition of "disease" or "a disease" there is an arbitrary judgement of abnormality [Mar76] [Eng76] [Krä80] [Sca80] [Too81] [Mil85].

On the other hand, we would like to expose another quite different duality of opinion with regard to the concept of disease. This has been discussed by Ridderikhoff [Rid89] [Rid92].

There are two opposing views already present in what is possibly the oldest recorded medical thinking, that of the Hippocratic age. Hippocrates himself was an exponent of a way of thinking, which we will call Coan, after the island of Cos where his medical school flourished. In his writings, the patient and his or her symptoms and other circumstances which are deemed relevant are described minutely. From these circumstances a prognosis is made, based on the experience of the doctor. No specific name is given to the disease; the patient's symptoms **are** the disease.

> "Remarkably, modern readers cannot make any sensible diagnosis from the Hippocratic descriptions. The description is too much patient-oriented and not - as we are accustomed nowadays - disease-oriented." [Rid92]

This view has percolated through the ages, and its vestiges can still be found in homeopathic medicine. But the idea that the patient's symptoms by themselves may sometimes constitute the best description of his disease is widespread among general practitioners and is also very important in psychiatry; indeed the currently popular "DSM III R" classification of psychiatric disorders is nothing but a collection of lists of symptoms.

The other school of thought may be called Cnidian, after the island of Cnidos, which is close to Cos, but where in ancient times a different view of medicine prevailed. The Cnidian school adheres to the idea that a disease should be clearly defined. Galen (Galenus) is an exponent of this school. According to this philosophy, a disease is a predefined entity with specific composing elements, and a new case can be diagnosed by comparing the symptoms of the patient with those of the disease.

> "Modern readers will be astonished to find the portrayals of Galen very much comparable to current descriptions of diseases as found in textbooks.[...] In fact, the Cnidian conception was the first attempt to systematize and classify diseases into a taxonomy." [Rid92]

This concept is still the mainstay of current clinical medical thought, although the taxonomy of diseases which has sprung up in the last few centuries is characterized by fragmentation and haphazardness: different systems are used concurrently, based on pathology, pathophysiology, etiology, biochemistry, microbiology, etc.

And although this Cnidian standpoint is taught in medical schools nowadays, there is, as noted above, still a gap between theory and practice, especially in the cases of general practice and psychiatry.

This gap is caused by lack of knowledge. In many cases we simply do not know what causes the complaints or symptoms of the patients, and we are reduced to descriptions of the symptoms for want of an underlying systematic theory. By combining these and assigning a name to them, such diseases as "sudden infant death syndrome" and "irritable bowel syndrome" have come into being. They are often referred to by acronyms, which serve better, perhaps, to hide our ignorance: SIDS, IBS.

"Diseases without a real substrate, correct legitimation, or proper classification. They are convenient names without explanatory or predictive meaning" [Rid92]. Another way of coping with such patients is by using the Hippocratic patient-oriented or complaint-oriented approach of diagnosing the patient's illness as a unique disease. However, this puts the doctor in a

difficult position when he has to give a diagnosis to placate the insurance company.

Campbell, Scadding and Roberts did a survey of medical men and lay people on their interpretation of 'disease', and from the results argued that the general public's definition was 'essentialist', where diseases are specific, well-defined entities with a single cause. In contrast to this, doctors perceived disease more as a description of an aggregate of abnormal biological phenomena without implying a specific etiology [Cam79]. However, the disease as-an-entity notion also has a considerable hold over the minds of many doctors. In an editorial in the British Medical Journal the anonymous author writes:

> "Unfortunately, many doctors remain blissfully unaware of this logical structure and continue to be influenced by outmoded assumptions. One still hears eminent physicians say "I'm sure he is a schizophrenic even though he doesn't yet show any of the typical symptoms."; and chest physicians still argue whether a patient has chronic bronchitis (a clinical syndrome) or emphysema (a structural abnormality)."[Ano79]

The former doctor still thinks of an underlying disease model in a case where medicine has, temporarily at least, given up the position of a specific entity 'causing' the patient to suffer schizophrenia, and retreated into a list of defining symptoms. In the latter case doctors argue on the assignment of the patient to groups belonging to different taxonomies.

For our current purposes in this work it is sufficient to define a "disease" as the coincidence of one or more specific circumstances or conditions in a person. While we will commonly use the term "symptoms" in this work for these circumstances, it is important to keep in mind that they may include signs, symptoms, tests, x-ray investigations, behaviour, or whatever else makes up the disease. This definition completely sidesteps the important questions of who is to define the circumstances which make up a disease, and what criteria this person or agency should use to select these circumstances. But it has the advantage that such a concept of disease can be made fully operational. To determine whether a patient has any given disease or not, we look for the set of circumstances which define it. If they are present, he or she has the disease by definition. If they are not, he or she has not. Finally, the available information may be insufficient to warrant either conclusion. This position also has the advantage that either the 'Coan' or the

'Cnidian' viewpoint of 'diagnosis' may be used, each where most appropriate.

1.1.2 Nosologies: taxonomies of disease

Taxonomy is the combination of classification, nomenclature and identification [Jac92] [Jac92a]. *Classification* is the ordering of diseases (in the case of nosologies) into groups, on the basis of both similarities and dissimilarities between them. *Nomenclature* assigns names to these groups according to more or less formal rules. (Rather less than more formal, in the case of medicine.) *Identification* is the assignment of a new specimen (disease or patient) to one of the predefined groups on the basis of the observed similarities and differences.

Classifications of objects may be made on the basis of a single distinguishing characteristic (monothetic classifications), or on the basis of several (polythetic classifications). Sometimes the objects to be classified fall naturally into several completely disjunct groups; in other cases there will be a considerable overlap and any line of demarcation drawn between the groups must to some degree be arbitrary. All varieties of this spectrum occur in current medical taxonomies of disease; examples are e.g. infections by different species of bacteria, and degree of malignancy of tumours. (Although even in the case of bacteria there are disputes between bacteriologists over the classification of closely related groups. It is not so easy to find any disease which has absolutely no marginal examples, no cases where there is not room for doubt somewhere.)

The *objective* of nosology is to describe classes or entities which enable the doctor to predict as accurately as possible what a particular patient belonging to that class can expect in the future, and in what degree the natural outcome of the disease can be influenced by medical science. Prediction of outcome, and then the attempt to influence that outcome favourably, is the real objective of diagnosis.

The development of nosologies in the last few centuries was made possible by the advance of medical knowledge. When, in the renaissance, doctors started to examine the human body after death, gross morphologic abnormalities came to light and these led to a pathological nosology. The subsequent rise of physiology and pathophysiology gave birth to a pathophysiological nosology. The discovery of bacteria led to a nosology of infectious diseases. The current 'total' nosology of medical science hardly deserves this name, because it is a combination of all of these, with a few

additional classifications thrown in for good measure. Every disease is placed in the classification in which it seems to fit best, but it may also exist, with slightly different connotations, within another system, e.g. the example of emphysema vs chronic bronchitis given above. Medical science suffers from a lack of knowledge and accordingly frequently has to shift its frame of reference to fit the diseases it endeavours to classify.

In practice, the doctor is forced to work with these nosologies, however inadequate they may be, and with his inadequate knowledge of them. His knowledge is a personal knowledge: apart from what he remembers, or thinks he remembers, from the lectures at medical school, he uses a large body of acquired knowledge in the form of his own experience with patients and with diseases, which is different from that of every other doctor, even if the scientific medical education is not. There is no 'gold standard' definition for most diseases, although in recent years such definitions are being attempted with ever increasing frequency. There is no general consensus between doctors over what the characteristic points of any disease are. There are large differences between the frequencies of even the commonest diseases in different countries; diagnoses which are extremely common in Germany, e.g. "Kreislaufschwäche", ("weakness of the circulation"), are practically unheard of in Anglo-Saxon countries; and vice versa. The journalist Lynn Payer gives a number of amusing examples of diseases which seem to be specific to certain countries.[Pay88]

1.1.3 Diagnosis

The word "diagnosis" can also be used to mean several things, the most important of which are a process, and the conclusion of this process. [Fei73]. We use it in both these senses: The process of diagnosis is the art of finding all diseases which can be established in the patient; the conclusion is the disease which seems most likely after sifting as much information as possible. But the raison d'être of a diagnosis is not simply to 'find the disease': it cannot be viewed as completely distinct from the goal of treatment for the patient [Whi81]. Diagnosis has also been defined as "a provisional formula for action."(Cohen)

Although there is some truth in this, regrettably this dictum is too often viewed as a license to stop thinking about the nature of the patient's problem too soon. Such 'diagnoses' as "IBS" (irritable bowel syndrome), "COPD" (chronic obstructive pulmonary disease), and "PID" (pelvic inflammatory disease) are examples of this. While it is clearly (?!) not necessary to

determine which particular virus, out of many dozens of species, is responsible for the patient's 'common cold', on the other hand if we label the death of an infant with the term "SIDS" (sudden infant death syndrome), we are liable to forget that this 'diagnosis' is nothing but a testimonium paupertatis, just another way of saying that we do not know what caused the death, and to think instead that 'SIDS' is a disease in itself, something you can die of.

Making a diagnosis is a process of identification: we lump the patient together with a number of other patients, under the heading of "disease" X, which is usually called "the diagnosis" for short. But the diagnosis need not coincide with any recognized disease; it may be the point at which the doctor stops his inquiries for the moment [Fei73]. In that case diagnosis is considered a judgement.

Of course, there is always the possibility that the patient does not have a disease at all. This is a frequent occurrence in general practice [Mec78] [Kel76]. If we are to attempt pattern-recognition as a method of assigning patients to disease groups, as we will in this book, there will always be patients who do not conform to any of the patterns we know. According to the taxonomy, these patients then do not have a disease, which is not to say that they do not suffer, or do not feel ill. If any taxonomy is used statically, and not reviewed from time to time, no progress in the classification of disease can be made. The cases which cannot be classified within a taxonomy are just as important for its development as the cases that can be, and perhaps even more important, because they point out the imperfections of the system. Such patients should stimulate the doctor to step out of the system, and consider the system of classification itself.

1.2 Diagnoses based on different types of evidence

Diseases are rather arbitrary entities, although they should always be made up of a specific set of circumstances, such as symptoms, signs, laboratory measurements, and x-ray investigations. As we have seen, these circumstances shift and change with the time, the place, and the person of the doctor.

To be useful, any definition of a disease should be able to be operationalized: it should always be possible to determine unequivocally whether a particular patient does, or does not, suffer from the disease. This may often or even generally not be practical, but it should at least in principle be possible if a disease definition is to be useful. In the final analysis, a disease definition without such a *decision procedure* is a meaningless entity.

However, from a more practical point of view, there may be a number of different symptoms and signs which are not part of the definition of a disease, but which are known to be associated with it in a sufficiently high degree to be usable for a probabilistic diagnosis, especially if the strict determination meets with problems of a practical nature. For instance, to diagnose a myocardial infarction the principal requirement obviously (from the name of the condition itself) is that a portion of the patient's heart muscle has died. Since this is very hard to determine directly, at least without doing more harm to the patient by cutting him open, the diagnosis in hospital is made by a number of 'in lieu' arguments, i.e. a history of chest pain, characteristic findings in the ECG, and a rise in the blood levels of certain enzymes. It should be borne in mind that a diagnosis of 'myocardial infarction' reached by this path does not necessarily correspond to the actual disease state of necrosis of a portion of the heart muscle: there may be some patients who do not have the dead heart muscle but who do have the other symptoms; likewise, there may be patients who do have the dead muscle cells but not the chest pain, ECG findings, and rise in enzyme levels, or maybe just one or two of these. This phenomenon of incomplete overlap provides us with a strong argument not to name diseases after hypothetical causes: if the hypothesis is later revised, the old name will probably persist and cause much confusion. [Ash72]

For the general practitioner, enzyme measurements and an ECG are usually not immediately available; he will have to go by such information as he can gather from the patient during the consultation, such as the chest pain, sweating, possibly irregular heartbeat, shock, irradiation of pain to arms or neck, etc. These signs and symptoms do not constitute strict evidence of myocardial infarction at all, but they are all the information that the doctor has available when deciding to call for the ambulance on suspicion of an infarction, or not.
The criteria by which doctors arrive at conclusions differ with the circumstances and with the person of the doctor. Different general practitioners can have entirely different ideas about which signs and symptoms are the most important indicators of what seem to be clear-cut examples of disease. Ridderikhoff found that for a patient with acute myocardial infarction, 20 physicians who correctly made the diagnosis agreed with one another on the presence of just one symptom: pain. On both the symptoms of blood pressure and perspiration, the next two most mentioned symptoms which were found in the patient, only 11 of the twenty agreed with one another. All other symptoms were mentioned eight times or less [Rid86][Rid86a].

1.3 The goal and the process of diagnosis

The purpose of the diagnosis is one of categorization: we want to group patients together on the basis of similarity, and we want to group them not just for the sake of this similarity in their signs and symptoms, but because we know, suspect, or at least hope that in doing so we will also group them together in terms of cause, prognosis and optimal treatment.
The process of diagnosis will therefore in this work be viewed as a task of categorization: the patient is assigned to a disease name, which is the patient's diagnosis. As many signs and symptoms as possible in the patient should be explained or at least covered by it, and conversely the patient should minimally show other signs and symptoms which do not agree with that diagnosis. Finding a solution within the limits of both these constraints is one of the main difficulties in finding a diagnosis.

1.4 Different types of diseases/diagnoses

A disease entity, a grouping within the classification of diseases, may be based on widely different kinds of arguments. It may be, in its most primitive form, a single symptom, or a small group of symptoms. In the age of Hippocrates, this was a common occurrence: e.g. fever. Nor has this type of diagnosis entirely disappeared in our time: Cough, IBS.
On the other end of the spectrum, we have the highly specific diagnoses, e.g. diagnoses which point to the exact location on the chromosome and the exact nature of a genetic defect or to a specific germ to explain the cause and effects of the disease in question.
Most diagnoses/diseases find a place somewhere in between the ends of this scale. Diagnoses tend to move from one status to the other: they start out as syndromes, collections of signs and symptoms that are observed to occur together in many patients, gradually acquiring more specific causal overtones as more becomes known of the mechanisms at work. This process of refinement continues until the problem seems to be understood and solved to the degree necessary to cure the disease, or else to explain the cause of the disease and the reason why it is not curable.

Due to this natural process of refinement the exact definition of the disease might become blurred. While some doctors may still work with the old definitions, others who are more in touch with the forefront of science, or

who are perhaps just in a position to order the latest tests, will have replaced their own private definitions of the disease in question with a more advanced one. This may be the cause of many medical misunderstandings. Traditionally there is no agency which defines the current standard for a diagnosis, although in the last decades some progress has been made in this respect as groups of experts set out to define diagnostic criteria for diseases.

1.5 Summary

We have discussed different views of disease, diagnosis and the current state of nosology. It is argued that the prime purpose of diagnosis is the classification of the patient into a group of similar patients, thus allowing an accurate prediction of the outcome and the identification of methods to influence this outcome favourably. We try to classify the patient on the basis of similarities and dissimilarities between his or her signs and symptoms and those of a pre-defined pattern called a disease. The standardization of these patterns and their classification into a nosology at present is chaotic at best.

Chapter 2. Computers in general practice

2.1 Introduction

When computers were first invented, many people predicted them to be able soon to perform all kinds of complicated tasks that had hitherto been the exclusive domain of man. The world champion chess programme has been predicted many times "within a few years from now" since before 1960. And although nowadays even chess Grandmasters are occasionally beaten by a computer, and the goal no longer seems to be so far out of reach as to be ridiculous, the prediction has not yet come true.

The goal seems to recede with every step taken toward its realisation, yet there is undeniably a steady progress in the quality of chess-playing programs.

The development of programmes for other difficult tasks like understanding natural language, or diagnosing an illness, seems to be in a similar situation. So far, tasks that can be reliably performed by computers are especially those tasks that can be exactly described in formal terms, e.g. database management. Computers can outperform humans by a very large margin in sheer speed and volume of the data they handle in these instances.
But for many problems of a more complex nature we are still forced either to oversimplify the problem or else to reduce the solution domain to a very small area, e.g. the causes of high blood pressure. Within these small domains, and even in some quite large ones, computers are indeed able to compete with human experts. Yet we find that these programs are rarely, if ever, used in practice, even if studies have shown them to be very reliable, at least in a laboratory setting. The reasons are complex and multiple, e.g. cost, (un)availability, difficult access, distrust, unfamiliarity with computers, and unfamiliarity with the programmes themselves.

2.2 The use of computers in general practice.

Even before the advent of personal computing, there were attempts, e.g. by Preece et al. and Bradshaw-Smith, to use the computer in general practice. [Pre70] [Bra76]. The latter describes a system for a group practice, operating on exclusive telephone lines from a local hospital computer centre. It was intended to replace the NHS envelope system. Gruer even describes a very early off-line batch processing system, which aimed at bringing together all medical data of a patient in a single computerised record. [Gru72] The

record was split into permanent data and episode data. The records were extracted at regular intervals using a special output programme for each purpose, and the resulting lists could be studied by the doctor. The main applications of the system were considered to be patient surveillance, morbidity registration, drug use, work-load analysis and use of services such as lab tests.

As personal computers became generally available in the early nineteen-eighties, a few enthusiastic family physicians who were ahead of their time started using them. Within a decade, a computer became a commonplace item in general practice, rather than a rarity. These early machines were by today's standards extremely limited in working memory (RAM,ROM), mass storage (floppy disks or, rarely, hard disks), and display capabilities (monitor screen).
Yet within these limitations physicians were soon finding useful applications for these machines. One of the first applications was the electronic address book and appointment scheduler; once this was implemented, doctors started to look for ways to extend the benefits of this newly acquired level of organization of their information. One of the first countries to experiment on a major scale with computer systems in general practice was Great-Britain, which had been ahead of the rest of Europe in the use of personal microcomputers. (The BBC microcomputer project was designed to provide both a cheap personal computer and training on how to use it.)

Around 1982, there were systems available for applications such as an appointment system, population register, disease or risk register, repeat prescription register and 'various other logs'. [Met82]
Doctors started to use their computers to write prescriptions (Preece) or monitor repeat prescriptions (Difford) [Pre84] [Dif84]. Note that by 'writing prescriptions', what is meant is that the doctor determined what to prescribe; the computer just printed the recipe.

There was some debate about whether doctors should program their computers in BASIC or Pascal [Mul84]; that they should program their own systems was fairly natural, as there was very little software readily available. If you needed something, you either wrote it yourself or went without. This of course soon started to change.

A number of studies emerged reflecting the increasing use of fledgling administrative systems by general practitioners, eg. monitoring a population at risk for a specific disease or other factor: high blood pressure, cervical smears, flu shots.

There were already some early concerns about increasing computer usage, too:

- Would not the computer form an impediment to the doctor-patient interaction?
  Answers to this question could at his stage only be intuitive; Metcalfe thought it would not. [Met82]

- How reliable is the computer technology?
  Emrys-Jones gives a tale of mistakes not to repeat when installing a computer. From this we can conclude that any doctor or practice acquiring a computer needed in that period a remarkable single-mindedness, a willingness to learn programming skills, a close working relation with a reputable dealer to solve the numerous hardware problems, and a fair amount of money to achieve an acceptable result. [Whi82]

- It is not enough to have the computer keep the records: they must be entered and updated meticulously to be of any use.
  Sprackling lists a number of problems encountered [Whi82].
  He found that it was hard to motivate personnel to keep the database up-to-date if they were not trained well and were not involved in the use of the database. He found that after one year, quite a lot of patient records contained either errors, especially wrong addresses, or omissions, especially missing diagnoses, or both.

Reading these early publications, one feels that buying and maintaining a computer system at that time must have been worthwhile at least as much for the sense of adventure and pioneering as for the actual medical benefits, and probably more so.
Metcalfe made an important point about the use and usefulness of computers in general practice:

> "The step to the narrative record will not be worth the investment in resource or energy if it is purely to store the current undisciplined paper record in electronic medium. It will only be worthwhile if two preconditions are met: firstly, that some standardized form of recording such as problemorientated medical records are accepted and, secondly, if it is accepted by the doctor that he is handing over a certain amount of control to the system which will prompt, cue and remind him of things that have to be done". [Met84]

Of course, long before the first computer was installed in a general practitioner's office, much had already been said about the application of computer technology to medical problem solving, and much research had been done in academic centres where large mainframe computers were available. Shortliffe et al give a review. [Sho79]

2.3 First use of the computer as a diagnostician

From the use of the computer as a fast and reliable record-keeper to its use as a diagnostician is a big step, even a bigger one than was at first thought. It presupposes an accurate medical description of patient symptoms and disease entities, knowledge about diseases, together with machine with the speed and memory capacity to handle both. There turned out to be a big gap between AI research on the one hand, and medical knowledge and medical problem solving on the other. Mcmullin points out a number of obstacles to various methods which might be used [Mcm83]. Attempts to simulate the thought processes of the clinician disregard the fact that we do not know accurately what they are. They are not learnt from a manual nor are they ever set out explicitly in the ordinary exercise of the skill [Pol58]. Pattern-matching approaches work only in well-defined and limited areas of medicine where the classification of data presents no serious problems [Mcm83].

In the early period of research in medical computing, the most common approach to the problem of making a diagnosis was pattern recognition [And68] [And68a] [Boy68] [Bol75] [Led59].
Even before computers were actually used, there were some pattern-recognition based approaches using mechanical sorting devices. [Bro59]
However, this method was soon eclipsed by the artificial intelligence techniques that were rapidly being developed. The two fields of common origin grew apart and each came to have its own domain: Pattern recognition being used mainly for mechanical tasks that could be transformed to fit the method as closely as possible, such as ECG interpretation or white blood cell recognition in smears, while AI concentrated on more 'fuzzy' problems like diagnosing illnesses in specific domains [Gel89].
The type of AI research done in university centres was very often limited to small specialty domains. The broad range of complaints the general practitioner sees was felt to be too large and too ill-defined to be harnessed by programs using the AI approach. Furthermore, the transfer of research results to the family practice setting has always been strongly limited by

available technology and cost. It is only now that cheap powerful systems are within reach of the average general practitioner that research into computer use in general practice has really taken off. The possibility of a computer in the doctor's office has given the GP a potential direct benefit from such systems - he does not have to wait anymore for the results to be mailed back to him after evaluation on the large computers of a university research centre.

2.4 Modern developments

As has been stated countless times, the power of personal computers has dramatically increased in the last decade. Instead of at 2 to 4 MHz they now run at 20 or 40; instead of using an 8 bit databus, they now use 16 or 32 bit databuses; instead of 20 to 64 Kilobyte of memory, they now have 1 or 4 or more Megabytes; instead of 160 Kilobyte single-sided floppy disks they use 1.2 and 1.44 Megabyte ones; and instead of no harddisk at all, 100 Megabyte harddisks are now quite common. In lay terms, this means that a personal computer in the year 1992 may be from ten to a hundred times faster and more powerful than one from 1982, as well as being considerably cheaper.
By this development much more ambitious applications came within the reach of the personal computer. Problems which used to require large amounts of data space and processor speed, incompatible with contemporary personal computer technology, can now easily be implemented on a fast personal computer.

But there are other obstacles to be overcome.
Shortliffe [Sho89] cites a number of psychological obstacles to implementation of decision-support systems, as apparent from from surveys of physicians' attitudes. Among these obstacles are:

- Fear of loss of rapport with patients.
  Physicians do not like to sit behind a terminal interrogating the patient, having to divide their attention between the two.
- Fear of loss of clinical control.
  The doctor wants to be responsible for the patient, and not to delegate this responsibility to a machine.
- Inertia: physicians feel their lives are complicated enough already without having to learn a new and complex computer programme and/or system.

- Nonacceptance of machine capabilities.
  Doctors just do not believe that a machine can help them do things which they find difficult.
- Suspicion of Artificial Intelligence.
  Human experts do not trust the data that have been fed into an expert system, and are keenly aware of differences between human experts.
- Fear of legal liability. Who is responsible for managing the patient, and who is liable in case of errors?
- The challenge of data entry.
  Doctors do not like typing; and they do not like rigorous menu systems which present questions that must be answered. A free-format voice data entry system might be ideal.
- Age. Most physicians regard support systems as something for younger generations.

As Shortliffe [Sho87] also points out, any medical computer programme can in a wider sense be viewed as a medical support system. He goes on to define three subcategories, ranging from generalized to patient-specific:

a) Tools for information management. e.g. A hospital information system.

b) Tools for focusing attention. e.g. A clinical laboratory systems that flags abnormal values.

c) Tools for patient-specific consultation. Such programmes provide customized assessments or advice based on sets of patient-specific data. They may follow simple logics (such as algorithms), may be based on statistical theory and cost-benefit analysis, or they may use numeric approaches only as an adjunct to symbolic problem solving.

It is with this third type of programme that we will be mainly concerned in this study.
As has been said earlier, in many cases administrative systems are also used to store medical data about patients, and this area is growing quickly. This medical information is usually typed in by the doctor in a free-text format, generally under a few different subheadings. This has the advantage of legibility over the old written record, but it is still of limited use for research purposes.

2.5 The development of computer use in general practice in the Netherlands.

General practitioners in the Netherlands are still among the Dutch professional groups with the lowest usage of computers in their daily work. But the number is rising rapidly: in 1983 about 0.6% of practices used a computer, in 1988 about 20%, and in 1990 about 35%, mainly for administrative purposes. [Höp90]
In 1989, The Dutch association of general practitioners (NHG) published a set of standards for patient registration database systems, specifying a full relational data model, coding tables, and response delay times. This standard was based on experiences since 1983, when experimental projects had started. This standard has been updated a number of times. A number of conforming or near-conforming database systems are now available [Höp90] [Wes89]. The model is called WCIA-HIS.
Although in 1992, about 50% of general practitioners were estimated to use a computer with a conforming HIS , only about 1 in 10 of these was estimated to use a medical module in addition to the basic administrative module. These numbers are expected to rise with the introduction of financial incentives in the future. [Has92]

2.6 Modern programmes used in general practice

Systems currently used in general practice are typically used for administrative purposes: to record names, addresses, telephone numbers, visits, medication, and insurance data of patients. Moreover, they are used to write bills and prescriptions, to schedule appointments, and to generate reminders. They offer the possibility to enter medical data about the patient in a free-text format, usually with some structuring (the POMR, or Problem-Oriented Medical Record, is popular.)

They are also rather expensive, both in purchase price and in maintenance fees, at least when compared to off-the-shelf applications used in business. This is due to the high cost of software design for a relatively small market, and the need for instant repair and backup facilities if the system should break down. Maintenance is typically a large part of the budget for software and hardware.
One of the most popular systems in use today in the Netherlands is MicroHIS. Another system in use today in the Netherlands is ELIAS, an

administrative system for general practice which has been designed in a modular fashion. A core system may be introduced in the practice, with the option of adding modules later as needed. It offers interfaces to existing hospital information systems. While it was originally an administrative system, there are now modules available for drug treatment, medical data record keeping, etc.

2.7 Problems with the free-text approach

Free-text medical records suffer from the following disadvantages:

1) Personal terminology:
   Every doctor uses his own terminology; what one doctor calls a myocardial infarction, another may refer to as a coronary infarction, ischaemic heart disease, a heart attack or just MI.
2) Ambiguous terminology:
   If two doctors use the same word or term for a disease or symptom, they do not necessarily mean the same thing. There are many terms that are so ill-defined that it would be better not to use them at all.
3) Varying precision of description:
   Doctors use widely differing levels of description of diagnoses; one may content himself with describing a patient as having COPD (Chronic Obstructive Pulmonary Disease), while another may diagnose the same patient as having bronchitis superimposed on existing asthmatic constitution due to a bacterial infection. Both intend to describe the same thing but one of them takes much more trouble than the other to write it down, and also gives a much more accurate representation of the patient's state.
4) Unstructuredness of records:
   The information about the patient is often written down in a more or less unorganized way, with symptoms, complaints, problems, diagnoses and solutions interspersed. Most computer systems make at least an attempt to split the data into sections. e.g. problem, findings, evaluation and plan, but this is only a first step on the road to structuring medical information.
5) Use of the keyboard.
   The text approach requires typing skills from the general practitioner, who must sit facing the computer, rather than the patient, in order to type. Typing takes time and attention, and the typist is liable to make errors which take time to correct.

These are major obstacles to extracting, structuring and using the medical information on a patient in free-text format. The more structure a programme forces on its users, the less flexible it becomes and the more the users will feel constrained by the programme, but the better the generated data can be used and evaluated by other programmes, e.g. for research. This is an example of a trade-off where it is hard to find an optimum.

2.8 Structuring patient data

There are basically two ways in which structured patient data may be obtained:

a) The data may be entered into the system by the doctor in a structured way. This is the easiest type of system to design and implement but it has the disadvantage that it is more work for the doctor. This is not necessarily a bad thing if the doctor (or rather his patient) also benefits: if the doctor has to think a little before writing things down, this may improve his patient management. But the burden of proof for this lies with the systems designer, and the doctor will frown upon anything that makes his work more complex or time-consuming.

b) Attempts can be made to increase data structure by processing unstructured data. This is an area in which especially in the last decade a large amount of research has been done.

Various ways of manually structuring or flagging the information in patient records have been tried. Three examples:

1) There is sometimes a possibility of putting an electronic "tab" on a record if the patient has a specific disorder or complaint: for instance high blood pressure. In this way, the doctor can manually mark a few specific diseases for patients he wants to keep track of. This requires both an explicit judgement from the doctor: I want to keep an eye on this one, and a specific action: press a key, or select a menu to put a tag on the patient. Doctors may forget to do this, or just not care enough to bother. It does not modify or structure the data in the record, just flags it.

2) There are also systems that attempt to standardize the terminology used by doctors by offering them lexicons and coding systems in which they can look

up terms to find a preferred one. There are currently classifications available for diseases, reason-for-encounter, and medication [Lam86].

3) Structured data entry.
The data the doctor enters into his system may also be structured right from the start. This may take the form of questionnaires: the programme must have an answer to a number of questions it poses, and the doctor or the patient has to provide these. This is a considerable limitation of the doctor's freedom. He may resent having to fill in lists of symptoms that may seem completely extraneous.
The work of the Dombal is a good example of this; in a number of experiments using the computer to diagnose acute abdominal complaints the methodology used was to gather data by means of forms, with lists of questions that should systematically be answered. [Dom78]
On a related note, there have also been attempts [Qua86] to transfer this burden to the patient by letting patients fill in electronic questionnaires before seeing the doctor in a hospital. Though most patients did not mind this and were perfectly able to do so, a few of them could not learn to handle even the simplified data entry machinery. Also, the questionnaire was necessarily a global screening one, checking the performance of a number of bodily functions and general well-being. These data were not specific enough to base a diagnosis on; rather they were meant to alert the doctor to possible problems that might not otherwise be discovered in a specialist consultation where questioning is generally in the line of the expected problem field.

Automatic structuring of patient data.
As an example of b), there have also been systems developed that attempt to extract structured information out of the 'raw' patient records by analyzing their contents with lists of synonyms for medical terms. These lexicons quickly grow to enormous size, and cannot deal with the problem of doctors not meaning the same thing when using identical terms. Nevertheless, some success has been achieved in this direction, especially when used on hospital records for retrospective research.

2.9 Use of structured data by other programmes.

2.9.1 Hypercritic.
Some programmes process medical information extracted from patient records in order to evaluate the medical treatment of patients and alert the

doctor to mistakes or omissions. One example of this is Hypercritic, a critiqueing programme for the management of high blood pressure in general practice [Lei91].
The purpose of this system is to advise general practitioners on the treatment of patients with hypertension.

Hypercritic uses data stored in the patient records, thus avoiding a consulting-style interaction with the user. It examines the patient record in an attempt to discover the physician's actions and decisions. These are called events. The events are then scrutinized and where appropriate, comments are assigned to them. These comments are called critiqueing statements or recommendations. They are all suggestions to the physician to modify his actions. To generate these statements, Hypercritic has its own medical fact database where information about hypertension and its treatment is stored. A main characteristic of this system is therefore its background activity: the user does not have to call it, it does not ask questions, it just monitors the doctor's actions from available data. This makes it very easy to use: just install it and forget about it. If the user makes a perceived mistake, Hypercritic will pop up and remark upon it. This is just a private reminder to the doctor, who is free to follow it as he likes and is not belittled or put to shame before a colleague or a patient. These are all big advantages.

Problems with this approach are that doctors must enter the data before they can be monitored, and that Hypercritic must be able to fit the actions of the physician to its database of the known actions. There is a considerable translation effort, and the data may be incomplete. The knowledge domain is extremely limited. Hypertension, though common, is just one health problem, with a fairly standardized management which has been researched extremely thoroughly.
To generalize this effort to other medical problems would be a huge task.

2.9.2 Use of structured data by other programmes.

Expert systems, which use production rules and an inference engine to draw conclusions from data, are highly dependent on the availability of specific data: if a rule states "if x and y then z", the programme has to know whether x and y are present before it can conclude z. Such programmes often require specific information to be entered before any conclusion can be drawn at all. Furthermore, a number of these systems use Bayesian reasoning to produce probability estimates for specific diagnoses, based upon prevalence data and diagnostic predictive value of test results.

In general practice, often very little is known about the prevalence of diseases. Such figures are usually based upon hospital frequencies, which do not reflect the prevalence in the population. Diagnostic tests which are quite accurate in a hospital setting, may have little or no value to the general practitioner because of high false-positive rates and low prevalences; also, the false-negative rate in general practice may be much higher because patients come to their doctor before the disease has had a chance to evolve to its full-blown form; the general practitioner may see the early, hard to recognize signs and symptoms of a disease.
Besides, especially the general practitioner is faced with a limitless number of conditions, many of which are not 'medical' at all in a strict sense.

Accuracy figures for the diagnostic performance of doctors are hard to come by and difficult to interpret. How should a diagnosis be judged ?
Let us suppose that a patient suffers from anaemia caused by excessive menstrual blood loss, causing an iron deficiency. She goes to see her general practitioner, who diagnoses 'anaemia', not otherwise specified, and stops at that point. He is not wrong, but he could have done much better. How much should we subtract from his score for not inquiring into the cause of the anaemia?
Besides the accuracy of 65% quoted by De Dombal [Dom72] for acute abdominal complaints, others have found similar values. Ridderikhoff found an accuracy of 47% for general practitioners solving a common disease presented to them in the form of a simulated patient [Rid86]; and an only slightly higher accuracy for general internists solving the same problems.

Such figures indicate a large potential benefit from improvements in diagnostic thinking. Whether this benefit is attainable in practice remains to be seen. One possible approach is to try and use the possibilities of the computer to improve diagnostic accuracy.
We propose a method of using a computer in general practice for diagnostic decision support, which is much different from all systems which have been described above. It is based upon three principles:

a) Strictly structured coding of patients' symptoms, complaints and test results.
b) A very simple user-interface which offers the doctor complete freedom in deciding which symptoms to enter and in which order.
c) An extremely simple, yet robust algorithm for generating a differential diagnosis.

2.10 Evaluation of decision support systems.

If a diagnostic decision support system is to be accepted, it should be evaluated for its adequacy first. According to Taylor [Tay90], most evaluations of decision support systems are still limited to diagnostic accuracy, whereas in a real-world setting, usefulness and acceptability should be taken into account too. The programme for diagnosis of abdominal pain complaints, designed by de Dombal and others in the early 1970s, has been tested extensively over the years and was shown in 1972 to have a 91.8% accuracy compared with 79.6% for clinicians. However, in a prospective trial involving 6962 patients, accuracy fell to 42-59%, compared with 65% for the physicians [Dom72] [Sut89].
The cases in this last study were from very different populations, and the programme did worse than any of the physicians in the comparison. This phenomenon is often called 'brittleness': programme performance deteriorates spectacularly when the input is taken from a different source than was used in the development environment.

One suggestion to explain this discrepancy is that many of the benefits in the previous trials may have come from standardization of terminology and feedback to physicians rather than from the decision-making power of the programme itself. This theory is supported by the finding that physicians using the programme improved their own diagnostic accuracy of the acute abdomen while they used it, and that this accuracy fell again when they stopped using it. [Dom87]

The above may be seen as evidence to support that

a) general practitioners see a different kind of patient compared to emergency room doctors; and
b) standardization and structuring of terminology in itself may have a beneficial effect on diagnostic accuracy.

2.11 The doctor-patient-computer triangle.

Some writers have voiced concern over the patient's attitude to the computer, and the disturbing influence the computer may have on the consultation. Moser wrote: "I do not want to have to glance up from a computer printout

as I address my patient. I want eyeball-to-eyeball contact." [Mos75]. This is of course a legitimate concern. However, if the doctor can refrain from looking at his paper record in those first five minutes he can also refrain from looking at the computer printout (or rather the computer screen, nowadays). One might say that if the doctor does not find it a hindrance to work on his notes while listening to the patient, there is no reason why he should not work on his notes with the aid of a keyboard. But this presupposes that working behind a computer screen is a task of a complexity comparable to writing notes on paper. For many physicians, this is clearly not true.

Furthermore, patients may dislike the idea of the computer in the doctor's office; however, research shows that the large majority of them does not. Some even consider it a positive development, showing that their doctor is abreast of modern developments. There is also concern about the level of confidentiality of electronically stored patient data. Experiments where doctors made their entire record system available to their locum while they are absent on holiday, even after getting approval from the individual patients concerned, have been criticised as an unnecessary breach of confidence.[Höp90]
Although the locum could theoretically look up any patient in the old paper record system, if he went through the trouble, this is not quite the same thing as having the entire administration of a colleague on a compact-cassette-size tape in your pocket. There is no guarantee against illegal copies being made, which would be unpractical with paper records but which is a matter of a few minutes with electronic media, and which is also undectectable.

2.12 Use of medical information of patients for other purposes

There is a conflict which always crops up sooner or later when patient data are stored in a computer. Although the original aim of record-keeping is to help the doctor manage the patient's problems more efficiently, there is always the temptation to use the information for other purposes. This is a process that should be recognized and monitored carefully, as patients' privacy may be in jeopardy; also there is the more fundamental question of whether a patient's medical data should be used for research at all without his or her informed consent.

There is no consensus on this point. The feeling of the author is that it is allowable, but only if the patient's privacy is guaranteed and he or she is in no way inconvenienced or disadvantaged, and there is a clear benefit from this research for medical science or patient care. Many others hold stricter views.

The system described in this work is partly justified by the possibility of using the patient data for research. However, actually using patient data in this way may turn out to be problematic because of ethical constraints.

Chapter 3. Modelling the diagnostic process

3.1 Introduction.

The diagnostic problem-solving process may be viewed in two ways:

a) As the identification of the patient with a predefined entity within the taxonomy of diseases with the goal of prediction.
b) As a process of reasoning to justify the assumed resemblance.

Both views are objects of study in this chapter. They represent the processes which can usually be observed in human problem-solvers. After all, it is the problem-solving of physicians which we want to improve. Not by attempting to make radical changes in the physicians' behaviour, but by supplying nudges and hints which may help the doctor avoid making errors on the difficult path from patient's complaint to diagnosis.

Many models have been proposed for the diagnostic process; models originating from various disciplines such as cognitive psychology, philosophy, information and decision theory, and artificial intelligence. Some models are intended to simulate what actually goes on in the doctor's brain; others are cast in a form so as to represent a way in which computers might arrive at a diagnosis. These latter models represent a logical way of reasoning which seems rather remote from the day-to-day working pattern of the physician. The first results with the use of these formal approaches were very impressive in the achieved level of diagnostic accuracy [Dom72] [Dom72a] [Dom87]. In day-to-day practice however, the effects were less striking. Howell complains that attention is focused too much upon the methods people ought to use, and not on 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." [How82]
The principles of modelling, including the principles of normative modelling, have contributed greatly to the understanding of processes that are used in medical problem-solving. We shall proceed to discuss a number of these models.

3.2 Medical Decision Making

Several attempts to gain insight in the medical diagnostic process and to optimize it have been made and are still being made. Among these attempts are the Problem-oriented-medical record, medical audit, protocols (how to diagnose and act in certain cases prescribed by experts), flow charts, and several variations of medical decision making. So far, the results of all these methods are discouraging when applied in practice.
Medical decision making is based on Decision Theory, a group of related constructs that seek to describe or prescribe how individuals or groups of individuals select a course of action when faced with several alternatives, having a variable amount of knowledge about the determinants or the outcomes of those alternatives. The theory can be divided into two types: on the one hand, a theory concerning descriptive behaviour (how people do behave), and on the other hand a theory concerning prescriptive behaviour (how they should behave). Decision making can be defined as the process of thought and action involving an irrevocable allocation of resources that culminates in choice behaviour.
The decision-maker invariably has to choose from a number of alternatives, either diagnoses or therapeutic actions. The quality of the decision depends on how well the decision maker is able to acquire information, analyse it, and evaluate and interpret this information so as to discriminate between relevant and irrelevant bits of data; it also depends on how well the decision maker is able to cope with the stress which invariably accompanies important decisions.
People are inclined to reduction and simplification, in order to avoid data overload, a tendency which after Mischel is called "cognitive economics." [Mis79]
Cognitive economics produces several dangers such as: erroneous routes of simplification, unpredictable growth of self-knowledge, and unchecked rules for self-regulation with maturation.
Medical decision making is based on a theory which has not originated within the medical world. It can be viewed as a multi-step 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 steps can be described as:

1) ascertain the need for a decision;
2) establish decision criteria;
3) allocate weights to the criteria;
4) develop alternatives;
5) evaluate alternatives;
6) select the best alternative.

A basic thought behind this approach is the assumption that medicine as a discipline starts from clearly defined entities (like the elements in chemistry) which can unequivocally be distinguished. It assumes causal relationships between the composing elements such as symptoms and signs, and a clear notion of the probabilities of the distribution of diseases as well as symptoms and signs. As we have discussed in chapter 1, this notion applies to only a few groups of diseases and certainly not to the majority of illnesses as they are presented in the consulting rooms of general practitioners. The lack of success of medical decision making in this field must partly be attributed to this deficiency in the discipline of medicine itself. It is essential to both the disciplines of medical and information science, to strive for one line of research, one common view of the medical process, one shared language. Both disciplines have to analyse their methods and contents in a rational and objective way. For the medical part we may agree with Taylor when he states: "It is, therefore, more promising to begin projects of this kind with an analysis of the decision 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." [Tay76]

3.3 Cognitive science

Cognitive science is a complex of disciplines, among which are cognitive psychology, (artificial) intelligence, problem-solving and reasoning processes. They all attempt to describe what actually goes on in the human brain during intellectual processes. Cognitive science is mainly approached from the viewpoint of problem-solving, but intelligence tests and the justification of outcomes also belong to this domain.
Within the context of this book we are mainly interested in the problem-solving activities of the medical doctor when facing a patient's problem.
Within this process four activities may be distinguished:

-observation;
-validation;
-identification;
-justification.

Observation

The first step in the diagnostic process is certainly one of observation. It is one of the doctor's foremost abilities: to listen and observe, to sense and feel what is wrong with the patient. In a way, observation is the essence of the physician's art. The Oslerian concept: "Listen to the patient, he is telling you the diagnosis" is still largely valid. Regrettably, the art of observation seems to become more and more eroded as a method of fact-finding [Rei78]. Many doctors rely more and more on lab tests, and place a higher trust in millimoles than in their senses. But most diseases and syndromes are defined by their specific configuration of detectable phenomena as they may be observed in the patient. Identification of a disease in the patient requires clear and unbiased observation, looking and listening.

Traditionally the taking of a history has been viewed as the most important part of the diagnostic process. It is also the part in which the patient has the opportunity to express his sensations, interpretation, views, fears and needs. It is the part in which the doctor generates diagnostic hypotheses which may subsequently be tested. This part of 'fact-finding' has been somewhat neglected during medical training. Apart from the vocational training for general practitioner, hardly any attention is paid to this important aspect of the diagnostic process. It is perceived as time-consuming and less rewarding in terms of fact acquisition, at least of facts relevant to the doctor [Rid93].

During the training of doctors, much emphasis has been put on the observation of the body. Inspection, auscultation, percussion and palpation are methods by which the doctor can glean valuable information from the patient's body. In various studies these methods appear to be less reliable than would be desirable. Many signs are difficult to interpret and variable in their presentation. But the observer, the doctor, is also prone to error.

As Popper said, observation is 'theory-laden': we observe particular things in a situation because we have theories which assign relevance to some of them and not to others [Pop72]. There cannot be a purely neutral and indifferent collection of clinical facts. Each observation is person, time, and place-based. What is really worrying in the construction of a reliable database is that these highly personally-coloured data give rise to several

kinds of errors in the data-acquisition and -processing operation. These errors can be of many types:

1)	errors of omission:	neglecting to observe;
2)	errors of detection:	failing to observe;
3)	errors of interpretation:	dependance on the definition of cut-off point: the point where the calibration changes from normal to abnormal;
4)	errors in recording:	symptoms only relevant to the task and doctor are memorized and recorded (Nobrega et al., [Nob77]);
5)	errors of memory recall:	inaccurate recall of assumed similarity of clinical patterns;
6)	errors of retrieval:	failing to find memorized and recorded symptoms and signs, test-results, therapies, etc.
7)	errors of identification:	incorrect classification of the patient in the medical taxonomy (inaccurate diagnosis);
8)	errors of terminology:	the use of homonyms and synonyms in describing illness phenomena of the patient.

Errors of detection are hard to compensate. Several factors may cause the failure of observation such as short nights, fatigue, bad temper, noisy circumstances, time pressure, annoying patients, but also a lack of interest and training. The best observer is a trained observer: he who knows what and where to look for. Observation requires an open eye, skilful hands and a perceptive mind: being alert to all findings whether they fit your preconceived ideas or not.

But it is very difficult to follow the line of an open mind. As the philosopher Blaise Pascal stated: "We impose our conceptions, our ideas and thoughts upon reality which creates prejudice. Prejudice precedes our view, our observation, and will determine what we shall see." The first diagnostic idea the doctor gets in mind may guide him in his observations. It makes observation not 'neutral' but 'hypothesis-driven', which narrows the scope of his investigation.

In physical examination many valuable facts can be discovered by careful inspection, auscultation and palpation. Time pressure may detract from an optimal search for relevant clinical phenomena. Protocols and guidelines,

therefore, may serve the doctor in these circumstances. These protocols may also help the doctor to overcome the error of omission. By offering hints for observation the doctor may become attentive to several phenomena that otherwise would have been overlooked. Hints may also help in recalling certain test procedures or particular questions. The offering of lists of possibilities may assist in decreasing these types of observational errors. Clendening and Hashinger put it in the following way: "How to guard against incompleteness I do not know. But I do know, that in my judgement, the most brilliant diagnosticians of my acquaintance are the ones who do remember and consider the most possibilities" ([Cle47], as quoted in [Led58])

Interpretation errors depend on the distinction of health and disease. Since this distinction can scarcely be made, interpretation errors in terms of specificity (defining non-patients) and sensitivity (defining patients) have a strong impact on health care.

The term "recording errors" applies not just to the written statements of doctors but also to their memories. The unreliability of people's judgement when based on memory recall was convincingly proved by the studies of Kahnemann & Tversky [Kah72].

Apart from some forms of "forgetting", sometimes information does not get full attention. Certain facts, which appear to the doctor to be the leading facts, are given with much emphasis and in detail by the patient, while other facts, which appear to be subordinate or trivial, are partially suppressed. However, it will commonly happen that the leading facts turn out to be the ones that had been passed over as negligible.
Wagner et al. collected about 1000 papers on errors in medicine, of which 383 about errors in diagnosis. [Wag78] "Although we know it (making errors), nevertheless, we hold the most striking judgements, data and findings for pure gold, and build with these elements our judgements, albeit in general their range is unknown to us." [Wag64]

The main problems of observer 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?
[Gil73]

Transfer of data can be regarded as 'information transfer" if and only if the meaning of the data is the same to the sender and to the recipient. As medical terminology lacks overall consistency and uniformity, the abstract terms used in clinical medicine often hide a multitude of individual meanings and thoughts. In order to overcome the lack of a uniform and consistent terminology, several attempts have been made to set guidelines for the preparation of monolingual vocabularies. A classified vocabulary is the first step towards standardization of terms and concepts. These guidelines are just a means to an end: to ensure that words are used in such a way that readers and listeners understand them to mean what the author or speaker had in mind when he used them - or, as Young put it, "to get an idea as exactly as possible out of one mind into another". [You82] For the sake of understanding and retraceability, standardization of terminology is a prerequisite for sound and accurate medical practice. Such a standard will ensure data integrity, consistency, security, reliability, and ownership.

Validation

Information gathering is the accumulation of a profile of data about 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. But "facts" are only "facts" if they carry a particular meaning to the doctor at particular moments in a particular process. There is evidence that symptoms are not the elementary pieces in this process, as is usually assumed. Depending on the nature of the symptoms and the context in which they are found, their meaning may vary. Patients can notoriously be vague and inconsistent in their descriptions, the doctor's perceptive facilities may be sub-optimal, doctor-patient misunderstanding, language problems, misinterpretation of physical findings, and many more aspects may distort compatibility and consistency of physicians' databases. The suggestion that backfeeding the information to the examining physician would help both to reduce the differences between physicians, and between the separate recordings of one physician can barely be sustained [Lah78] unless the problem of inconsistency can be solved.
The variability and inaccuracy of data [Kom79] made physicians turn to information which appears more stable and accurate such as graphs, pictures and chemical tests. Unfortunately, this will not bring us much further, because diseases are described by their clinical picture, the specific configuration of symptoms and signs, as was discussed in the first chapter.

In the theory of clinical decision analysis the relevance of a symptom is often expressed by its degree of contribution to a particular disease. The diagnostic value of a symptom is disease-dependent and its relevance is likely to be different for different diseases. [Cum76] 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. [Wor72] The mathematical theory starts from the principle that symptoms are the atoms of medicine. When we know the configuration of the disease we only have to weigh the symptoms and everything can easily be fitted in the right place, using only a programme to find the "goodness of fit". Regrettably, it is not as simple as that. Every physician knows that the same symptom can have different meanings not only in different diseases but also in the same disease. Pain in the chest in a young person can be indicative of a pneumothorax, whereas in an elderly person it may point to a myocardial infarction. It can make a lot of difference whether a symptom presents itself in an acute, an intermittent or an insidious way, even if it concerns one and the same disease.
The present habit of doctors to order multiple tests does not seem to contribute to a clearer understanding of the problem or to help in diagnosing. Test results are only helpful if:
a) they are all normal, thus tending to exclude a disease;
b) they are all abnormal, thus tending to confirm a disease.
They are least helpful when some are positive and other negative [Gri81], as is often the case in real-life situations. The interpretation of the results is also influenced by the definition of the cut-off point, the point where the calibration changes from normal to abnormal. When this point is moved towards the range found in diseased people, specificity (defining non-patients) increases but sensitivity (defining patients) decreases, and when it is moved in the opposite direction, the reverse is true.

Another source of error is formed by the process "massing of data". [Lic78] If a doctor is confronted with a patient's complaint of stabbing pain on the chest with exertion he mentally completes the picture with all data concerning Angina pectoris. "Massing of data" (defining syndromes) enables the doctor to economize his information recording and processing but is liable to lead to "missing of data". The validity of such a procedure depends on

whether the doctor will actually complete his picture by testing his hypothesis.
"Massing of data" also influences people's estimation of occurrences especially when the occurrences have a catastrophic character. People respond strongly to emotionally charged events. The media dwell on potential catastrophes and not on successful operations ("fear sells"). The more there is published about the same disease, the more the physician tends to diagnose it. The studies of Tversky & Kahnemann [Tve74] made clear that people tend to estimate weights of events and their frequency by a number of erroneous heuristics. Estimates play an essential role in medical decision making. The validity of most estimates in clinical practice is still matter of debate.

Identification

Of this process Ledley and Lusted tried to give one of the first systematic accounts:

> "First, I obtain the case facts from the patient's history, 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 that the case can be fitted into a definite disease category, or that it may be one of several possible diseases, or else that its exact nature cannot be determined." [Led59]

While this is not a bad description of the way in which the diagnostic process is described in clinico-pathological conferences in medical journals, (CPCs in the New England Journal of Medicine were part of the authors' study material for their article) it has obviously little to do with the daily activities of doctors when seeing a real patient, and the authors admit that it is obviously "greatly simplified". We think Sober 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". [Sob79] According to Pauker et al. [Pau76] this description represents the difference between the expert in practice and the expert as often depicted 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 seems to rely much more heavily upon "guessing", his initial hypothesis typically being based on precious little data. These "guesses" are apparently prompted by pattern 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.

One of the most striking phenomena in the medical problem-solving process is the early generation of (diagnostic) hypotheses. "Specific diagnostic hypotheses were generated often with little more information than presenting complaints" (Kassirer, [Kas78]). This phenomenon was generally found in various different studies of physicians' problem-solving behaviour ([Bar72], [Boh75], [Els78], [Gro85], [Nor88], [Rid89]) and seen as part of the (specific) diagnostic strategy of the physician. This strategy, which is observed in actual experiments, seems to be characterized by the early generation of an hypothesis, in a process similar to that of template matching, followed by a search for symptoms to complete the template (current hypothesis). Obviously, this type of hypothesis should not be confused with a scientific one. It must be viewed as a reflection, an idea, a thought, based on many sources among which are presented evidence, background knowledge, specific experience referring to analogies with familiar pictures and configurations. Since these hypotheses cannot come from the stimulus perceptions, they have to come from the doctors themselves. This type of hypothesis, therefore, involves a personal investment, a sense of commitment. These hypotheses must be viewed as highly personal "patterns" like the ones we know from pattern recognition.
Pattern recognition may serve as the main characteristic of the behaviour of the doctor's strategy. It can be identified as a process of matching a patient's symptom configuration with one that the physician has memorized either from literature or through personal experience [Rid85].

The rationale behind the identification process is to search and find a disease entity, defined within the taxonomy of diseases. Identification will be defined as the allocation or assignment of additional unidentified objects (patient cases) to the correct class, once such classes have been established by prior classification. The definition of classification is the ordering or arrangement of cases into groups or sets on the basis or their relationships. [Sok74] All classifications aim to achieve economy of memory. The world is full of single cases. By grouping numerous individual cases (objects) into a

class, a taxon, the description subsumes the individual description of the objects/cases contained within it. By grouping the cases into a taxon we are already aware that, although not all the patients may yet have been grouped there, eventually all the patients within the taxon measles will get spots. If we combine 100.000 cases of measles into one taxon the average course of the disease can be deduced from such a taxon. The cases are arranged in systems in which the several taxa can be easily named and related to one another. The paramount purpose of classification is to simplify these relationships in such a way that general statements and predictions can be made about classes of objects/cases.

Prognosis is the ultimate goal of medicine. Prognosis permits the choice of the proper treatment and its evaluation. But preceding any prognosis there must be a correct identification of the disease-taxon.

Classifications that describe relationships among cases should generate hypotheses. In fact, the principal scientific justification for establishing classifications is that they are heuristic and that they lead to the statement of a hypothesis that can be tested [Rid89].

But it is at this point that problems arise. As in all biological classifications there is variability between the elements that make up a taxon. As the distinctions between the various disease-taxa are rather vague, the art of diagnostics is very much complicated by this variability. Individual differences may account for much uncertainty. Symptoms which clearly belong to one disease may appear in another disease. Disease of the lungs and disease of the heart may be very similar in their clinical picture but very much different in cause and course. Any medical diagnosis bears the possibility of alternative explanations. It is up to the practising physician to test and judge the various possibilities.

Another obstacle in the identification process is the variability of staging. Illnesses present in various disguises, depending on their stage of development. The first symptoms of a disease may be less distinctive than the symptoms with which it presents in its full-blown stage. The task of a general practitioner is complicated by his position. He is the gateway to more specialistic medical aid and, therefore, one of the first to see the patient. Moreover, a taxonomy of diseases in their first stages does not exist yet. General practitioners have to derive their clinical pictures from the knowledge of the full-blown equivalents as they are taught in medical schools and described in textbooks. General practitioners have to operate in that particular domain where symptoms and signs are vague and variable and disease-descriptions very much overlapping. More than their fellow-phys-

icians in the world of specialized medicine, general practitioners have to rely heavily on their intuition and personal experience.

A commonly heard statement is that general practitioners do not arrive at a specific diagnosis, but usually dwell on such entities as: headache, belly-ache, something wrong in the chest, or describe the illness by its dominant symptom such as: dizziness, diarrhoea, fatigue. But such a description is beneath the mark of any professional management. The goal of every doctor is to heal the patient by means of a treatment of which he knows all the secrets. It means that he has at least some idea about the kind of disease and the patient's immediate future. Any prescription is a justification of this thought. Unfortunately, most of these ideas remain beneath the surface of consciousness and certainly beyond the reach of any recording. This makes the doctor's activities vulnerable and un-reviewable. It does not permit the retracing of his mental steps and these therefore become incorrigible.

A logical stepwise diagnosis may be performed from two starting points: the identification process itself and the process of justification. Starting from the taxonomy of diseases it seems plausible to begin the identification process with a tentative hypothesis which covers a broad class of taxa. Taxonomy itself is hierarchically layered. In a biological taxonomy, for instance that of Linnaeus, several levels of specification can be distinguished. It begins with the two kingdoms (plants and animals, botany and zoology) and follows in increasing detail with phyla, classes, order, families, genera, and species. Attempts in this direction for the medical taxonomy have been proposed by Doroszewski [Dor80] and Ridderikhoff [Rid91]. The latter author defined a hierarchical partition ranging from global descriptions, like "anything wrong with.." to specified bio-pathological states.

Level 1: Human being.
Descriptions referring to global states comprising the total human being. It refers to descriptions like ill/not ill, mental/somatic. serious/not serious.
Level 2: Multiple organ dysfunction.
This class refers to psychopathological states comprising more than one organ system. It includes diseases of body systems or body parts. This class can be typified by examples like: unspecified infectious diseases, unspecified neoplasmata, unspecified ailments of body parts ("something in the chest").
Level 3: Organ system.
This class follows the normal medical distinction of organ systems. For instance: respiratory tract, digestive tract, musculo-skeletal tract.

Level 4: Organ.
This class refers to specified organs and tissue entities. It includes the nearest explanations to the specified disease entity. It can be typified as gall-bladder disease, anaemia, eczema, etc.
Level 5: Cellular.
This class describes a specified biopathological state. It refers to all specific diseases generally described in medical textbooks. It provides an exact circumscription of the lesion and its cause.

The idea that doctors will follow this hierarchical line of reasoning in their diagnostic process, by moving through a number of hypotheses of increasing detail, was tested in a study with 60 general practitioners and 8 internists solving 253 simulated patient cases [Rid86]. No such reasoning was found. The more specific the first hypothesis, the narrower the scope of investigation. In fact, 23% of the early hypotheses belonged to the levels 1+2 (respectively 6% and 17%), whereas 77% pertained to the levels 3+4+5 (respectively 14%, 34%, and 29%). Not every diagnostic process ended up in very specific diagnoses. Of all final diagnoses 8% scored in the levels 1+2 (resp. 1% and 7%) and 82% of the diagnoses were more or less specific (resp. level 3: 11%, level 4: 35%, and level 5: 47%). Less than half of the clinical processes ended with a diagnosis allowing precise prediction. This sums up the general practitioner's dilemma with the imprecise medical taxonomy in his domain. He is more or less groping in the dark; fortunately however, things generally turn out well.
The doctor's process of identification can be conceived as a sequence of "diagnostic guesses", each of which is subjected to justification. The "diagnostic guesses", however, usually follow a single line: a line within one tract. To generate a hypothesis in another tract seemed to require a larger leap of imagination than to look within the same tract. [Rid86]

Justification.

Once we have arrived at some diagnostic ideas, the next step is to make a choice and to justify that choice. Much research, and even more speculation, has gone on into the question of how a doctor selects a diagnosis for a given patient who presents himself, or is presented, to him. It is also one of the key issues in this study. How does the doctor decide what is wrong with the patient; why does he select any particular diagnosis in preference to another which may be objectively just as likely? The diagnosis is a decision which is

of crucial importance for the patient. Does the doctor take this decision optimally, or could it be improved? If the diagnosis is correct, the accepted treatment may be looked up in a medical book. If the diagnosis is wrong, the patient will have to get better without the benefit of a correct treatment, and perhaps despite the treatment for another disease which the patient does not have, and which may well harm him.
But making a correct diagnosis is prime creative challenge; and that there is room for improvement is shown by the fact that many diagnoses made by doctors, in fact many more than suspected by the general public, are wrong. According to Wardel & Wardle, Rosenblatt found only 44% accuracy for cancer of the lung; Gwynne found 33.4% false negative diagnoses in a sample of 1627 patients; Prutting found more than 50% inaccurate diagnoses in various series; in France, in a report of 100 autopsies, in 55.4% of diagnoses were missed ([War78]; Garland reported an accuracy of 44% for myocardial infarction. [Gar60] This is consistent with the findings of Lubsen and v.d. Does. [Doe78] Ridderikhoff found in a population of general practitioners a diagnostic accuracy of 47% for common diseases [Rid86], and de Dombal found an accuracy between 40 and 60% for acute abdominal conditions. [Dom72]
These low numbers may seem more alarming than they are in practice: most health problems presented to the general practitioner are over by next week, whether the patient takes an aspirin, or any other drug, or not. If the condition does not improve sufficiently, the patient will probably be back and the doctor can have another go. But the efficiency of the medical system improves if we "get it right the first time". And of course there do exist cases where the patient comes to serious harm because of a wrong diagnosis.
The quality of medical diagnostics has been subject of many studies. The subject has been approached from different angles of which the most important are: Decision theory, cognitive science and philosophy. Decision theory has already been discussed in the first section of this chapter.

3.4 Problem solving

Before a person can attempt to solve a problem, he must understand or 'assimilate' a description of the problem. The presentation of a complaint does not automatically imply its understanding by the problem-solver.
Cognitive psychology attemps to describe what actually goes on inside the brain when the subject is solving a problem.
There are at least two conditions to this understanding:

a) it is only a problem if it puzzles or worries somebody (assimilation)(Polanyi, [Pol68]);
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 of techniques to solve it. (acquisition) (Mettes, [Met80]).

In order to be able to attack a problem the problem-solver translates the problem into a familiar one and he defines a problem space. The doctor's problem space can be described as his particular medical knowledge relevant to the case and his/her particular way of reasoning in order to solve this clinical problem. As the physician's medical knowledge is partly experiential knowledge and the rules of the problem-solving are more or less idiosyncratic ones, the problem space also includes the prejudices and prejudgements of the doctor. Chapman and Chapman [Cha76] demonstrated how prior expectations can lead to erroneous observations and inferences.
Familiarization of problems may create bias. It means e.g. that the doctor translates the patient's problem into a task he can understand within the context of medical science and health care. For example, many doctors, especially general practitioners, are often confronted with problems beyond their special competence, e.g. problems of a social, educational or psychological nature. Frequently they try to translate these problems to domains in which they feel more confident, e.g. a clinical domain.

Animal studies led to the belief that problems are solved by trial and error. In medical practice there is very little room for trial and error. Perhaps because of this the solving process is often assumed to originate differently.

(1) Sequential: the solver always appears to search sequentially, adding small successive accretions to his store of information about the problem and its possible solution (Simon & Newell, [Sim71]). This process is assumed to take place in normative medical decision making.
(2) Configural (Gestalt psychology): the mind has the tendency to organize and integrate and to perceive situations, including problems, as total structures. In the Gestalt view, the insight that leads to a solution stems from the perception of the requirements of a problem [Sch63].

In Gestalt psychology the perception of the solution of a problem is like the perception of a hidden figure in a puzzle picture. It means that:
- a particular disease pattern existing in the memory is called for;
- between the call and the 'jumping to mind' no (logically) (re)traceable steps can be formulated, nor observed;

- the disease pattern does not necessarily represent an exact replica of the memorized pattern: it shows a number of gross characteristics which mirror the original one. [Rid89]

The implications of this view are decisive. Because of its irreversible and non-retraceable character the solution of the problem is perceived by the problem-solver as an actual discovery. But a discovery is only accredited as such when we believe that it comes from sheer induction, 'out of the blue'. The discovery makes the subject believe that the element is a "real fact" existing outside himself. "Any presumed contact with reality inevitably claims universality" (Polanyi, [Pol68]). While the sequential process can be made explicit by following the steps, the configurational one is unique, '*einmalig*'. This view is corroborated by the finding of the absence of falsifying behaviour in problem-solving processes by physicians. ([Bal80], [Rid89], [Was77]).

3.5 Clinical judgement

Clinical judgement 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. It is suggested that the clinical judgement of a physician is, at the very least, related to his underlying intellectual ability, to the quality of medical education and to the depth of his clinical experience. Knowledge that is gained by experience is not reducible to explicit rules, recipes, or basic principles, however. [Eng79] Even some single data and the notion of frequency evokes the physician's judgement on only a few alternative outcomes. Barrows & Bennet were amazed that in almost all physicians the judgement was reached before all the available data from the patient were obtained. [Bar72] Kleinmuntz demonstrated that data not related to the physician's mental hypothesis or diagnosis are totally forgotten by the physician. [Kle68] Physicians tend to cease further inquiry when a stereotypical pattern has come to mind. The doctor is satisfied with a marginally verified pattern and he feels quite certain about his judgement.
Goldberg comes to the conclusion that clinical judgements are:
(a) rather unreliable;
(b) minimally related to confidence and amount of experience;
(c) relatively unaffected by the amount of information available;
(d) rather low in validity. [Gol68]

Ridderikhoff summarized the findings of cognitive psychology with respect to medical problem solving as follows:

> "The rise and development of the cognitive science has given some insight and understanding of the mental processes involved in problem-solving. A number of theories have been developed ranging from the classic introspection method to the current theory of information processing. In the latter approach the process may be sketched as one of collecting data in a sequential order, interpreting and analyzing these data, and performing some kind of analysis upon them, which may eventually lead, by force of logic, to a valid judgement. However, the process as such was found to be virtually non-existent in routine clinical practice. Cognitive biases, uncertain and unreliable clinical data, sub-optimal and prejudiced data acquisition, early reduction of possible states of affairs, and the communication and interaction of problems between patients and doctors make the thought of a strict and rational process of information-handling rather unrealistic."[Rid93]

3.6 Philosophical approaches

Philosophy has since time immemorial tried to examine and define how we know something, and how we can proceed to something else, given that we know it. And how we are able to know whether the something else represents the 'Truth', or a real 'law of Nature'. "The aim of science is to find satisfactory explanation of whatever strikes us as being in need of explanation" [Pop83]. It is assumed that valid inference and logical reasoning from theories or from observations will help us find the Truth. To this end two main ways of reasoning have been defined: the inductive and the deductive way. Another method has been developed during the previous century and was called after Whewell(1837): hypothetico-deductive. [Whe37]

3.6.1 Inductive reasoning

Inductive reasoning starts from unbiased observation of Nature which, by force of our creative facilities, leads to a hypothesis which is then verified by the observation of Nature. This leads to what Braithwaite [Bra68] calls "a circular reasoning": verifying hypotheses based on some evidence with evidence derived, or at least related, from this hypothesis. It can be illustrated by the following dialogue: "Why is this person so pale?" - "Because

he is ill." - "By what evidence can you support your statement that the person is ill?" - "Oh, can't you see he is pale? And is that not always the case when people are ill?" [Rid89].
Moreover, unbiased observation is seriously questioned by many authors. It is best illustrated by a Popperian anecdote: "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 the strong urge to ask: 'What do you want me to observe'?" [Joh73]
Inductive reasoning concludes from detailed observations to general statements. If we have seen 5 white swans and we observe another white specimen we believe in the statement that all swans are white. If we have seen 5 patients with cough and fever, all suffering from pneumonia, and another patient presents himself with cough and fever we believe the patient to suffer from pneumonia.
But as Russell states:"however many cases of coherence between A and B we have observed, it cannot be a reason for their coherence in the future. Therefore, induction by simple enumeration cannot be a valid way of reasoning". [Rus56]
Inductive thinking, as a generalization from specific instances to higher rules, in the words of Medawar

> "cannot be a logically rigorous process. It cannot (as deduction can, properly executed) lead us with certainty to the truth. Mill believed it could do so, but John Venn and C.S. Pierce and others flatly disagreed with him, and it is their opinion which has prevailed. I shall waste no time attacking a position that is no longer defended. No process of reasoning whatsoever can, with logical certainty, enlarge the empirical content of the statements out of which it issues." [Med67]

Although the disadvantages and drawbacks of the inductive method as a method of science were broadly advocated, its overall existence cannot be negated. It is practised as an ordinary way of reasoning in daily life. People usually infer general statements from particular elements or single observations. We are so accustomed to this type of reasoning that it happens without our noticing. Inductive reasoning is a way of surviving in a world full of information and uncertainty. Rephrasing Bacon, we could say: "The last thing anyone would be likely to entertain is a state of uncertainty".

In the uncertain world of medical practice the doctor must be creative as well as economical with his time. Pattern recognition and a parsimonious testing of mental ideas are instruments for quick judgement. "This judgement

may very well be true and may 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" (Brehmer, [Bre80]). In the inductive reasoning the diagnosis is based on the analogy of the clinical picture of the patient with the one remembered from experience or learned from the classical taxonomy of diseases (see chapter 1). The elusive way in which the analogy is 'sensed' from the (possibly biased) observation, combined with the extremely perfunctory testing of the 'clinical picture' forms a major obstacle to feedback, evaluation and retracing of mental steps in the problem-solving process.

The inductive process of medical problem-solving may be sketched as one of successive "diagnostic guesses"; guesses based on presumed similarity of (parts of) the clinical picture of the patient with a memorized one. As was mentioned before, the memorized picture is mainly based on experience [Koc83] and is therefore a personal one which is expressed in individual terminology. It is also very personal in so far as people adore their own brainchild and are very reluctant to reject it.
The testing procedure in inductive reasoning is guided by the mental hypothesis. Informational cues and testing elements cannot be distinguished in the clinical process (Elstein et al., [Els78]). The acquisition of symptoms which test the generated hypothesis are used to generate another hypothesis in reverse [Rid89]. In his study of the problem-solving behaviour of general practitioners Ridderikhoff found that 75% of the acquired symptoms were directly related to a hypothesis. But if a hypothesis is the dominant force behind the acquisition of patient information, alternative explanations of the collected data will compete less successfully for the attention of the problem-solver. As had been already mentioned by Kleinmuntz [Kle68], in this study it was also found that doctors completely "forgot" non-hypothesis-related data; data which might have served the doctor in finding alternative explanations for the presented problem.

Sage [Sag81]) listed some of the concepts induction-led people employ. The concepts:

(a) are drawn from personal experience;
(b) involve elementary classification and generalization concerning tangible and familiar thoughts;
(c) involve direct cause and effect relationships, typically in simple two-variable situations;

(d) can be taught or understood by analogy, algorithms, effects, standard operating policy, or recipe;
(e) are 'closed' in the sense of not demanding exploration of possibilities outside the known environment of the perceived data.

Retraceability is a chief condition of a verifiable diagnostic judgement. Another doctor, or the same doctor at another time, should be able to retrace the reasoning which led him to his conclusion. If this is not possible, no learning from experience can take place. This quality of retraceability can only be found in two other ways of reasoning: deductive and hypothetico-deductive. The 'reasoning' of the DDSS, conforms most closely to the inductive variety, just like the doctor's. It is our philosophy to stick as closely as possible to the usual deeds and thinking of the practising doctor. The system is eminently retraceable, however: consecutive runs with the same data are guaranteed to produce the same result, and the reason why this diagnosis is suggested and not another may be immediately inspected.

3.6.2 Deductive reasoning

Deductive reasoning is a different method of inference. Instead of making general statements (diagnoses) plausible from the observed evidence, the deductive reasoning starts from the opposite side. "If a general statement or theory is true, then I must find some evidence to justify the hypothesis by simple testing". Deductive reasoning may start from bold and brave assertions and proceeds by a strict logical process of detailing and specifying inference. Deductive reasoning is usually exemplified by a syllogism:

All Greeks are mortal.
Hippocrates is a Greek.

Hippocrates is mortal

The reasoning is one of the "if ... then" type. It is sequentially and logically constructed so that the steps of the mental processes can be retraced. A more detailed argument must unequivocally follow from the preceding one which in its turn comes after a more generally stated argument. The most finely-grained argument (hypothesis) is open to testing by a controlled experiment. This particular hypothesis can either be verified or rejected. As a consequence of its logical reasoning in argumentation-steps [Sad74] the result of

the testing is valid for the complete construction, including the preceding general statement. Any mode of reasoning which ends with more than one hypothesis left (as some other plausible explanation) is contradictory to deductive inference.
An example of such an inference is given in the following:

I) 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;

II) The observed evidence (symptoms and signs) can only be explained by explanation B of the A origin. The disease can only be explained by e.g. an inflammation of the lungs;

IIIa) In case of explanation B there must be a C test of D level confirmative of the B, e.g. demonstration of the presence of pneumococci in the sputum of the patient;

IIIb) In case of explanation B there must be a C test of E level (E!=D) refutative of B, e.g., when microorganisms of the type Bact. pneumococci are demonstrated to be present there cannot be an increasing antibody titre for a different microorganism.

Tested against this particular case:

*f*1: In all cases of disease A the level D of test C is within the range R; e.g. over a certain number of microorganisms;

g1: In this case the level D of the test C is within the range R. (from [Rid89])

Variability of diseases and their causal relationships often prevent a clear deductive reasoning in practical medicine. Although this method should be a preferred one in scientific reasoning, the limitations of medical knowledge and the practical situations in which doctors find themselves interfere with such ambition.

Hypothetico-deductive reasoning

With the theory of hypothetico-deductive reasoning a bridge between the two antithetical conceptions seemed to have been built. Originating in the 19th century the method combines inductive and deductive reasoning into one single process. It starts with the inductive act of generating hypotheses from observation and in the second part justifies these hypotheses by deductive inference. This mode of reasoning is fairly generally assumed to be used by

doctors, although this has been seriously questioned by others (McCormack, [McC86])

Medawar listed a number of positive features of this method.

1) A clear distinction is made between discovery (a creative thought) and justification or proof;
2) The hypothetico-deductive scheme provides a theory with a special incentive. Our observations no longer range over the universe of observables; they are confined to those that have bearing on the hypothesis under investigation;
3) It allows also for the continual rectification of the hypotheses by the process of negative feedback;
4) Error is simply explained; scientific error is now an ordinary part of human fallibility;
5) The hypothetico-deductive scheme gives due weight to the critical purposes of experimentation: we carry out experiments more often to discriminate between possibilities than to enlarge the stockpile of factual information.[Med69]

It is our contention that a severe limitation of the search for hypotheses takes place. The search is limited to what is previously within our experience, which then determines the relevance which we ascribe to characteristics of the case at hand. We are not so sure that a restricted hypothesis generation does not influence the process by which a conclusion is reached. When this 'act of proof' is assumed to follow the ritual of deductive inference, the falsification of the generated idea is the logical implication. However, this presumes that our observations will be made with the goal of falsifying the hypothesis: either we may come to think that our observations may themselves have been faulty, or that they may have been made against a background of misconceptions, or our experiment(s) may have been badly designed. It is very unlikely that individuals would reason in this way. We follow Harre [Har72] in his statement that in essence the hypothetico-deductive method is an inductive method.

Several authors, however, believe the hypothetico-deductive method to be "a nearly universal characteristic of human thinking in complex, poorly defined environments" (Elstein et al, [Els78]). Needless to say, these authors consider medicine to be such an environment. These authors, however, have not provided a general scheme for such a conviction. Conversely, McCormack considers the claim of a hypothetico-deductive method in

clinical practice to be a myth to season the clinical decision process with a scientific flavour.[McC86]
When looking for the landmarks of the hypothetico-deductive method, Ridderikhoff found no such cases in his material. [Rid86]

3.7 Probabilistic approaches

The number of techniques to analyse data has increased enormously over the past decades. These techniques are almost all based on 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 probability in real-world situations? When we try to answer this question we reach the foundations of probability theory and, as in many sciences, it is precisely here that differences in opinion appear. The foundations of this concept have been object of much discussion [Rid89]. Different interpretations have resulted in different axiomatic systems.
The usual distinction in probability theory is:

I) objective probability; and
II) subjective probability (also called: inductive probability).

The former meaning refers to the mathematical type of probability, or statistical probability. In its simplest form it can defined as "the relative frequency of an event 'a' occurring within a reference class (also called 'population') 'b'; in formula:

$$p(\mathrm{a},\mathrm{b}) = r$$

Sometimes another formula is used:

$$p(\mathrm{a}) = r$$

or, the (absolute probability) of 'a' being 'r', also called the 'prior' or 'a priori' probability, when a specific reference class is taken as 'understood'.
For this type of probability there is no place in general practice. For there is no point in saying "measles is a common disease" unless you specify "in childhood", and even then not every child with symptoms similar to those of measles has the disease and not all analogous symptoms refer exclusively to

measles. The expression p(D), the 'absolute' probability of disease D, cannot have any relevance within the framework of objective (statistical) probability, unless a conditional reference class is taken as understood or will subsequently be added. In daily medical practice such a reference class cannot be adequately defined, hence the objective type of probabilistic reasoning cannot be a basis for the justification of certain hypotheses in routine practice.

However, the notion of randomness (patients and diseases present themselves in a completely haphazard way) and uncertainty (as a consequence of any problem-solving process) led several scientists to believe that the decision-making process has something to do with probability. Given a prior probability of a certain hypothesis and the weights of the evidence related to this hypothesis we will eventually be led to some notion of strength (posterior probability) for the decision. But this notion of probability is not identical to the definition of objective probability. It refers to the other form of probability: subjective or inductive probability. Inductive probability is a measure of the strength of support given to a hypothesis (h) by the evidence (e). In the notation of Carnap [Car60] it takes the form

$$C\ (h,e) = r$$

which is the degree of confirmation (C) of 'h' on the basis of 'e'. The hypothesis may be any statement of an event, e.g. a diagnosis. 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, our inductive thinking in everyday affairs and in science. It expresses in quantitative terms our confidence in the outcome of a particular process. This kind of probability, which must not be confused with the objective probability of statistics, should be understood in the context of a defined virtual sequence of events. We have to assume an orderly trend which is not assumed to be present in objective probability. Inductive probability is a way of judging hypotheses concerning unknown events. It is guided by our knowledge of observed events. The greater the relative frequency of an event the more plausible an assumed hypothesis. It seems plausible in daily practice, that from observation of what has in the past been the consequence of a certain course of action, one may make a judgement as to what is likely to be the consequence of that course of action another time. Our confidence that a certain therapy will work in a present

case of certain disease is higher the more frequently it has worked in past cases. But this inference is logically false to the facts. The outcomes of the various types of probability (objective or frequentistic and inductive) are not interchangeable as was shown by Feller (see [Pop83]). It disregards sources of evidence unrelated to the particular hypothesis; and assumes orderly trends where real life provides us with randomness and sometimes chaos. This means that replacing probability estimates from research in (diseased) populations into a formula intended for subjective (inductive) probability, or vice versa, is very hazardous. It creates one of the more difficult problems to be solved in medical decision-making. If one is not aware of differences between the two types of probability, decision-making can produce outcomes far removed from reality, apart from the variations due to the personal knowledge and backgrounds of the physicians.

Various flaws and weaknesses may occur in the diagnostic process of the practising physician. Unlike the structured and well-prepared discussions of medical judgements as described in the clinico-pathological conferences in various medical journals, the practising doctor is unpreparedly confronted with a large number of patients with a multitude of ailments. Eddy & Clanton propose a multistep model based on these clinicopathological conferences, but they must admit that "It is impossible to state precisely the extent to which the diagnostic methods displayed in CPCs are used in practice. The CPC is an artificial forum." [Edd82] Besides, the general practitioner is the one who has to decide about the least documented diseases in their vaguest forms. The worrying statistics of diagnostic accuracy, however, must be a good reason to strive for improvement. If modern equipment and methods can present a helping hand, by all means let us grasp it.
This is the raison d'être of the Diagnostic Decision Support System. It endeavours to help the practising physician in the various stages of the diagnostic process: observation, validation, identification and justification, each with its specific difficulties. The next chapter will give an outline of the details.

Chapter 4. Problem formulation

4.1 Introduction

In this chapter we will outline the problems we face when designing a system to assist the doctor, and further we shall enumerate the main properties of a well-designed system and its expected benefits for the physician.

The job of the doctor consists for a large part of taking decisions concerning his patients. In this process, the doctor uses among other things an amount of patient information that should be accessible at all times. As a problem-solver and a data processor, man is restricted by factors such as:

- a finite memory
- an imperfect registration
- a subconscious preference for current and recent events
- a limited rationality caused by (over)simplification of reality
- a personal, unique knowledge
- stereotyped patterns of taking actions
- a number of prejudices which may blur observations

Especially in the work of the general practitioner, as opposed to clinical medicine, these limitations are hard to avoid because the doctor is hardly able to use options such as consulting colleagues, reading all new literature, taking time to reflect before acting, etc, when solving a patient's problem. Nevertheless, the patient demands from his doctor at all times a responsible and reliable decision, in spite of the above limitations.
Another problem is that doctors tend to be rather disorganized. The registration of their patients and contacts is usually incomplete (in the sense that another doctor probably would not be able to do very much with the records), and their ability to create chaos out of order is notorious. The reason for this may be that compared to his patient's problems, administration seems quite unimportant to the doctor. This lack of organization, however, may cause problems when a new doctor takes over the patient, or when a replacement sees the patient in the doctor's absence. Other problems arise when the patient has not visited the doctor for a while, and the doctor cannot remember exactly the medical history of this patient, which might be

of great importance for the current diagnosis. What is needed is a system that will:

a) keep an administrative record of the 'raw' patient data, i.e. his or her complaints, symptoms, signs, test results, etc, stored in a structured and uniform way, and
b) be able to use these structured data to aid the doctor in arriving at a diagnosis.

4.2 Problem definitions and requirements.

From the earliest beginnings of the project, three prime requirements have been set for the system to be developed.

1) **It should be used simultaneously with the patient-doctor contact.**

This starting-point is in contrast to the traditional state of affairs: most diagnostic systems are meant to be used especially or exclusively in difficult cases, and after the patient has been seen. But most doctors struggle particularly with the so-called 'easy' cases: cases which present little hard evidence, cases hardly mentioned in the textbook literature, cases easily passed over as negligible. It is not so much the rare disease which is difficult, but rather the vague and nebulous presentation, which is very frequently seen in general practice. Consequently, separate diagnostic support programmes are hardly used at all, as most doctors do not even take the time to stand up and consult a book from the shelves in their consultation room. Many errors are made, and perhaps especially in those cases which at first sight seem to be easy and not to warrant special attention. This simultaneity of the doctor-patient contact with the use of the DDSS is an essential part of our strategy to capture the patient data at the earliest possible moment, before rationalization, reasoning, and memory recall errors of the doctor have a chance to distort them.

2) **It should be a logical extension of the problem-solving process of the doctor, not a replacement of it.**

This means that the system must follow the lines of thought described in chapter 3. There is no sense in pushing people into a type of behaviour which does not come naturally. The system is meant to adapt itself to the

usual method of diagnosis as has actually been observed in general practice [Rid86].

3) **It should include a full administrative system, replacing the paper record and obviating the need for it.**

In the absence of an overall theory for such requirements, the basic principles had to be developed within the proper framework of the health care system, or in our case the general practice.

From this, we derived the following requirements:

- *Because of the position of family medicine in health care, the system should cover very broad areas of medical diagnostics.*

- *The construction of a medical taxonomy for general practice being in its infancy, the system has to overcome the problems of varying terminology and nomenclature.*

- *The system should provide the possibility to detect and flag errors in reasoning and judgement.*

Through a number of successive refinement stages, these have been the guiding principles from which a large number of other consequences naturally follow, such as the required simplicity of the user-interface and real-time performance, which means that results are available immediately without the doctor having to wait for them.

Our goal in the widest sense is to improve the prognosis for our patients. But the prognosis can only be accurate when an accurate diagnosis has been made, and an accurate diagnosis depends on a large number of factors, some of which can be influenced, while others cannot. We want to improve accuracy, but we cannot change human nature: we cannot fundamentally alter the reasoning modes of the doctor. We must take into account the circumstances under which the doctor has to work: any system that increases the time needed for a patient will be very hard to defend, even if substantial benefits can be shown. Our formulation of the main problem then becomes:

How can we best design and use the DDSS to improve diagnostic accuracy of doctors without distracting from their overall efficiency?

4.3 Further analysis of the problem statement

From this overall question two concepts may be selected: Use and Design.
Use concerns the user and his environment, clinical as well as situational. It also concerns the patient. We may call this the clinical part.
Design encompasses the software in connection with the equipment employed; it is mainly the technical part. We shall elaborate on these two themes in the next sections: 1) clinical and 2) technical aspects. The clinical aspects pertain to the user's (doctor's) view of the system, while the technical aspect represents more the view of the software engineer and programmer. Inevitably, there is some overlap between the two views.

4.3.1 Clinical aspects

From the doctor's point of view, we want to assist the diagnostic decision-making process in four main areas: 1) data acquisition, 2) data storage and retrieval, 3) the reasoning process, and 4) the knowledge base used. In addition, there are a number of 5) constraints, deriving from the nature of clinical practice.

4.3.1.1 Data acquisition

- *The system should avoid bias of observation and perception in data acquisition.*

In our view data acquisition is composed of observation and validation. There are two approaches:

1) A normative approach such as in clinical decision making, protocols, etc. This means a particular way of sequential data acquisition and a structured manner of handling data in connection with frequency distributions of the relevant data; and
2) A human, subjective approach in the realization that data acquisition is subject to several types of error as described in the preceding chapter.

As stated before, we have opted for the human approach, assisting rather than directing the practicing physician. Each of the error types already

mentioned in chapter 3 has to be faced and met in the system as far as possible.

The problem of the validity of the data (reproducibility, precision and reliability) is countered in the system by the detailed structuring of the data entries. For this a Symptom Coding System (SCS) was developed, to be described in more detail in the next chapter. Use of this system at least avoids misunderstanding and promotes the use of standard terminology and therefore reproducibility. Precision promotes accuracy which, in its turn, might be tested by retraceability. In normal practice, observation and validation are both guided by the diagnostic hypothesis currently under consideration; the results of the observation are formulated in widely varying and unpredictable terms.

1) A very large number of possible observations (symptoms, signs, lab results) is defined in the Symptom Coding System (SCS) menu structure, thus guiding the doctor's attention and enabling him to observe items which might otherwise have escaped his notice. In our expectation, doctors will query their patients wider when confronted with the DDSS menu system. They are also expected to be able to encode all their observations within the SCS.
2) The validity is enhanced by the level of detail, and there is a greater level of detail offered by the SCS is than usually required in general practice. The availability of this accuracy should stimulate the doctor to probe deeper with his questions, and so to get a more detailed picture of the patient.
3) The terminology used for the description of the patient's symptoms is enforced by the menu structure of the SCS, and is therefore invariant over time for the same doctor and also between doctors. Doctors should be able to handle the standardized terminology which may (and indeed certainly will) differ somewhat from their own, without too much trouble.
4) If the doctor is of the opinion that the level of description offered by the SCS is not enough in specific cases, he has the option of attaching a note to any symptom, and add the required accuracy in the form of a free-text field. This should only rarely be necessary if the level of specification of symptoms is adequate.

- *Bias in validation by letting one symptom weigh heavier than another should be avoided.*

In numerical taxonomy, classes are made up of characteristics which are analysed mathematically to yield groups, the members of which resemble each other more than members of other groups. For this analysis it is important that no single characteristic is weighed more heavily than any other; 'all citizens are equal' to the algorithm. In medicine, this is a highly unusual approach, because every doctor instinctively assigns much more importance to some symptoms than to other ones. However, as in our case there is really no satisfactory pre-existing taxonomy, we will have to construct one ourselves. Therefore we will start from the premise that all items are to be weighed equally.

4.3.1.2 Storage and retrieval of data

An accurate medical record is vital for the retraceability of the diagnostic thought process. Memory is usually short and unreliable and paper records so summary that they do not serve to reconstruct the reasoning of the doctor, but only give the patient's initial complaint and the conclusion. The ideal record would retain all information the doctor has used in making his decision, and would be easy to recall and review. The DDSS should be designed in such a way that the doctor can enter every symptom, sign or complaint as he finds it, without being biased toward the final diagnosis. This is important, because it makes it easier to reconsider the diagnosis if necessary. The information which did not fit this earlier diagnosis will still be there in the record, and may serve as a starting point for reconsideration, while it would probably not even have been entered into a paper record. It facilitates a critical review of the doctor's decisions - by others, but first and foremost by the doctor himself. Quoting McIntyre and Popper:

> "Doctors are expected to learn from their experiences, and from their earliest days medical students are exhorted to learn from their mistakes. To learn only from one's own mistakes would be a slow and painful process, and unnecessarily costly to one's patients. Experiences need to be pooled so that doctors may also learn from the errors of others. [...] But errors need to be recorded and to be analysed if we are to discover why they occurred and how they could have been prevented."[McI83]

An accessible record, and especially an electronic one, makes inspection and retraceability possible to all who want to learn (and teach) from experience. The medical record present in the DDSS allows such a possibility. Absolutely necessary for the practical use of this posibility is the option to anonymise

the record, i.e. to strip the personal and private elements from the case record. To this end the administrative items such as name, addres, type of insurance, etc, are completely separate from the typically medical parts in the DDSS. These latter elements are exclusively included in the SCS. For compatibility reasons the administrative part of the system follows the standard model which has been specified by the NHG, the Dutch College of General Practitioners.

The SCS as a recording system only applies to symptoms and other medically relevant information about patients; there are a number of administrative items, such as patient's name, address, type of insurance etc, which have already been implemented in many systems for the general practitioners currently on the market. The DDSS is not a full implementation of the NHG standard, but the subset it includes does conform to its structure. Given the time and resources, a full implementation can undoubtedly be made.
In summary:

- *The medical record present in the DDSS should make errors visible (and retraceable).*

4.3.1.3 Reasoning processes

As we have seen in chapter 3, besides the observation and validation mentioned above, the reasoning process proper is especially concerned with the problems of identification (of the disease sought) and justification (increasing the personal certainty for the correctness of this hypothesis).
Hypotheses within the reasoning process must be viewed as reasonable explanations for the acquired evidence. In problem-solving the doctor is assumed to consider the various explanations which arise from the evidence so far. By stimulating medical students explicitly to consider several explanations, so called differential diagnoses (DD), they are challenged to reconsider all the available possibilities they can find in their repertoire. Regrettably, this potential quickly erodes during the years in practice.

Doctors use a 'configurational' mode of reasoning which is especially prone to errors of incompleteness, or oversight. Whereas a sequential process can be made explicit by following the preceding steps, a configurational one is implicit and therefore not retraceable. The sudden 'recognition' of a particular disease profile makes the doctor fail to consider alternative diagnostic hypotheses. Doctors also let their observation be guided by their hypothesis; if the hypothesis is rejected, the symptoms

already acquired to support it may easily be discarded as well. Validation in medical practice is reduced to finding concordant symptoms for the current diagnostic hypothesis. This search for concordant symptoms may even go so far that the doctor fabricates 'evidence' which the patient has not given, assuming things to be present without having asked for them or distorting patient's answers to fit his own ideas.

The DDSS should be designed to meet the problems in the reasoning process in the following ways:

1) To offer diagnostic hypotheses on the basis of the symptoms entered, thus reducing the probability of oversight; doctors should consider more hypotheses in their decision making after having seen the system's DD.
2) To offer an objective measure for the likelihood of a diagnosis, on the basis of the overlap between symptoms of the patient and symptoms belonging to the disease. Doctors should ask for more symptoms to confirm the 'diagnosis' = currently considered diagnostic hypothesis, if they see the degree of concordance between their hypothesis and the symptoms.
3) To use all symptoms entered in the evaluation of every hypothesis, not just the ones only gathered to support the current hypothesis. Doctors should reconsider a conclusion that does not explain many or most of the symptoms.
4) To offer the option of looking up symptoms for the suggested hypotheses, to assist the justification process, and the search for evidence. Doctors should check missing symptoms in their configuration.
5) To guard against unwarranted interpretation or even fabrication of symptoms by suggesting detailed options for symptoms. The doctor should be made aware that there are many ways of stating a symptom and should be encouraged to inquire closely to distinguish between them.
6) To improve the reliability of the reasoning: the same data should give rise to the same diagnostic hypotheses.
7) To improve the retraceability of the process by allowing to retrace step by step the sequences of hypotheses and symptoms.

In the experiment to be performed, each of these hypotheses and questions should be reviewed.

4.3.1.4 The knowledge base

In order to offer more than purely administrative support,

> *the system should provide an extensive medical knowledge base that complies with the doctor's knowledge domain, the structure of which reflects the natural progress of his decision-making process.*

In order to obtain a suitable knowledge-base that is as complete as possible and complies with practice, knowledge must be used from reliable source(s) within the appropriate domain of health care, e.g. from a group of practising general practitioners.
The way in which this knowledge is obtained (knowledge engineering) is of essential importance to the maximum attainable quality of the results. This presents the following problem:
The individual doctor's medical knowledge is far removed from the image of the uniform medical knowledge, shared by all doctors. Instead of being uniform, medical knowledge is highly personal and is among many other things dependent on and derived from personal experience, recent successes and failures, and the field in which the doctor works. "The first thing you think of is the last thing you missed". The personal character of this knowledge creates another problem: in practice, each doctor seems to have a terminology of his own, and in many cases, he has his own interpretation of the medical terms used. If one should ask ten doctors for a medical description of a certain diagnosis, it would appear that there exist ten different descriptions of the disease. To make comparison possible, a set of standard terms must be used for symptoms, signs, complaints, test, diseases, etc.

> *In order to develop algorithms that make sure that when determining a diagnosis, one or more comparable diagnoses are selected from the knowledge base in which similar symptoms and signs appear, any terminology and nomenclature must be uniform.*

Our task is also to construct a knowledge base which is tailored to the specific field within the health care system the user has to work in.

The DDSS was designed to minimize differences in terminology, while at the same time offering the option to create specific knowledge bases for a specific medical field, such as general practice. The DDSS can offer a standardized database of templates for diseases in the presentations which

usually occur in the field where the system is being used. One absolute prerequisite for any such undertaking is the standardization of terminology; this is enforced by the use of the SCS. Diseases can be defined in terms of their observable symptoms, which may differ according to the position of the doctor who sees the patient: the general practitioner may have to rely on less obvious patterns for his diagnosis than a specialist. Such a target-specific knowledge base should ideally be constructed from the collective experience of the target physicians themselves. The DDSS was designed with the idea that a large number of medical records gathered by different doctors should be pooled and extracted to produce such a specific knowledge base. For obvious reasons, this knowledge base could not be constructed before the first try-out of the system. Instead, we had to make do with a simple ad-hoc database constructed from medical textbooks.

4.3.1.5 Constraints of clinical practice

As it is the doctor who must use the system, not his secretary or assistant, and because the doctor must use the system *during* the consultation and not afterwards, the system should distract the doctor as little as humanly (or mechanically...) possible, in order not to disturb the doctor-patient contact, and the doctor's concentration.

4.3.1.5.1 Time

The system should not interfere with the consultation time.

The system should not require much time to use, and there should be clear immediate practical benefits to its use if it is to be accepted at all. The doctor should be distracted from his interaction with the patient as little as possible. This is a very hard constraint to fulfil. To the inexperienced computer user, having to use a computer programme at all is a formidable distraction in itself, and only long familiarity with the system can really reduce this extra load on the attention of the doctor. In the experiment the consultation time is compared with the time usually spent in new case problems.
Using the system should be so easy as to become largely sub-conscious over time, completely integrated in the problem-solving process. This degree of expertness is obviously not attainable in the sort of experiment we can perform for a first test. Nevertheless, it may be possible to draw some conclusions about how easy or difficult it would be to reach it. Ideally,

doctors would not find the use of the system disturbing. We cannot ask the patients yet.

4.3.1.5.2 Ease of use

As the psychological contact with the patient during the consultation is of extreme importance, the machinery should not interfere with this process. The DDSS is conceived as an equipment which may even stimulate the exchange of information between patient and doctor. When both people have become somewhat familiar with the various screens (which is quite possible with the frequent contacts in general practice), they can discuss and specify the particular symptoms and signs which the patient actually experiences. A basic principle of the DDSS is that observations are processed without interpretations. This contributes to the validity of the entered data and may, in reflection, serve to remind and to test cues. It is essential for the doctor to involve the patient in the process and to give the patient the responsibilities (s)he deserves. Visualisation of what is sometimes intuitively perceived may make vague and nebulous feelings of the patient suddenly explicit and recognisable. The involvement of the patient in the data gathering process may certainly help to decrease differences in observations and perceptions in doctor-patient consultations. As the collection of patient information is not determined by special circumstances and conceptions, DDSS can be adapted to any situation in any environment.
As the use of colour may help to find the right path in the right context, the DDSS should use this medium unobtrusively.
The ultimate goal for data-entry perhaps is voice-recognition. As this technology is still in its infancy, we chose the next easiest method: a three-buttoned mouse with fixed menu data entries.

The DDSS should not detract from the doctor-patient contact.

Needless to say, the DDSS should not require so much attention from the doctor that his ability to communicate with the patient, both on a verbal and nonverbal level, suffers.

4.3.1.5.3 Ease of documentation and information exchange

In an era of openness and access to data, the electronic recording of patient data can be a very useful method. Patients' requests for inspection of medical records and other relevant documents can be easily complied with.

Information which is relevant to particular purposes, e.g. for specialists, can easily be transferred electronically. If anonymised, cases can be pooled in knowledge bases for supportive, or teaching and learning functions. Consultation among physicians is very much facilitated by this electronic medium and by the use of standard terminology. Doctors do not have to rely on their scanty notes and summaries of the case but can peruse the complete story of the patient. This may result in a better understanding of the patient and more accurate diagnosis.

The system must be private to doctor and patient but allow easy communication when necessary.

4.3.1.5.4 Learnability

Besides being easy to use, it should be very easy to learn. Ideally, users without any previous computer experience should be able to sit down and use the system without having read the manual. This is probably not feasible in practice, but for a user with some general computer experience the manual should not really be necessary in order to be able to use the system. The teaching programme and the learning process for the test subjects should be scrutinized for points which appear to be difficult to grasp.

4.3.1.5.5 User documentation

The system should be self-documenting as much as possible, but there should also be a printed manual, in which any difficulties the user might encounter are explained. Use of the manual and assistance from the observers of the experiment should be carefully monitored.
The experiment should yield detailed information about all these aspects of the system and its use.

4.3.2 Technical aspects

By this we mean the "information technology" side of the system. The software engineer faces a number of problems with regard to the user. These problems may be summarised as follows: 1) Ease of learning; 2) Ease of use; 3) Reliability and robustness; 4) Tailoring the system to the user's needs; 5) Productivity.
We shall succinctly treat each of these items.

4.3.2.1 Ease of learning.

Rubinstein and Hersh claim a "Ten-Minute Rule", which is that a beginning user should be able to learn how to use a system in just ten minutes [Rub87]. Part of this exaggeration is the presupposition that users will bring a considerable knowledge they already have of their own field. The ten-minute rule emphasises the important condition that architects of systems should be aware of what is helpful for the user and know how to adapt the system to the customary tasks in daily life.
The DDSS is based on direct observation of practising doctors, so that the user will find a system which (s)he is more or less familiar with.
Learnability has two aspects: a) ease of learning to handle the hardware; and b) ease of learning of the software.
The learnability of the hardware is directly connected with ease of use (see next section).

The learnability of the software is a matter of
- having an overview of the organisation of the SCS (including terminology);
- understanding the concept of a 'tree' - structure; and
- understanding the screen information.

This leads to the following problem:

To create a system in which most doctors will easily find the right data-entry for a particular observation despite the varying medical nomenclature employed in medical practice.

4.3.2.2 Ease of use

A number of aspects may be distinguished: a) user-friendly equipment; b) user-friendly interface; c) presentation and documentation.

a) Developments in the computer industry largely determine the choice of equipment. The dominant position of (Microsoft) DOS operating system computers almost forces our choice here. Most general practitioners who already possess an information system make use of this type of computer. For reasons of convenience therefore the DDSS was developed for this type of computer. Although the personal computers are not traditionally thought of as user-friendly, with the addition of peripheral equipment

such as a mouse and colour screens user-friendly programming has become a distinct possibility.

b) User-friendliness in the software is implemented by simple start/restart functions, menu-oriented user interface, uncluttered screens, use of colour, and on-line help functions. Simple communication between user and machinery is a prerequisite for optimal employment of the system in daily practice. This can be reached by adhering to the following rules:

- The user-interface should be extremely simple. The most immediate consequence of this is that the user should not have to type in any more data than is strictly necessary. This practically forces the system to be menu-driven.
- The user should be able to use the shortest path to perform the things he would like to do, and he should feel in control of the operations.
- He should have an idea of what is going on inside the programme, which need not be technically accurate, but should be consistent with its behaviour (the "myth" of the programme).
- The function of keystrokes should be consistent within the programme, and where possible consistent with common usage in other programmes.
- It should be possible to summon help at any point in the programme, and the help given should be relevant to the context in which it is called.
- The system should guard against user error, without becoming patronizing. Accidental deletion of data should not be possible. Actions associated with data loss should be confirmed.

c) The presentation of the system and its various parts must be connected with the intuitive feelings of the user. The various categories and screens must follow the customary way of the physician's problem-solving. It helps the user very much to be fully oriented as to what is going on during the process which strengthens his/her idea of full grasp.
- Screens should be uncluttered, giving only the information the user needs at that point in the process. Colour and other graphic devices may be used to support the presentation. Colour especially is a powerful director of attention, and background colour may be used to reinforce unconsciously the orientation in the menu tree.

User documentation has already been mentioned in the clinical section.

In short,

the DDSS should be designed in connection with equipment already in use with general practitioners and should adapt to rules customary in medical problem-solving.

4.3.2.3 Reliability and robustness.

The reliability of the system should be high, for nothing is more aggravating than the breakdown of a system upon which the doctor is dependent without a fall-back system in place. The programme should be tested and be as bug-free as possible. Ideally, user error should never lead to a system crash. If an error-condition occurs, the software should silently and automatically recover and allow the user to proceed. If the error is unrecoverable, the system should give clear diagnostics which are understandable to the user and pinpoint the source of the error to the programmer. There should be clear procedures for error recovery and restarting the system if the user ever encounters a fatal error. Errors encountered during the first practical testing of the system will give an indication of places where improvement is desirable.

- The system should perform correctly, giving reproducible and accurate results. Data storage, modification, and retrieval should be possible, while accidental data loss should be guarded against. Database integrity and consistency should be guaranteed. Any errors in this area should be recorded and analysed.
- The system must perform within a specific time, depending on the application. The time needed in typical and worst-case applications should be measured.
- The system should recover gracefully from error conditions, whether the error is made by the user or present in the programme itself. Any difficulties in this regard should be logged and analysed.
- The system should perform on a specific hardware platform, the specifications of which form part of the system. The hardware should be easily and cheaply available, and the system should be tested on a number of different machines.

The programme and its infrastructure should, in case of problems, not leave the user with a host of subsequent problems.

4.3.2.4 Tailoring the programme to the user's needs

Practitioners of various disciplines populate the medical world. What they have in common is a medical knowledge in the form of the taxonomy of diseases. Because of its comprehensiveness many special domains have been created; each domain having its own special characteristics.

The DDSS embodies two principles:

- standardisation of nomenclature by strictly maintaining the principle of processing patient data without interpretation and valuation;
- every domain needs a specific knowledge base. The support function can only operate optimally if the patient data correlate with the domain-specific disease- and syndrome-descriptions.

The construction of domain-specific medical knowledge bases is a future development of DDSS. For the experiment we made use of disease descriptions as found in various medical textbooks to generate a very simple database.

4.3.2.5 Productivity

Computer-based support systems may be valuable, but most of these technologies are not sufficiently adapted to the processes in routine medical practice. Computer applications have been characterised as inflexible, with a poor interface to its user and of no readily perceivable benefit to doctors or patients. So far, the use of computer programmes lengthens rather than shortens the time spent on a average case. There is little to show for the enormous amount of work in this area. [Fri77] [McM83]
Without obvious benefit in terms of productivity, the use of these systems will be minimal. For a practising doctor the system must at least fulfil two aspects of the diagnostic process:

a) it must be supportive in accomplishing the task (or at the very least not interfere with it);
b) it must be quick; time consumption should not exceed the 'normal' consultation time.

The first aspect refers to the customary mode of problem-solving which is discussed in the section on reasoning processes. The DDSS allows any

specific order of information gathering any individual physician wants to employ.
The second aspect points to two elements: the effect on consultation time and the time users have to wait for answers of the system. Nothing is more annoying than to wait in front of an empty screen. The software engineer should be alert to this impediment.
The DDSS should enhance productivity in terms of speed and support as well as accuracy of medical diagnostics.

4.4 Summary

From the descriptions above, we have seen see that the goal of the DDSS is to improve the diagnostic process at many points: better observation, both in depth and in scope, and a better recording of the observed facts, leading to a better hypothesis generation and evaluation, and ultimately to a better knowledge base from which to derive diagnoses. We have also specified a large number of constraints for the system. To what extent have these goals been met in the current implementation of the Diagnostic Decision Support System? The construction and structure of the system is described in chapter 5, the experiment to test these questions and assertions is described in detail in chapter 6.

Chapter 5. The design of the DDSS

5.1 Introduction

In this chapter we are going to look at a specific decision support system. We will review the construction of the DSS that was used in the current study and explain the decisions that were taken during its design and implementation.

The doctor's diagnostic decision-making consists of two activities:

(1) determining the exact symptoms, signs and complaints of a certain (ill) patient; and
(2) finding an appropriate diagnosis compatible with the found symptoms for this individual patient

The most important operational demand for the system is that the DDSS can be integrated flexibly with the doctor's daily occupations, and does not affect the doctor-patient relationship in any adverse way. The responsibility for the decision must therefore at all times remain with the practising doctor, who must be in control of whether and how much support is given during the diagnostic process. When desired, the computer system can be an aid and a guide during this process.
It must be taken into account that for the doctor with a busy practice, efficiency and user-friendliness of the automated interactive system are of great importance. Therefore, the system should act as a logical extension of the usual diagnostic process of the practising doctor. It means that the typical succession of events in the diagnostic process must be followed by the system. It should contain the same logical partitions of history, physical examination and tests, and the usual medical catch- and keywords, standard menus with uniform terminology, and 'real time' responses. When seeing a patient, the doctor will be able to go through the series of menus, thus recording the patient's set of complaints. Based upon these data, the system will select a number of possible diagnoses, thus making sure that the doctor thinks of all possible diagnoses. The system automatically keeps a log of all actions so that the medical files are updated automatically. Other aspects of the patient as well as a number of administrative facts can be stored.

It is also very important for the system to be reliable. Even though it will always be the doctor who takes responsibility for the diagnosis, the system

must furnish a reliable overview of possible diagnoses. Reliability is also important with respect to the confidential data of the patients. The system should be sufficiently protected against misuse, while at the same time minimally inhibiting legitimate use. The patient data must be protected to ensure that no one but the doctor has access to them. The system must also be safe to use in the sense that it should be impossible to delete data accidentally.

As we have discussed in the previous chapter a number of problems and requirements have to be conquered in designing a practical and user-friendly decision support system.

5.2 A definition of a decision support system

The term "Decision Support System" refers to a class of information systems that support users in decision-making. "Decision" accentuates the decision-making: judgement and taking action. "Support" emphasises that the computer is merely an aid, not a replacement of the decision-maker. It promotes a dialogue between user and computer without frustrating the creativity, the experience and capabilities of the user. "System" refers to a structured approach of the problem. Especially in the medical domain with its ill-structured problems a DSS of the type we have developed provides possibilities which stricter systems such as expert systems cannot offer.

Decision Support Systems (DSS) have been characterized by van Schayk as follows:

> "The *function* of decision support systems is supporting management decision-making, especially in ill-structured problems, in order to improve its effectiveness.
>
> The *technology* that has enabled the rise of decision support systems includes on-line accessible database systems, interactive programmes, and data-communication.
>
> The *development method* of decision support systems is basically different from transaction processing systems and management information systems. Since decision-support systems are aimed at ill-structured problems, the information needs cannot be exactly defined beforehand. Therefore, a linear design method, where the specifications of the system

must be given in advance, is not appropriate for the design of decision support systems. An incremental design method is more suitable for the building of decision support systems."
[Sch88].

Generally a DSS contains three main components:

- a user-interface for communication between user and system;
- one or more data-subsystems for the various types of data to be stored;
- problem-processing subsystems containing models and programmes for processing and generating solutions for the current problem.

If necessary a fourth component, a knowledge subsystem, can be attached, especially in cases where an extensive amount of specialistic knowledge is required. This knowledge subsystem might be a knowledge-base or an expert system.
Decision Support Systems in ill-structured problem situations have met with little success so far. This lack of success can be attributed to a number of factors.

a) Many DSS have been developed in a way similar to that of well-structured problem situations, in the belief that more and better information automatically will lead to better decisions. The availability of huge amounts of data does not guarantee more effective decisions. Successful systems should smoothly adapt to the routine practice of the user. [Sch88]
b) Coordination between human decision-making processes and those of machines has drawn little attention so far.
c) There is no visible gain in effectiveness and productivity for the user.

An effective DSS yields better decisions, which means more than just flexibility in data management and overview of the information. However:

- a better decision cannot always be expressed in numerical terms, but may be the result of weighing reasonable options intuitively;
- most DSS need a considerable period of experience in order to produce the desired effectiveness;
- as most DSS are based on un-analysed decision-making methods, their effectiveness and productivity are not provable to the user.

The first step towards developing a DSS is a thorough analysis of the steps in the problem-solving process. The more the system fits the real processes the better an effective DSS can be developed, especially in the case of interactive systems with ill-structured problems as in general medical practice.
Our DDSS is based on previous research into decision-making processes in general practice [Rid86]. It fits as closely as possible the actual problem-solving processes of general practitioners in routine practice.

5.3 Developing the DDSS

In the development of the DDSS a prototyping approach was chosen. "Prototyping or adaptive design has been suggested as an effective approach for developing and implementing DSS. Empirical research has shown this design strategy is effective in establishing meaningful user involvement and high user satisfaction" (Henderson, 1985). A prototyping approach serves various goals such as:

- gradually increasing understanding of what the problem is and how to formulate it.
- defining functional specifications in practice;
- verification of the system with users during the development process;
- developing and defining the friendliness of the user-interface in conjunction with the users;
- finding and defining problems to be conquered.
- realising and testing a preliminary version of the system with potential users.

Next to these items two almost mutually exclusive objectives have been formulated: A strictly structured design with standardized terminology and a high degree of flexibility for the adaptation to developments in such an innovative domain of science as medicine.
The first attempts, written in a high-level database language, showed that such an interface might be feasible, but were much too slow. Interface and calculations were speeded up by writing them in the programming language 'C', while the database structure and -language were retained for those parts of the system implementing the existing standard for information systems in Dutch general practice.

5.3.1 Acceptability considerations

Because of our prime requirement that the system must be used simultaneously with the doctor-patient contact, the primary need of the system will be speed and ease of use.
The system is only acceptable to the doctor if it can operate in accordance with his own style and working pattern. As most doctors use a method of iterative hypothesis generation with subsequent hypothesis-driven data collection, the system should allow the doctor to follow this line of thought. It means that he or she must be able to step from one category in the system to another and vice versa within the same line of thought (hypothesis). This requires short paths, with few actions necessary to perform the things he wants to do, a flexible search through menus and options, and he should still feel in control of the operations. Any loss of control should be compensated by easy access to on-line help functions.

Further acceptance criteria are:

- the user should at all times in a general sense be aware of what the system is doing;
- generally, response times should be limited to two seconds. If more time is needed, the system should notify the user and give an indication of the progress;
- the time needed to perform a consultation with the system should not be longer than without the use of the diagnosis system;
- the system should sufficiently be protected against abuse in a way which does not disturb the legitimate user too much. The patient data must be protected to ensure maximum confidentiality;
- under strict conditions data might be shared with other health care workers for better patient care;
- accidental deletion of data should not be possible. Actions associated with data loss should require separate confirmation.

In the experiment which will be described in the next chapter we will concentrate on a number of hypotheses with regard to the programme and its users. We would like to introduce some of these hypotheses at this point. They all refer to qualities the system should have:

H1 Doctors do not find the use of the system disturbing.
H2 For users with some general computer experience the use of the system is easy to grasp.

H3 The use of the manual and assistance of the observers during the test is minimal.
H4 The user is able to find the shortest paths in searching for specific items in the SCS tree.
H5 The user has an intuitive feeling of what is going on in the system. The system follows his usual routine.

H6 The system is proof against accidental erasure of data.
H7 The use of colours will help the user with his the orientation within the menu tree.

5.3.2 User interface considerations

The user interface for an application such as the one under discussion must combine many, sometimes mutually exclusive qualities. It should be so simple to use that a manual is hardly necessary, yet powerful enough to follow the user's thoughts. It must enable the user to select any one of a hundred thousand items within a few seconds, without leaving him confused. It must show him enough information, but not more than necessary for the task at hand. It should be possible to use the programme without resorting to the keyboard at all.

- *Menu oriented*

We decided on an interface which we hoped would distract the physician from his rapport with the patient as little as possible.
Faced with a choice between two general interface styles: menu and command-oriented, we selected the former, for a number of reasons.
A good user-interface enables the user to apply operations to his or her data in the form in which he thinks of both the data and the operations. It requires few actions to initiate an operation, these actions are readily apparent, and the user can form a clear mental picture of what is happening.
A good user interface does not overload the user with information he does not need, but it must be able to show the state of the data at any time if he chooses to inquire after it.

- *Patient data input and storage*

Input is done largely with the mouse. Although it is certainly possible to use the entire system without using the mouse, this is not the normal or recom-

mended way. The mouse is of the three-button variety, it is used to select items from menus and its three buttons allow the user to activate several options for each menu item.

Only for the input of administrative data for the patient, such as name and address, a minimal use of the keyboard is needed. The system assists the user by checking its database as soon as the user begins to type a name, so that it is rarely necessary to enter more than a few letters of the name for any patient who has been seen previously. Once the doctor has identified himself to the system, he only needs the keyboard to identify the patient and to enter data manually which he cannot find in any menu but which he feels must be put in the record. This option is only offered as an escape and should not normally be necessary. In principle, the medical status of the patient is entered entirely by means of the mouse.

During normal use, the doctor has one hand permanently on the mouse, with three fingers on its three buttons, so that he does not need to look at the mouse at all, but can instead divide his attention between the patient and the screen. The screen is kept simple and uncluttered to reduce the amount of attention needed. With a keyboard, such efficiency in data entry would only be attainable if the doctor were to be a good touch-typist, which is very rare.

In the symptom data entry phase, the system always presents a menu which contains the subnodes of the SCS (Symptom Coding System: explained in the next section) node that is currently selected. One of the options is highlighted. The highlight may be moved to another item by moving the mouse. By pressing the middle mouse button, the highlighted node becomes the current node, the old menu is cleared, and a new menu is drawn, showing the subnodes of the selected node. If the selected node is a terminal node, or leaf node, nothing happens.

The user can determine from the screen where he is in the SCS tree, because every choice he has made to arrive at the current node is displayed in a separate window. This window displays the current path in the SCS tree. As the maximum depth is about 10 levels, there is no need to scroll this window: the full path always remains in sight.

By pushing the left mouse button, the user retraces his previous step and returns to the next higher menu level.

Finally, by pushing the right mouse button, the currently highlighted node is added to the patient data, along with any higher nodes which may be necessary to attach the selected node to the existing patient data tree.

In order to keep the display uncluttered, the user cannot see the full list of symptoms he has already entered, but this is available in a scrolling window at the push of a single key (F2).
This display may be purged of items that may have been accidentally entered incorrectly, by moving the highlight to the offending items, clicking the middle button, and pressing a confirming 'y' key on the keyboard. All this happens with instantaneous feedback on screen.
At all times, the keys that have useful actions associated with them are shown in a separate 'info line' at the bottom of the screen.
By hitting a single key (F4) at any time during information gathering the user may call up the diagnostic module, which produces a differential diagnosis within a few seconds and puts it on the screen.

Pre-trial experiments showed that young adult volunteers, even persons with no previous computer experience, could be familiarized with this system of data entry in about 10 minutes.

- *Use of colour*

For clarity and attractiveness, all menus are in colour. The colour scheme that is used most often is light yellow on a green background for all menus, and light yellow on blue for the background screen. The background colours differ for the main branches of the SCS: physical examination, laboratory and x-ray results, and patient history. This was done to provide an unconscious clue as to the current position in the tree. Warning messages are displayed in light yellow on a dark red background. For all menus operations and messages in the system, a similar layout and operation was used to minimize user confusion.

5.4 The origin of the symptom-coding system

In earlier research by Ridderikhoff [Rid86], when designing the experimental set-up for the research into decision-making of general practitioners, it was found necessary to measure the amount of information which doctors extract from their patients when they make a diagnosis. For reasons of comparison the patient information had to be "frozen" in a preconceived database which could be used as a constant frame of reference. To this end every item had to be coded to determine the relation of this item to other items asked during the interview. For this purpose an ad-hoc system was developed which turned out to be quite powerful in retrospect: all information of the simulated

patients could be fitted into the system, and more than 98 % of the symptoms, signs etc. the doctors enquired after in their patients could be fitted into the system by an observer. [Rid93].

The coding system itself originally derives from system theory (Weinberg) which states that complex structures can be split into less complex substructures, in a continuing process of refinement and increasing detail. This was performed using the levels of organ system -> symptom -> symptom-aspect -> symptom-subaspect as starting points, refined or modified where the nature of the symptom at hand made this a necessity.
An important concession to the modus operandi of the physician was that symptoms are not uniquely defined within the system: e.g. a headache can be coded in several ways in the current system. This causes some problems in the diagnostic algorithms, which can only partially be solved. The advantage of this approach is that the user has a simpler programme interface to deal with.

The current study used the same method of gathering data as an earlier study into the problem-solving behaviour of doctors [Rid86], i.e. consultations with simulated patients. An adaptation and extension of the symptom coding system used in Ridderikhoff's research was used in the DDSS computer system. Whether doctors themselves could be taught to use this system as efficiently as the observer of the experiments in the earlier trial, was of course one of the things to be tested, as we will see later.

5.5 General description of the symptom coding system

The symptom coding system (SCS) creates a systematic approach to medical information about the patient at the symptom level. For 'symptom' in this context, and in most of what follows, one should read 'any information which may point to a disease or health problem in a patient'.

The SCS is conceived as a hierarchical system. When expanded schematically it resembles a tree. (In such a tree diagram, it is customary to speak of the "highest levels" of the tree when the part is meant that in a real tree would be the lowest, i.e. the stem. The "deepest levels" are the tips of the outermost branches) Every node has zero or more child branches, and every node has exactly one parent node, except the root node which has no parent. The symptom description proceeds from the general to the specific level. The data representation is further specified in section 5.7

On the first levels (the stem of the tree) a tract is selected, in accordance with the usual method of questioning employed by doctors. Examples of tracts are cardiovascular, pulmonary, or urinary tracts. Within such a tract there is a relatively small number of 'main symptoms', such as pain, feeling tired, dizziness, coughing, etc. Some of these are specific to the tract - e.g. coughing is specific to the pulmonary tract-, others are more general: pain is a symptom that may be associated with any tract.

Each main symptom has an associated submenu, where the symptom is further qualified. For example, the symptom 'Pain' has a localisation, a pain type (sharp, dull, aching), an onset, an intensity, etc. associated with it.
The specification of these symptom-aspects corresponds to the level of detail that makes doctors think of specific diseases. "Pain" by itself is fairly non-informative, but "Pain of a sharp nature, with an onset about three hours ago in the upper abdomen" will probably trigger one or two ideas in the mind of the average GP.

Every symptom may be specified to several levels, depending on the specific symptom selected. Some descriptor branches are up to 10 levels deep, others only 5. There is no strict limit to the extension of a branch, beyond the general principle that the tree should remain more or less balanced in an informal sense, and that the level of detail attainable should be more or less equal for all symptoms. [Wil76]

The full tree specifying all items of information that may be associated with the patient within the system has now grown to about 100,000 items.
The condition of almost any patient presenting himself to the general practitioner may be approximately described by a collection of branches from the tree. One of the points to be investigated is whether this approximation is sufficient to satisfy its users, i.e. the doctors.

Even with this enormous number of possible symptoms, symptom aspects, and symptom subaspects, there are large areas where the tree is still very limited. Especially the field of psychiatric complaints has proved to be hard to encode. We have also restricted ourselves to those symptoms which might reasonably be encountered by the general practitioner behind his desk during surgery. Only tests usually requested by GP's, and no special investigations were included. Results of x-ray investigations can only be coded in the most rudimentary way (preferably normal/abnormal), based on the interpretation of the radiologist.
Pauker and Kassirer estimate the number of medical terms in the field of internal medicine at about 4 million. They concluded that this made a direct storage and search of the symptom space impossible. [Pauxx] Viewed in this

light, our efforts toward organization have yielded a remarkable increase in efficiency.

How does it work? The doctor selects items from menus. Each menu is a list of all 'deeper' branches attached to a specific node. If he selects a particular branch, the next menu consists of the items attached to it, if any. At any level of specification the doctor may decide to enter the symptom into the database record for that patient.
It is the doctor who decides how detailed the information should be for this particular patient, and this particular symptom. For instance, he may decide to code only that the patient has a cough. In other circumstances he may decide that the type of cough is important under the circumstances, and he may then include information about sputum production, colour of sputum, duration of the complaint, amount of sputum, whether coughs come in fits or single, whether coughing takes place at night or is associated with a specific position of the patient's body, etc.
It may be argued that the doctor does a lot of interpretation in this way already, pre-processing the data of the patient into the forms required by the tree structure. This is true, but it is likewise true of every other diagnostic system where the doctor enters patient data. And because of the specific terminology used by the system, interpretation is less intuitive than usually occurs. At least the doctor uses the same term for the same symptom as his colleagues.
The SCS tree was designed in such a way as to minimize observer interpretation of patient data. For example, it is not possible to enter a symptom like 'aortic valve diastolic heart murmur' into the system. Such a 'symptom' is an interpretation of an observation without the possibility of tracing the composing elements on which the judgement is based.
In the SCS, this is coded in a way to preserve the direct sensory input as much as possible, eg:

heart/murmurs /localisation/thorax, upper left;
/timing/diastolic;
/loudness/(3/5)

Such a coding system is more elaborate and slower, but more objective than 'aortic valve diastolic regurgitation murmur'.

Any data entered into the system are used by the computer to try to find matches between these data and the built-in disease database. (This knowledge base is to be developed for a specific medical discipline over time.)

Whereas a doctor is inclined to look for alternative hypotheses within the tract to which his attention is currently focused, the computer has no such handicap, and it can compare the symptoms against thousands of hypotheses in a few seconds.

5.6 Some difficulties encountered in coding the SCS

One difficulty is the size of the tree. Firm discipline is required to keep the coding scheme internally consistent when the tree grows to tens of thousands of nodes. A fundamental change in design for a particular problem may necessitate the modification of branches in various other parts of the tree. This task can to some degree be automated, however.

Another difficulty is the problem of the representation of time-related phenomena. This is a problem which is not specific to our system, but which seems to occur in most diagnostic programmes and related systems, including the doctor's written patient records. The solution adopted in our system, as in many others, is to regard the current status of the patient as the one to be encoded. Intervals, onsets, durations of complaints, etc. are encoded with respect to the moment of presentation.

A third problem is the arbitrariness of encoding data on several levels; given two characteristics, each of which might reasonably be selected as the higher level, often an ad hoc decision had to be made which level of detail should be encoded first. Sometimes it was instead decided to encode both at the same level, which gives rise to the fourth point:

The fourth problem is the ambiguity which occurs in specific cases when different instances of the same symptom are to be encoded: e.g. if a patient has two different pains, the specification of aspects of these pains cannot be retraced unambiguously to the belonging main symptom, i.e. if someone has both a sharp pain in the thorax, and a dull pain in the abdomen, then after encoding this could not be distinguished from a patient with a dull pain in the thorax and a sharp pain in the abdomen. This may seem a big disadvantage, but it is less so in practice; most patients do not have more than one such main symptom at a time. The advantage of this way of coding is that it makes the tree more accessible to the user so that data entry can be performed more quickly.

5.7 Data representation

The symptom tree is *conceptually coded* as a tree of varying order and depth. A tree is a graph where every node has exactly one parent node and zero or more child nodes, except the root node which has no parent and zero or more child nodes.
Any vertex, or node, may have any number of child nodes (within reason, in the current implementation the maximum allowed number of child nodes is 128, the maximum occurring number about 35); any branch may continue to any depth. To give a very approximate idea of the current dimensions of the SCS tree, the average depth of the symptom tree is about 7 levels; the average branching factor is about 6.

The symptom tree is *implemented* as a binary tree, which is a tree where any node has zero or two children and every node has exactly one parent node except the root node, which has none. The symptom tree, of varying order and depth, is coded as a binary tree, using the scheme where any right child node is the brother with the next higher number, and any left node is the first son node. Brother nodes on the same level thus form a linked list of right children in the binary tree, and all children of a given node are found as a linked list of right children on the left child node of the binary tree. See figure 1 and 2.

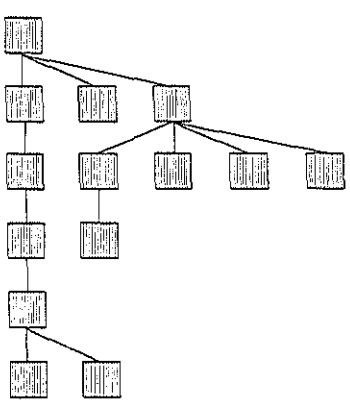


fig. 1 An example tree

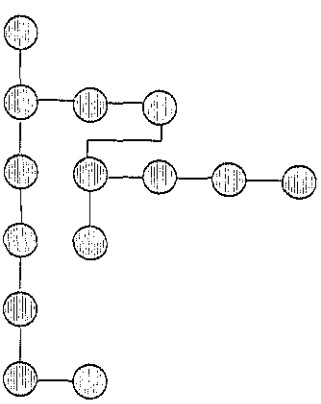


fig. 2 Translation to binary tree format

Because of the need to extract subtrees from the main symptom coding tree, both for patient and for disease data, the logical structure of these subtrees cannot be used to reconstruct the path in the main tree by itself, and every node in a subtree has to be numbered. It carries the number of the corresponding node in the main symptom coding tree. See figures 3 and 4.

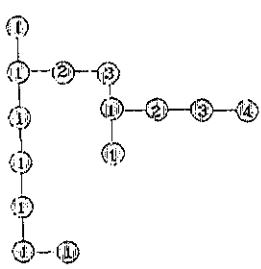


fig. 4 Numbering scheme used for tree nodes

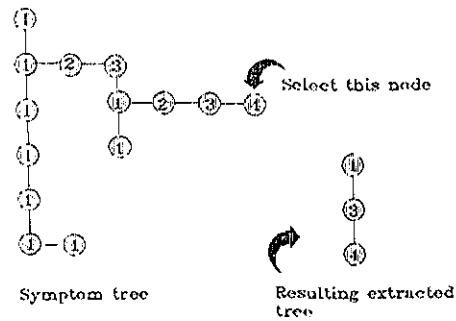


fig. 3 Information loss without numbering

Brother nodes are sorted, i.e. higher-numbered nodes are always to the far end of the chain. Every node has an associated record of information, which consists of a number of fields:

- the number of the child node with regard to its parent; (the first child is numbered 1, the second 2, etc.)
- the name of the node, a text field;
- the significance bit;
- a weight;
- a presence bit;
- a comment field.

The text belonging to a node serves to identify it to the human observer; the number to identify it to the computer. The significance bit determines whether this node is an 'information-carrying node' or just a 'structural node'. (The difference is explained below.)

The weight can in principle be used (but it is not currently) to assign various weighing factors to different symptoms in the context of specific disease patterns. But such a valuation assumes an unambiguous notion of the disease and a subsequent notion of the importance of its composing symptoms and signs. This is not a general characteristic in medicine and certainly not in general practice.
Some envisaged applications require that weighing of symptoms should not be used. The (numerical) taxonomic process to determine natural groups of symptoms in order to create taxonomic disease classes and entities is such an application. Besides, in most cases, too little is known about the weight of specific symptoms in general practice to justify weighing of findings at this stage.

The presence bit can be used in patient data sets to determine if an item of information has been inquired after. If it is set, it means that the item has been entered as present. If it is not set, this means that the item has been enquired after but that it was not present. (If the item has not been enquired after at all, the symptom does not appear in the tree for that patient.)
This is important for later extensions: the system could use disease patterns to suggest questions which would be able to differentiate between two hypotheses in its collection.
It would also enable doctors to check whether an inquiry had been made and found to be normal, or whether the inquiry had not been made at all. In the first case, it would be present but with the presence bit cleared, in the second case it would not be present in the tree at all.
In principle, only abnormal findings are entered in the current system.

In addition, every node has an associated comment field, where the doctor can enter a line of text. This text cannot currently be used in any way by the system but it can be recalled on screen and if desired modified by the doctor. This was included to provide an escape possibility against incompleteness of the symptom coding system.

5.8 A simple description tree as an example

By constructing a description tree of a simple every day object we may illustrate most of the characteristics of the system in a more readily comprehensible form.

On the left we give the structure of the tree in text; on the right we give the coordinates of every tree node.

```
object                  1
   seat                 1,1
      colour            1,1,1
         red            1,1,1,1
         green          1,1,1,2
         blue           1,1,1,3
      material          1,1,2
         wood           1,1,2,1
         metal          1,1,2,2
         plastic        1,1,2,3
   legs                 1,2
      number            1,2,1
         1 leg          1,2,1,1
         2 legs         1,2,1,2
         3 legs         1,2,1,3
         >=4 legs       1,2,1,4
      material          1,2,2
         wood           1,2,2,1
         plastic        1,2,2,2
         metal          1,2,2,3
      colour            1,2,3
         red            1,2,3,1
         green          1,2,3,2
         blue           1,2,3,3
```

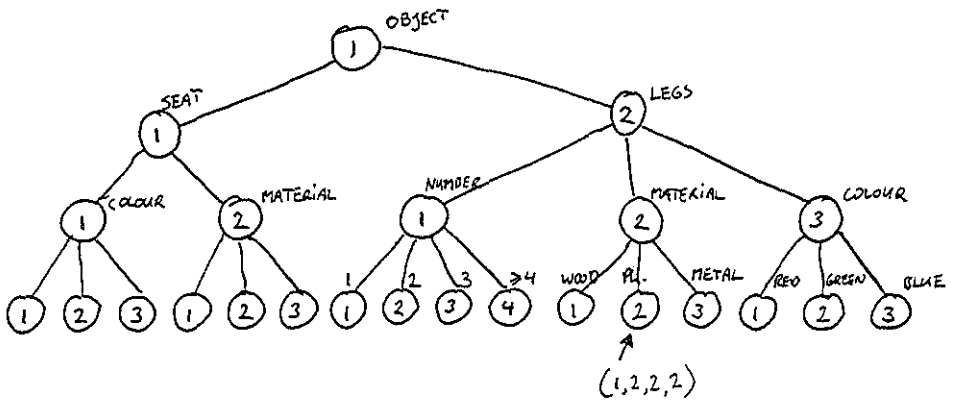


fig. 5 Graphical illustration of 'chair'-tree

If we want to represent an object with a red wooden seat and four green metal legs, we can pick the following subtree out of the above:

```
object                  1
   seat                 1,1
      colour            1,1,1
         red            1,1,1,1
      material          1,1,2
         wood           1,1,2,1
   legs                 1,2
      number            1,2,1
         >=4 legs       1,2,1,4
      material          1,2,2
         metal          1,2,2,3
      colour            1,2,3
         green          1,2,3,2
```

This gives us a description of our chair that is rather dull but fairly specific, and that can be used to identify chairs that look a bit like it.
Note that there are *descriptive* and *structural* nodes in the above tree. An example of a structural node is 'object', which exists only to connect its branches into a tree. It does not, in itself, convey any useful information

about our chair. In the same way, the nodes 'colour', and 'material' are in themselves non-informative, structural items. On the other hand, 'red', if specified with either seat/colour or legs/colour, is an informative item.
'Seat' is arguably informative: it conveys that the object does have a seat. It would depend on the context in which the system is used whether this is informative: if there are only chairs, this seat would not qualify as informative, but if the objects included tables, it would. Whether tree nodes are informative or not depends on the context. In our system, significance may be optionally attached to any node, depending on the disease.
If we want to generate a template to find similar chairs, we may assign to each node a 'significance' status:

name	coordinates	significance
object	1	0
seat	1,1	1
colour	1,1,1	0
red	1,1,1,1	1
material	1,1,2	0
wood	1,1,2,1	1
legs	1,2	1
number	1,2,1	0
>=4 legs	1,2,1,4	1
material	1,2,2	0
metal	1,2,2,3	1
colour	1,2,3	0
green	1,2,3,2	1

If we give a candidate object 1 point for every matching significant node, an unknown object can gain at most 7 points when compared with this tree pattern, in which case we can be sure that it has a red wooden seat and four or more green metal legs. It is probably a chair. If the comparison results in fewer points, there may still be a resemblance.
A computer, given this template and a number of objects to test against it, can easily make a list of objects which match this template, or pattern, to a greater or lesser extent, in order of the greatest resemblance. This even applies if a fallible human enters the characteristics of these objects, leaving out a few. The results become less reliable, but the method still works. This is precisely what happens in the DDSS.

Within the system, there is a collection of 'chairs' which specify characteristic patterns for common diseases. The doctor enters another pattern, which consists of his findings in the patient to be tested. The computer

compares these findings against the known disease patterns and sorts them in order of best match. The more data the doctor enters, the better the chance of a correct diagnosis.

5.9 Implementation decisions.

An earlier prototype of the symptom-coding system had 'dBase III+' as its programming language. This showed the feasibility of the idea to code symptoms but suffered from a few drawbacks:

- it was extremely slow, even when compiled with a dBase compiler (Clipper)
- it was hard to modify; the initial design called for a number of orthogonal characteristics for every symptom (i.e. every possible subsymptom appeared with every symptom). But many subsymptoms are meaningless when combined with a different main symptom, and the number of permutations that turned out to be necessary would have led to insurmountable problems.

Accordingly, the decision was made to represent the symptoms in the form of a tree. This was a natural choice, as a tree is the natural structure for a hierarchical system. For the symptom hierarchy, this design had the advantage of extreme flexibility: any node may have any number of branches (within certain limits which are mainly determined by the capacity of humans to pick an item from a menu in a reasonable time) and any branch may be extended to as many levels as is necessary to code the underlying medical information.

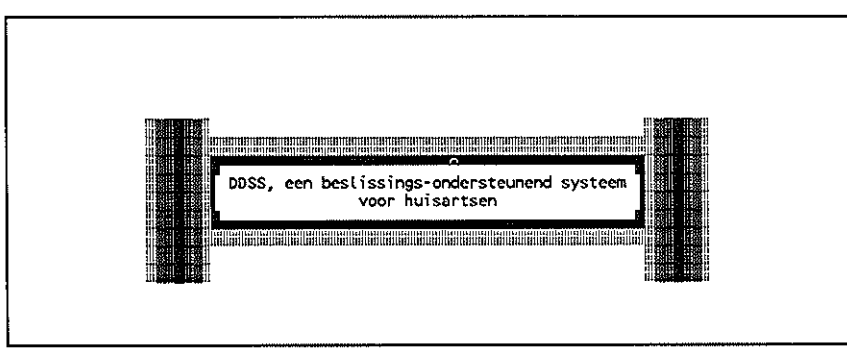
It is possible to represent trees economically in terms of system resources, and comparison of trees can be done with efficient algorithms.
The trees with the symptom structure were constructed with an ordinary word processor or editor (Wordperfect or Qedit) in a predefined format, and then translated to computer data files by a compiler programme. The input for the compiler programme must be plain ASCII text. Some slight difficulties were encountered with the translation of special accented characters from Wordperfect-format to ASCII format.
Changes in the tree do not necessitate any changes in the programme unless a major restructuring of the upper branches should become necessary, which, so far, has not been the case. Usually it is sufficient to change the tree file

and recompile it. This has only a local effect on the structure of the tree: data gathered with the old tree will become only partially invalid with tree updates, namely only if such data contains elements which are part of the changed menu and any nodes below that.
This graceful degradation of performance means that it is possible to correct minor deficiencies in the tree without rendering masses of gathered data invalid. It is also theoretically possible to transform existing data from the old form to the updated one, but a tool to do this has not been written.

For reasons of speed and economy of memory the SCS system was coded in C. A surrounding shell programme was constructed with Clipper 5.1 to take care of the more routine database tasks.
The tree resides on disk, and branches of it are swapped into and out of memory as needed; the total space occupied by the tree on disk is about 2 megabytes, or only about 20 bytes per node. This is a consequence of several tricks to economize on the memory needed. The swapping occurs with a hardly noticeable delay: any menu is presented virtually instantaneously.

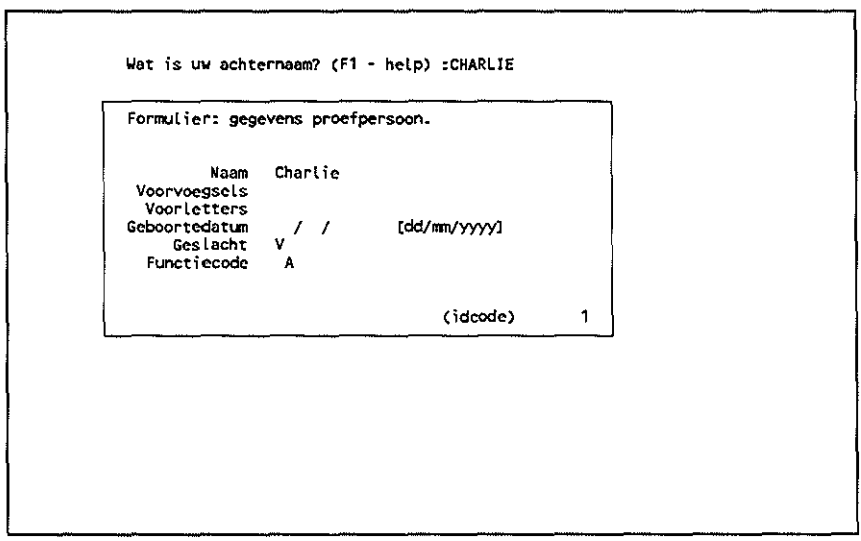
5.10 A user's view of the whole system.

In this section we will describe the way the system presents itself to the user. On typing "ddss" at the MS-DOS prompt, an introduction screen appears.



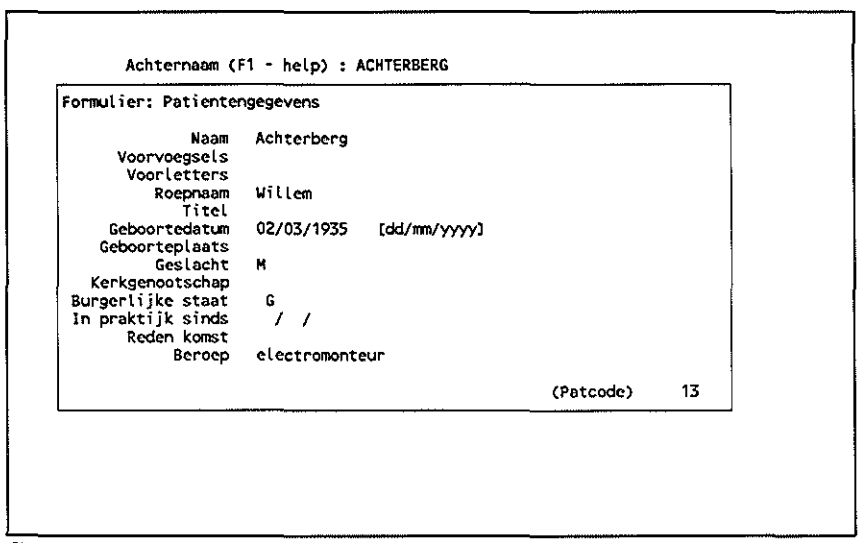
Screen 1 startup screen

After the system has been started, its first action is to ask the user, i.e. the doctor, to identify himself. If this is the first time the user logs on to the system, he is asked to fill in a few identifying fields. Otherwise just typing in the first few letters of his or her name will suffice to determine uniquely which doctor is meant.



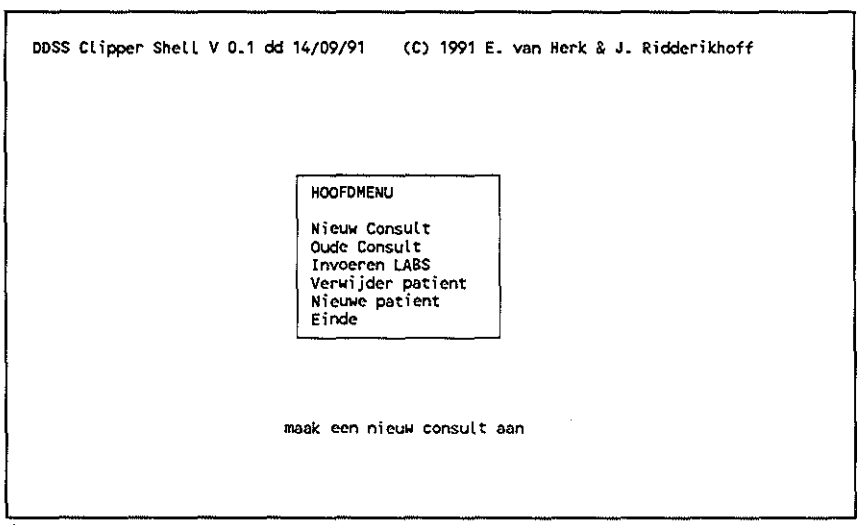
Screen 2 doctor identification

The system now proceeds with a screen for the selection of the patient from the database of all patients. The database structure underlying both screens has been designed to conform to the requirements of the national Dutch College of General Practitioners (NHG). This facilitates compatibility with systems already in use in general practice.



Screen 3 patient registration screen

After looking up the patient in the database, or creating a new record for the patient if he was not already present in the database, the system shows the main menu:



Screen 4 main menu

The items offered are:

New consultation
Previous consultations
Data storage for laboratory tests
Erase patient
Next patient
Exit

Most patients are already known to the doctor. After verifying the administrative details he is free to start a new consultation, i.e. the current episode for which the patient has come to see him. By selection of this option, the doctor can gather and store a full set of data for this patient in order to arrive at a conclusion.
Some consultations however may be connected to previous visits of this patient to the doctor, e.g. follow-up after therapy, or periodic visits because of chronic ailments such as diabetes mellitus, rheumatic disorders, etc. The doctor can retrieve the details of past consultations by selecting Previous

consultations, and look at the details. (But he cannot modify the data of old consultations.)

It is also possible, when "new consultation" has been selected, to load the data of a previous consultation as a starting point, and edit those symptoms which no longer apply. This has no effect on the data for the previous consultation, as the edited version will be saved only as part of the current consultation data. In this way the user has the ease of not having to re-enter largely identical data, but the option of retrospectively changing his bookkeeping on the patient, which is generally undesirable [Mar91], is avoided.

The results of lab tests ordered for the patient during a consultation will usually arrive later, when the patient has gone. The "data storage" option is meant to offer a separate interface to add these results to the consultation data, but is not yet functional in the current implementation.

Patients sometimes move to another city, or to another doctor, or they may die. In such cases the patient data can be deleted from the database. Note that it is not possible to delete individual consultations; only the entire set of all consultations and other data for that patient can be deleted. This option "Erase patient" is of course protected by suitable admonitions of "are you sure?" etc, and requires repeated explicit confirmation from the user.

When the next patient enters the consulting room, the doctor can choose "next patient" to go to the patient selection screen and identify the patient by typing the first few letters of the patient's name. He can also page through the database one record at a time, to verify that the patient record is correct, which is especially important in the case of several patients having the same name, initials, date of birth, etc.

Finally, the option "Exit" leaves the programme, after asking for a confirmation from the user that this is what he or she wants.

Normally, however, at this point the user will select "new consultation". The next menu offers three options:

Patient data
Diagnosis
Treatment

In most cases the doctor is unable to arrive at a diagnosis immediately. He or she needs information to come to a conclusion. Once this information has become available the doctor may decide on a diagnosis. By clicking on the option "Diagnosis" a diagnosis may be selected and added to the patient

data. About seven hundred different diagnostic names may be selected from the menu system under "diagnosis" so far.
The menu option "Treatment" offers a similar menu system enumerating therapies and other types of action (referral, treatment with medicine, etc).

Finally, the SCS proper can be found under the option "Patient data", and this will usually be the first selection the doctor makes at this menu. Under this heading all encodeable information about the patient's condition is to be found. The SCS follows the usual order of categories commonly used in general practice. The submenu accordingly looks like this:

Social	(family, occupation, work, hobbies, etc)
Old medical history	(vaccinations, old operations, accidents)
Medical history	(for the current episode)
Physical examination	
Laboratory tests	(biological, biochemical, immunological and physical tests)

The menus in each of these categories have different background colours, to aid the user in orienting himself within the system. The bottom line of the screen shows the options from which the user can select at each point. A box on the right of the screen shows the choices he has made to arrive at the point in the menu tree where he is currently located.

If our user now follows the usual order of events, he will first of all select "medical history" at this point, which offers a new menu showing a list of organ systems of the human body. (Apart from a few mild psychological problems the system is currently not usable for mental ailments.) First an organ tract is chosen, according to the usual method of questioning the patient. Examples of tracts are cardiovascular, pulmonary, or urinary tracts. Within each tract there is a relatively small number of 'main symptoms', such as pain, feeling tired, feeling dizzy, coughing, etc. Some of these are specific to the tract in question: "coughing" can only be found in the respiratory tract, but others are more general: "pain" is a symptom which occurs in practically every tract.

Each of these main symptoms now has an associated submenu, where the symptom is further qualified. As stated above, the symptom "pain" has an associated localisation, type (sharp, dull, aching), type of onset (sudden, slow), duration of onset, intensity and many other facets associated with it, some of which have yet deeper levels of description, in submenus of their own. Every symptom the patient exhibits may in this way be specified to an

arbitrarily deep level, or at least as deep as the symptom tree allows the doctor to go.
At every point in the consultation, the doctor is looking at a screen with two windows. The left window shows the current menu, with all selections that are available at that level.

The screen at this point is fully controlled by the mouse. Moving the mouse highlights (visually selects) an item in the current menu. Pressing the left mouse button undoes the previous selection, returning the user to the next higher menu level. In this way, from any point in the tree, the user can return to the top menu by a few quick successive clicks on the left button without having to think about it. The middle mouse button, on the other hand, takes the user to the submenu of the item that is currently selected on the screen, thus going deeper into the menu tree. Finally, the right mouse button adds the currently highlighted item in the menu to the data gathered for the current patient.
When all complaints of the patient have been communicated to the doctor and stored in the dataset for that patient (in so far as they represented abnormal findings) the doctor usually proceeds with the physical examination. He returns to the top menu by clicking a few times on the left mouse button, then selects "physical examination".
While the menu system under "medical history" was organized by organ system, because that is how a doctor usually performs his questioning of the patient, the menu system under "physical examination" is organized topographically. Again, this structure was chosen because this is the way most doctors perform a physical examination: they listen to heart and lungs of the patient, both on the thorax, before proceeding to the abdomen. The body parts are arranged alphabetically. Let us suppose the item "abdomen" is chosen. The submenu associated with the abdomen shows the items:

general
organs
rectum and anus
muscles and tissues
skin
signs, reactions, tests

If from this menu, "general" is selected, the system first offers a selection of the method of examination to be used: inspection, percussion, auscultation, palpation, or some specific instrumental measurement or investigation such

as the circumference of the abdomen, or skin fold measurements, or rectoscopy. If the doctor selects "palpation", the next submenu shows a number of signs which may be detected by means of palpation, such as localisation, form, structure, dimensions, single or multiple lesions, pulsations, and symmetry, among others.
In general, for every type of measurement, a special dimension scale is available. In this way lenghts, heights, weights, circumferences, angles of joint movement, deformities, temperatures, reflexes and sensibilities, acuteness of vision and hearing may all be measured. Every measurement is stored in a category, not as the actual value found, although this may be added to the data in the form of a comment. For instance, systolic or diastolic blood pressure can be either very low, low, normal, high, or very high.

In the "Laboratory results" partition of the menu system the user may choose from the options:

blood
faeces
sputum
urine
microbiology
genital infections
smears
imaging techniques

Most of these are subdivided into three partitions in the next menu layer. Below them are about three additional levels of menus to allow accurate description of many laboratory test results. The item "imaging techniques" refers to radiology, scans, etc.

After the doctor has selected from all menus those items which together give an accurate representation of the patient's condition, he will probably want to make a diagnosis, a judgement. Going back to the top menu, the user selects "diagnosis" and looks up the item sought in the diagnosis menu system, which is organised in the following four sub-items:

General impression
multiorganic, non-specific
diseases ordered by organ system
mental and social ailments

"general impression" refers to such general statements as ill vs not-ill, serious vs not serious, somatic vs mental.

"multi-organic, non-specific" refers to three separate groups:

- Symptoms/complaints, i.e. "diagnoses" which are nothing more than a repetition of the patient's main complaint. These 'symptom diagnoses are usually just working-hypotheses for the doctor.
- assorted illnesses sorted by group without further qualification, like allergy, neoplasm, trauma, intoxication, growth disorder.
- diagnosis by localisation: "something wrong with" head, abdomen, breast, leg, and so on.

"diseases ordered by organ system" offers the most specific descriptions of diseases, by the names under which they are documented in the medical literature.

As was mentioned before, the system is not currently applicable to serious mental illness e.g. psychosis and depression. Nevertheless, the system does provide the doctor with a few names of the more prevalent disorders with a large psychological component found in general practice: tension headache, fatigue, hyperventilation, dyspepsia, etc.

When the user finally returns to the top menu, he can choose the "treatment" item. Three main branches are offered: consultation, referral, and treatment by the GP himself. Consultation and referral lead to different alternatives ranging from specialist consultation to alternative medicine healers to physiotherapy to abortion. Under "treatment by GP" we find advice, diet, surgery, and drug treatment. The drugs are arranged by organ system, following the lead of most pharmaceutical books, and are present only as generic names. This was done because including brand names would unnecessarily expand this part of the tree to a multiple of its present size. Every drug is represented only once, helping to reduce the chance of double prescription.

Now the consultation is over. The user returns to the main menu by repeatedly pressing the left mouse button or the <esc> key, and is asked to confirm that he wants to close this consultation and that he wants the results saved. (the default for the first question is no, the default for the second question is yes.)

Now that we have described the system from the user's view we would like to introduce at this point a number of additional testing hypotheses for the experiment to be performed with it. They are concerned with the behaviour of the doctor.

H8 By the presentation of the menus doctors will question their patient broader (more categories of symptoms and signs) than when unguided by the system.
H9 Doctors will be able to encode all their observations within the SCS.
H10 By the presentation of the menus doctors will be stimulated to probe deeper with their questions, and so to get a more detailed picture of the patient.
H11 Doctors are able to handle the standardised terminology without trouble.
H12 It will only rarely be necessary to make specific notes in the free-text field for specifications beyond the levels of specification within the SCS.
H13 The doctor will be able to retrace his previous steps of data acquisition at any moment of the process.

5.11 The system's reasoning process

In the first chapter we have argued that there is not much evidence for the often postulated logical and systematic approach to the diagnostic problem in doctors, as described in an ideal form by e.g. Ledley and Lusted, Elstein and by Blois. [Led59] [Els78] [Blo83] There are however grounds to suspect a much more haphazard approach.
As argued before, doctors employ a more heuristic way of problem-solving. A number of standard steps can be observed: hypothesis generation based on little data, hypothesis-driven data acquisition, and a mixture of probing and testing cues. The process impresses as one of iterative pattern recognition.

This pattern-recognition approach was quite common in the early days of computing, both because it is easy to implement for simple cases and

because the limitations of the machinery at that time made more elaborate programmes impractical. Moreover, the theory for other approaches was still being developed. Such programmes, applied to well-circumscribed areas, like kidney disease syndromes, enjoyed a modest popularity until they were generally displaced by the more modern expert-system approaches. Examples of programmes that operated mainly on some form of pattern-recognition, sometimes completed with weighing methods of some kind, are RECONSIDER [Blo81] and DXplain [Bar87].
Like ours, these systems do not produce diagnoses, but rather lists of differential diagnoses based on the quality of the match between patient symptoms and its knowledge base of disease characteristics.

In family medicine, the range of possibilities for diagnosis is practically unbounded. People come to their general practitioner with extremely diverse complaints and conditions, many of which are not medical in a strict sense. To set up an exhaustive differential diagnosis, then set about eliminating them one by one from the list is simply not practical under these circumstances.
With the current state of formalized knowledge, it is also quite impossible to catch the work of the general practitioner in the rulebase of an expert system. Therefore, the expert-system approach to the work of the general practitioner as a whole seems unpractical.
On the other hand, if it were possible to standardize the description of a patient's symptoms to a sufficient degree, it might be feasible to let a computer scan the symptoms for matching diseases, and thus to generate a first-approximation differential diagnosis.

Such a standardized symptom description is one of the main characteristics of the DDSS.
Pattern recognition is a bag of tools for a bag of problems (Kanal). [Kan68]. Template matching is a robust method, in the sense that it is not very sensitive to missing or partially incorrect data. When implemented efficiently, it can be fast, because the basic operation, comparison, can be done very efficiently on a digital computer. This is an advantage over reasoning systems which need all elements in a chain of reasoning in order to be able to arrive at a conclusion. A pattern search may still produce a reasonable match, when the chain of reasoning has broken.

5.12 The pattern recognition algorithm on symptom trees

Given two trees, the pattern recognition algorithm consists of determining the common nodes in them. Comparing trees is a well-known procedure which has been studied extensively. Research by Day has shown that optimal algorithms exist which require O(n) time, where n is the number of nodes common to both trees to be compared [Day85].

In the case of the DDSS, the trees to be compared are both subtrees of the complete SCS tree.
The algorithm currently implemented is recursive, and can be described in pseudocode as follows:

```
- "X.nr" is a notation that refers to the number of X in the sequence of
  all its possible brothers.

- Brother(X) is a function that returns the next higher-numbered brother
  (right node) of X.
- Son(X) is a function that returns the first son (left node) of X.
- Valid(X) is a function that returns TRUE if X exists
- Informative(X) returns TRUE if X is marked as an informative node.

A and B are pointers to nodes. At the start of the algorithm, A is
initialized to the root of the first pattern to be compared, and B to the
root of the other pattern.
The integer variable Counter keeps tally of the number of matching nodes,
and is initialized to 0.

compare(A,B)
{ while(valid(A) and valid(B))          /* otherwise, a dead end was */
  { while(A.nr < B.nr)                  /* detected                  */
    { A = brother(A) ;
    }
    while(B.nr < A.nr)
    { B = brother(B) ;
    }
  }

/*  The preceding sequence synchronizes A and B until a match is found or
    none is found to exist.
 */

  if(valid(A) and (A equals B))         /* we have a match */
  { if(informative(A))
    { increment counter;
    }
    compare(brother(A),brother(B));     /* same process recursively with */
    compare(son(A),son(B));             /* left and right subtrees.      */
  }
}
```

Performance of the comparison algorithm.

For the synchronization part, (first half of the algorithm) all nodes on the same level of A and B are scanned, requiring in the worst case a time proportional to the number of nodes on that level for A plus the number nodes on that level in B. (Since the nodes are sorted, the algorithm never has to go back to a node already seen.) In the typical case where one node in tree A has to be compared with one, two or three in tree B, on average half that number of comparisons are necessary. Only for the matching nodes, the lower levels are evaluated. The time needed for a full comparison between two whole trees A and B therefore depends on the number of brother nodes at each level where at least one match occurs and the number of matches, and is the product of both. As the average number of brother nodes is a small fixed number, even in the larger of the two trees, the time needed by the algorithm is proportional to the number of matching nodes in the two trees.
Since there are only rarely more than a few nodes selected at any level, it is reasonable to estimate the time required for a comparison of two trees as a function of the number of matching nodes, multiplied by a small constant, in the order of two to three. For example, if there are 5 matching nodes in the tree, (a fairly typical value), the algorithm has typically to visit about 10 to 15 nodes to determine this. From this we see that the comparison algorithm by itself can be extremely fast.
However, in the current implementation, the speed of comparison is I/O bound and depends mainly on the speed with which tree data can be read from disk. The current system compares the data of a typical patient, as entered by our test subjects, with its database of diseases at a rate of about 10 diseases/sec; (80386 cpu, 25 mhz, hardddisk 16 ms access time).
But since profiling the programme execution showed that rather more than 80% of this time is currently spent in reading single bytes from the tree into memory, with a few very simple optimizations (block reads, disk buffers) we expect that this can be raised to at least 50/sec, and with a bit more optimization effort to 100 or 200 per second on similar hardware - which should be enough to check a database of 1000 or 2000 diseases against the current patient within a few seconds.

The results of the matching process are kept in a list, and sorted in order of decreasing score. The disease pattern with the highest score is presented first. There is a minimum necessary score which must be reached for a diagnosis to be shown at all. Currently this is set to three, but this arbitrary

choice should perhaps be set higher. If no diagnosis matches the patient's symptoms to at least that score, a message appears on screen to the effect that there are insufficient data for a meaningful conclusion. Besides the matching score, the highest possible score for that diagnosis is also shown. Not all diagnoses have equal numbers of symptoms which can be matched; some may consist of 20 items, others of 200. A matching score of 5 might be more significant for a disease in the former category, but this is not taken into account currently. We decided that neither the absolute number nor the proportion of non-matching nodes should be taken into consideration. On the one hand, in absolute numbers this result may vary widely with the extensiveness of the description of the disease itself; on the other hand the proportion of symptoms which do not match will generally exceed 90 %.

Diagnoses found are presented in a menu and the definition of such a diagnosis may be examined by the user. The nodes that make up the database definition of the diagnosis pattern can be viewed in the same format as the symptoms for the patient during symptom gathering.
Matching items could be highlighted on screen, thus giving the user a better feel for the quality of the match, but this is not implemented presently. (A suggestion from one of our test subjects.)

Whether the subject decides to accept a diagnosis is his own affair; the purpose of the system is to give him ideas and to guide his data collection. Having introduced the diagnostic algorithm, this is a good point to introduce the remaining hypotheses related to the diagnostic support which will be tested in the experiment:

H14 The doctor will consider more hypotheses in his decision making after having seen the system differential diagnoses.

H15 In undecided cases doctors will ask for more symptoms in order to confirm or reject particular hypotheses.

H16 By offering the degree of concordance (number of analogous symptoms) between the symptoms of the patients and symptoms belonging to a disease, the doctor is will be driven to ask for more symptoms of the patient (Confirmation = hypothesis given the evidence (Carnap).

H17 Doctors will use the option of looking up symptoms for suggested hypotheses to assist in their justification process and check for missing symptoms.

5.13 The DDSS knowledge base.

General practitioners have their own domain of medical knowledge, as they operate in an area of the health care system which is mainly populated with vaguely defined illnesses and disorders. Therefore most general practitioners carry their own part of medical knowledge and their own vocabulary. Regrettably, all this knowledge, predominantly based on personal experience, has not been condensed into a generally accepted and tested domain-specific taxonomy of diseases. The support function of the DDSS operates optimally only when the patient data can be compared with domain-specific versions of the disease- and syndrome-descriptions.
Developers of computer-based decision-support tools frequently adopt pattern recognition techniques as the basis for their programmes. The principal challenge in the creation of any clinical consultation programme lies in creating a computational model of the application domain. The difficulty in generating such a model manifests itself in symptoms that workers in the expert systems community have labeled "the knowledge-acquisition bottleneck" and "the problem of brittleness". [Mus89] The process of knowledge acquisition traditionally concerns the elicitation and encoding of given professional's relevant expertise to create the knowledge base of an expert or decision support system. Brittleness refers to the failure of such a system to offer appropriate advice on specific classes of cases not used in the construction of the knowledge base.
The problem of the knowledge acquisition is that experts (in our case, experienced general practitioners) simply cannot explain what they know or how they solve problems. Knowledge engineers who try to communicate with the medical experts have great difficulty in extracting the specific elements of expertise. And even if they succeed, they still cannot be certain about the general validity of this expertise. Brittleness also refers to the difference of one expert to another in particular cases. In the ill-structured world of primary health care exact definitions of ailments and illnesses (most problems seen in general practice have not yet evolved to the recognisable diseases described in textbooks) are hard to find. In our view this situation will last for many more years.

To bypass the problem of applying pattern-recognition methods to undefined illnesses we had to adopt another option. As was common in earlier times, we elected to define illnesses by their specific symptom configuration. A disease is recognised as such when it is predefined in the taxonomy of diseases. How can doctors diagnose a particular disease? Because they look

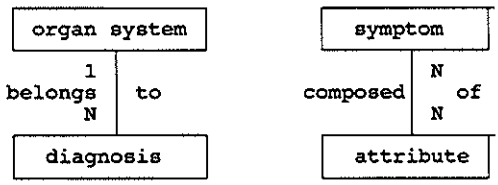
for some analogy between the symptoms as presented by the patient and the constituent elements of a predefined disease. The greater the similarity, the more a diagnostic hypothesis may be confirmed.
By means of the presentation of the alternative diseases in their symptom configuration the doctor is enabled to judge the value and the accuracy of his diagnosis. It will give him ample opportunity to (re)consider all the possibilities offered in the differential diagnosis, "lateral thinking." [Bon70]
This conception differs in two ways from other expert or decision support systems.

(1) Most (medical) knowledge systems are directed at restriction of the possible alternatives, preferably to the best one. DDSS takes another approach. It will confront the doctor with as many explanations for the present symptom-configuration as possible. It leaves the doctor to decide which one is the best fit.
(2) By the presentation of alternatives in their constituent composition of symptoms DDSS can offer the doctor an encyclopedic function.

Since the knowledge representation as is foreseen in the DDSS knowledge base is of the type of "frame representation", the knowledge base (and its subsequent pattern recognition process) is not restricted to particular domains, but can be adapted to any area in the medical world. A "frame" is a structured representation of an object or a class of objects. A frame represents an entity which may contain a number of descriptive attributes, also called "slots". In a hierarchically more detailed level attributes may be defined by so-called "facets". With help of such a representation a taxonomy for a specific knowledge domain can be formulated.
The Symptom Coding System is a perfect mechanism and interface to process and store symptoms and attributes automatically. The number of symptoms and attributes is variable which makes the system very flexible.
Three relations can be formulated:

a) The relation between diagnosis and symptom;
b) The relation between symptom and attribute (symptom-aspect);
c) The relation between attribute and facet (symptom-subaspect).

This can semantically described as follows:



If necessary the "frames" may contain a diagnosis-label. Pattern recognition, however, will be performed by comparing symptom-configurations with symptom-configurations.
For the time being, such a knowledge base still lies in the future. The tools are there. We only need the resources and the co-operation of many general practitioners.

Chapter 6. Methods

6.1 Introduction

In this chapter we will examine in more detail some of the hypotheses which are to be tested in the experiment and expand on them and on the questions which we hope to be able to answer with the results (section 6.2). Further on, the set-up and the proceedings of the experimental sessions will be explained (6.3). Then the selection of subjects (6.4), patients (6.5), and diseases (6.6) for the knowledge base will be described. Section 6.6 is an explanation of how the subjects of these sessions were trained in the use of the DDSS before the experiment, and finally section 6.8 tells what information was collected during the experimental sessions.

6.2 Hypotheses and questions.

From the discussion of the diagnostic process in Ch. 3, and the general problem formulation in Ch. 5, a number of specific hypotheses and questions have been formulated, which as we have seen previously may be arranged under the headings of (1) data acquisition, (2) storage and retrieval of data, (3) the reasoning processes, (4) the knowledge base, (5) constraints of clinical practice, and (6) technical aspects. For each of these six headings we have formulated a number of specific points, to be tested in the experiment. In the text, "subjects" refers to the doctors who tested the system.

6.2.1 Data acquisition

During a session, the specific medical questions asked are in themselves not very significant, since they will differ with the doctor, the patient and the circumstances, and yet may refer to the same symptom. The important thing is that there should be no great discrepancy between the number of questions asked and subjects' ability to locate the corresponding item(s) in the SCS. From the number of questions asked estimates can be made of the relative difficulty of the cases, etc.

H4 The user is able to find the shortest paths in searching for specific items in the SCS tree.

Subjects should be able to locate the object of his question in the SCS; any path found from node a to node b in the tree is necessarily the shortest path. But users may be tempted to explore other menus merely by seeing them, discarding their original course.

There are at least two possible modes of questioning the patient using the DDSS with the built-in SCS: a) subjects may follow the structure of the SCS in determining the order of their questioning, so minimizing the number of actions necessary to encode the symptoms found in the system; or b) subject may elect to follow his or her own customary routine, and attempt to encode the symptoms in that order, which requires more navigation through the SCS menus. Or c) subject may use a mixture of both. This strategy may also change with the specific patient seen or with the experience of the subject with the system. This behaviour may be diagnosed by looking at the logged data of the symptom entry during the consultations. Analyzing different parameters with respect to this behaviour may give important clues as to how the system is working with the user.

H8 By the presentation of the menus doctors will question their patient more widely (more categories of symptoms and signs) than when unguided by the system.

Doctors would be expected to query their patients more widely when confronted with the DDSS menu system. "Widely" to be interpreted as asking for a larger range of symptoms. This can only be tested subjectively by asking the subjects; an objective comparison is not possible without a control group. However, a previous investigation has been done with the same patients under similar circumstances, except that no computer system was used. [Rid89] To a certain extent the results of both studies may be comparable.

H10 By the presentation of the menus doctors will be stimulated to probe deeper with their questions, and so to get a more detailed picture of the patient.

The SCS should also stimulate the doctor to probe deeper with his questions, and so get a more detailed picture of the patient. The level of detail of the data gathered may be gauged by looking at the patient record produced; an objective measurement of whether this is deeper than without he use of the system is not really possible without a control group, but we can ask subject's opinion about his behaviour in this regard.

6.2.2 storage and retrieval of data

Symptoms are the result of questions which "hit the mark", i.e. of questions which return an answer about an abnormal phenomenon in the patient. If the diseases as doctors know them, or rather the mental pictures the doctors have of diseases, were identical, the "not-having" of a specific symptom might be a useful indicator. However, as explained at length before, this is not the case. Between the question and the search for a symptom (-aspect) in the SCS there should normally be a relation: the presence of the symptom in the SCS. On the other hand subjects may not be able to locate a specific item in the SCS by virtue of its sheer size. This requires an understanding of its structure which the subject cannot be expected to come by all at once.

H9 Doctors will be able to encode all their observations within the SCS.
This is not the same point as is referred to in H4 in the previous section: that refers to questions asked, this one to answers recorded. There may be a discrepancy. This can be determined directly from observation of the subject during the interview, from his own remarks about it, and from secondary sources such as the use of the comment facility to annotate symptoms.

H6 The system is proof against accidental erasure of data.
We can observe if accidental data loss does actually occur in the experiment, and its cause.

H7 The use of colours will help the user to orient himself within the menu tree.
There is no way to test this except by the impressions of the observers and the opinion of the subjects themselves.

H11 Doctors are able to handle the standardised terminology without trouble.
We clearly expect subjects to be able to use the system and to locate nearly all symptoms under their own steam, not needing any help. They will be able to translate their idea into a coding within the SCS. This may be checked by direct observation during the experimental sessions: misunderstandings quickly become clear. Some areas may give more trouble than others.

H12 It will only rarely be necessary to make specific notes in the free-text field for specifications beyond the levels of specification within the SCS.

If the description offered by the SCS is sufficiently complete, doctors should only very rarely feel the need to add a comment. Comments are easy to count from the log-files; subjects may also be asked.

H13 The doctor will be able to retrace his previous steps of data acquisition at any moment in the process.

Symptoms entered and the time needed for this may have a relation with the difficulty of the case. Doctors will probably want to make use of the option to review the data they have gathered for this patient.

6.2.3 the reasoning process

H14 The doctor will consider more hypotheses in his decision making after having seen the system differential diagnoses.

The differential diagnosis list of the DDSS should lead doctors to review their conclusions for the patient. This could be seen to occur in three ways: (1) after seeing the differential diagnosis of the system, the doctor decides to ask more questions to the patient, or (2) changes his own differential diagnosis, or (3) assigns different probabilities to his list of possible diagnoses.

H15 In undecided cases doctors will ask for more symptoms in order to confirm or reject particular hypotheses.

Doctors might be expected to ask for more symptoms to confirm the diagnosis if they see the degree of concordance between their hypothesis and the symptoms. When a diagnosis scores low in the dd of the system, doctors might be expected to ask for additional symptoms. Subjects may research symptoms of the patient to different levels. (symptom -symptom-aspect, subaspect etc.) These parameters may have a bearing on the number and correctness of their diagnoses.

H16 By offering the degree of concordance (number of analogous symptoms) between the symptoms of the patients and symptoms belonging to a disease, the doctor is driven to ask for more symptoms of the patient (Confirmation = hypothesis given the evidence (Carnap).

When seeing a very low score for a diagnosis they have made, doctors are expected to reconsider a diagnosis that does not explain many symptoms. Diagnoses with low scores in the system DD might be expected to be subject to review sooner than high-scoring diagnoses. This information can be extracted from the log-files and the observer records.

H17 Doctors will use the option of looking up symptoms for suggested hypotheses to assist them in their justification process and to trace missing symptoms.

When confronted with the symptoms that make up a diagnosis in the database, doctors are expected to look for missing symptoms in their configuration. We may conclude that this has happened if a doctor examines the symptoms that make up a diagnosis in the knowledge base of the computer, then decides to ask for symptoms which are present in the description. This information may be extracted from the log-files.

The medical record present in the DDSS should make many errors retraceable. We were not able to ask the subjects to review their own decisions later. We may try to reconstruct why some conclusions went wrong by inspecting the patient record and the logged performance of the doctor.

6.2.4 the knowledge base

Is the present knowledge base adequate for the collection of patients seen?
What are the particular problems encountered in extracting textbook knowledge for the knowledge database?
What are the problems encountered when attempting to encode the symptoms found within the SCS? It should be possible to encode nearly all symptoms found in textbooks within the SCS. This is a first test of the completeness of the SCS.
Does the knowledge base contain errors in the description of the diseases;
Does it contain gaps in its description of the symptoms?

6.2.5 constraints of clinical practice.

These are of two kinds: the system should be easy to learn, and easy to use.

6.2.5.1 How easy is the system to learn?

H2 For users with some general computer experience the use of the system is easy to grasp.

After the three-hour training programme the subject should be able to use the system on his own, possibly with help from the manual.

The learning programme and the learning process for the test subjects should be scrutinized for points which appear to be difficult.
Use of the manual and assistance from the observers of the experiment should be carefully monitored.
It may be possible to draw some conclusions about how easy or difficult it would be to reach expert level. If we look at the distribution of the level of expertise in the subjects, and the level of those who have designed it, and who can therefore to be assumed to be intimately familiar with it, we may draw some tentative conclusions.

6.2.5.2 How easy is the system to use?

H1 Doctors do not find the use of the system disturbing.

How is subject's interaction with the patient influenced by having to work with a computer during the consultation? We cannot ask the patient, but we can observe the interaction in the experiment and we can also ask the doctors themselves.

H3 The use of the manual and assistance of the observers during the test is minimal.

For a user with some general computer experience the manual should not really be necessary in order to be able to use the system.

H5 The user has an intuitive feeling of what is going on in the system.

Use of the on-line help facility should be carefully monitored.
The sequence of events takes place in time. By keeping track of the time in minutes and seconds one can get a clear picture of what happens during the process. The many combinations of the observable elements of the process (questions asked, information stored, hypotheses formed, and conclusions) with respect to time are good indications of the problem-solving process of the subjects and of the working of the system. One of the primary hypotheses of the system is that it is usable during the consultation. This also means that it should be able to function within the time which a doctor

normally has available for a new patient. A question to be answered is how much time the support takes during a consultation.
Subject should feel in control of the operations.

6.2.6 some technical aspects

The system should perform correctly, giving reproducible and accurate results. Data storage, modification, and retrieval should be possible, while accidental data loss should be guarded against. Database integrity and consistency should be guaranteed. Any errors in this area should be recorded and analysed.
Breakdowns of the system should be carefully recorded and analysed in terms of technical (hardware) malfunction, software malfunction, and user error.
It must perform within a specific time, depending on the application. The time needed in typical and worst-case applications should be measured.
It should recover gracefully from error conditions, whether the error is made by the user or present in the programme itself. Any difficulties in this regard should be logged and analysed.
It should perform on a specific hardware platform, the specifications of which form part of the system. The hardware should be easily and cheaply available, and the system should be tested on a number of different machines.
The system should guard against user error, without becoming patronizing. Accidental deletion of data should not be possible. Actions associated with data loss should be confirmed.
There should be clear procedures for error recovery and restarting the system if the user ever encounters a fatal error.
Errors encountered during the first practical testing of the system will give an indication of places where improvement is desirable.

6.3 Description of the experimental situation.

For a first test of the system it is necessary to control as much of the environment as possible, so that we can concentrate our attention fully upon the doctor, his modes of reasoning and his interactions with the DDSS system.
In order to provide a constant and comparable reference, simulated patients were used. The location of the experiment was also invariant: a quiet room in the Institute for Family Medicine, not the subject's own surgery.

Each of 20 subjects was given 5 patients, randomly selected from a collection of eight. Every doctor had two hours to complete the series, but was otherwise free to divide his time over the individual patients. This gave subjects 20 minutes per patient, slightly longer than most doctors would use for a new patient. The extra time was included to compensate for the unfamiliarity with the system and for the time needed to fill in the several forms required by the experiment. If too little time remained to see the last patient, their consultation was not started. So, we expected to gather a hundred cases.
The simulated patient was acted by an experienced doctor. Subjects could take the history of this "patient" in the normal way, but the physical examination had of course to be replaced by specific questions (e.g. "What do I hear if I listen to your heart"), and laboratory examination results could also be asked ("Let's do a Hb"), the results of which would be given if available. X-rays and other specialist examinations could also be requested and with the photograph the conclusion of the interpreting radiologist would be given if any such results were available. Subjects were told explicitly before the start of the experiment that every patient had a 'real disease'.
They were encouraged to perform a normal consultation, different only from their customary routine in that they were asked to enter all patient data into the system. When either they were satisfied for themselves that they knew what was the matter with the patient, or when they had arrived at the point where they would normally stop the interview in their own practice, they were asked to write down their own conclusions. This was the first stage. After that they were asked to consult the diagnostic module of the DDSS, which showed them a list of diseases which might fit the symptoms entered. Subjects were then given an opportunity to reconsider their conclusions or ask more questions if they chose. When this second stage was over, the subject's final diagnosis was noted and the next consultation would start.
This way of using the DDSS is not the only one imaginable; in the experiment, we enforced the order of the proceedings by requiring that subjects first do a complete consultation before viewing the system's differential diagnosis. In this way it is possible to isolate the effects of seeing this dd from the rest of the consultation. We can also imagine a mode of use where after entering a few main symptoms doctors would ask for the system's "opinion" to give them ideas for hypotheses, then proceed to check these. This mode of use was not tested in the experiment.

6.3.1 Description of the doctor's role

The doctor was expected to use the system as it had previously been taught to him in two training sessions of one hour, but was otherwise encouraged to perform just as he normally would in his or her own consulting room. He was only required to enter the patient's symptoms into the DDSS system, and encouraged to write down any hypotheses which he entertained at any given time as soon as he became aware of them. He was encouraged to interview the patient in his or her own style, not using medical jargon in his questions. (As the patient was himself played by a doctor, this could be a temptation, but the subject 'did not understand' jargon questions.)

6.3.2 Description of the simulated patient

The patients were all played by a single person, himself a doctor with 14 years of experience as a general practitioner. Each simulated patient's data were minutely and fully described in a specially-prepared written record, with which the 'patient actor' was intimately familiar. The written record was immediately available for reference if the doctor should ask a question which the patient could not answer from memory. Every patient introduced himself or herself in exactly the same words to each doctor, whom it was assumed he or she had never seen before. The conversation then was allowed to flow naturally, according to the doctor's questions and the patient's reactions to them. At all times it was attempted to keep the conversation natural, with the patient just answering questions if the doctor had a closed-questions interview style, or opening up a little and volunteering information in response to a more encouraging attitude of the doctor. The patient would express information in lay terms, not understanding questions couched in medical jargon. The translation of these terms then was the subject's task. The patient would introduce extraneous (irrelevant from a medical standpoint) material from time to time, just as real patients sometimes do.

6.3.3 Description of the observer's role.

The observer could not be seen by the doctor, being seated behind his back. The same observer was present during every consultation. His responsibilities were the correct set-up of the system, the presence of the required forms, the synchronization of the clock in the room and the clock of the system, and the registration of the actions of the subject in so far as these

were not captured in the system log files. He kept tally of questions asked to the patient. Finally, only in case of malfunction or other unforeseen problems he would help the subject, each time carefully making a note of the problem and of the intervention required to solve it.

6.4 Subject selection procedure

Subjects were all doctors or advanced medical students who had links with the Institute for Family Medicine in Rotterdam. They were only eligible to participate if they had no previous experience with the system, and had not taken part in its design or implementation. Candidate subjects were approached personally and asked to participate in the experiment. No attempt was made to form a representative selection of general practitioners representative of GP's in the region, although deliberately subjects of different sex, age and experience, both medical and with computer systems, were invited.

6.5 Patient selection procedure

The patients were the same as had been used for an earlier research project by Ridderikhoff, and have been described in detail there [Rid85]. In summary, they are based on real patients, whose data have been carefully collected and recorded with the goal of a simulated patient in mind. Besides medical, also social information was gathered and recorded. The eight patients were deliberately selected to represent four relatively rare and four relatively common diseases in general practice. 'Rare' and 'common' in this context means that an average family practitioner will see a patient with such a disease about once a year or about four times a year.
Of these eight patients, each doctor was confronted with a random selection of five. The selection of five from the collection of eight for all subjects was made by a randomizing computer programme before the identity of the subjects or the diseases of the patients were known. The first generated pattern in which not more than two doctors would get all four 'rare' or all four 'common' patients was accepted. Patterns were checked for bias after the experiment in several ways: a) the number of times each patient occurred in the total series, and b) the average order of each patient in the sequence. (It would be possible that some patient nearly always occurs as the first or last patient in the sequence of a subject.)

6.6 Selection of diseases in the medical knowledge base

The diseases in the medical knowledge base would ideally have come from general practice, but no such information was available. Instead, we chose to include the diseases of the eight patients in the experiment, supplemented by a number of diseases which would appear in the differential diagnoses for each of these diseases. This gave us a list of 45 diseases. For these 45 diseases information about their symptoms and signs was extracted from a number of commonly used general and specialist textbooks on medicine. These lists were then translated to the SCS coding system by the author and entered into the DDSS knowledge base, without any knowledge which diseases the patients to be used in the investigation were suffering from.

6.7 Description of the training scheme of subjects

Subjects were trained in the use of the system in two sessions of one hour to one and a half hour each. These three sessions should fall within the space of two weeks ideally.
In the first session, a general orientation in the purpose of the system was given, with an explanation of the SCS coding system and examples of how to use it to code patient data. Subjects were shown the DDSS in action and given a first chance to get some hands-on experience with it. They also received the DDSS manual (35 pages) to read at home if they wanted to know more. This was optional, not required. A manual was also on hand during the experimental sessions.
In the second session, subjects were instructed in the basic principles of the user interface. They were given symptoms to encode for themselves, with assistance being given if problems developed.
Test symptoms in the training stage were chosen so as not to bias subjects for the cases they might have to solve.

6.8 Description of information gathered during the sessions

During the training sesssions and during the experimental sessions, information about the subjects, the programme and their interaction was gathered. This information can be divided into four categories: (1) questionnaires, (2) observation (3) patient data files and (4) computer log files.

6.8.1 questionnaires

Every subject filled in two questionnaires: one before the first training session, and one after seeing the last patient. The pre-training questionnaire was intended to give an impression of the attitude of the subject towards computers in general and their use in medicine, and his or her experience with personal computers. The last questionnaire was intended to gauge subjects' reactions to the system.
In addition, every doctor was asked to fill in a form for every patient during each consultation, on which subject was encouraged to write down any diagnostic hypothesis that was entertained, however fleetingly. Directly after each consultation the subject was asked to give an opinion on how hard he found the problem and how realistic the patient seemed to him or her.

6.8.2 observation

During the patient consultations the observer could see the patient's face and the computer screen. For every question asked by the doctor, a tally was kept on a special form, timed accurate to the nearest minute. We therefore have a record of the number of questions asked during each consultation and of the temporal relations between them, but not of their content. In addition, any failures in the hard- or software, subject errors in using the system, and points where subject needed help from the manual or from the observer were recorded. Finally, anything which struck the observer as remarkable was noted.

6.8.3 patient data files

The patient's medical records were stored in the computer in the normal way during the consultation as they would be during routine use of the system. They were not used in the analysis as the next item gave the same information with much additional detail.

6.8.4 log-file of user actions.

Just for the experiment a special version of the DDSS programme was used which monitored a number of specific user actions by writing them to a special log-file. Stored in this way were:

- the times of entry and exit of the different modules of the programme,
- any use of the "Help" key,
- any symptoms entered, and
- any deletions,
- any editing of data already entered,
- any use of the support module and the dd given

Every one of these was registered together with the time at which it occurred, accurate to the nearest second. (Before every session the clock in the room was synchronized with the internal system clock). This gave us an extremely accurate record of the way in which the subject used the programme. The extra time needed for the construction of these log-files is imperceptible and the system is not slowed down at all.

6.8.5 Analysis

The data were entered into a spreadsheet programme with extended graphics capabilities, (Borland's 'Quattro Pro') which was used for the analysis of most data. Since no complex statistical analysis was necessary, this was perfectly adequate for the purpose. Log-files for the sessions were extracted with different filter programmes to look for a large number of different occurrences. The filter programmes were written in AWK, a pattern-action language which is very suitable to this sort of application. [Aho78]

Chapter 7. Results

In this chapter we will look at the course of the experiment on a broadly chronologic and step-by step basis, examining and describing each part of the process as it occurred. In addition, hypotheses formulated in the previous chapters will be discussed in the places where such a discussion seems most appropriate, and an attempt to verify or falsify these hypotheses will be made.

7.1 Description of participants in the experiment.

7.1.1 Recruitment of subjects

Subjects were taken from the group of doctors and advanced medical students who have links with the Institute for Family Medicine of the Erasmus University, Rotterdam.
Requirements for candidate subjects were that they had no previous experience with the system, and had not taken part in its design or implementation. Twenty-one subjects of different sex, age and experience, both medical experience and experience with computer systems, were approached personally and invited to take part. The resulting group is rather mixed, as the statistics below show. One of the subjects had to withdraw after the first training session because of lack of time. Data for this subject were not taken into consideration in the remainder of the chapter.

7.1.2 Description of the group of test subjects.

Gender: Our group consisted of 20 persons, 7 women and 13 men.
Age: The average age of our subjects at the time of the experiment was 38 years, ranging from 24 to 63 years.
Status: Subjects were a mixture of trainee doctors and experienced general practitioners. (see table 1.)

Table 1. Medical status of subjects

General Practitioner	12
'basisarts'	5
trainee GP	2
last-yr med student	1

Subjects formed a heterogeneous group of doctors, allowing us to assess the qualities of the system in its interaction with doctors of different age, skill and experience.

The total medical experience of our subjects, measured in years, also differed widely, with a large presence of both the very experienced and the very inexperienced. (table 2) This gives us an opportunity to check whether medical experience influences several parameters of the diagnostic process and the subject-system interaction.

Table 2. Experience as a doctor of subjects. (years)

Exp	Count
0-1	8
2-4	1
5-9	3
10-14	1
≥15	7

Table 3. PC ownership among subjects.

PC	15
No PC	5

Most subjects owned a personal computer (table 3). The possession or non-possession of a personal computer enabled us to assess the impact of this characteristic on several aspects of using the system, e.g. using the mouse, keyboard, software. Many subjects had some experience with specific information systems according to the HIS model for general practitioners. It will be interesting later on to see whether such experience is of help in using the DDSS system - the expectation is that it will not, because of the completely different nature of both types of program.

Table 4. Computer experience of subjects.

0-24	4
25-49	7
50-74	6
75-100	3

Table 4 shows a broad measure of the experience of our subjects with computer use on an (arbitrary) scale of 0 to 100, computed from their own reports of familiarity with different types of computer programme (word processing, databases, programming languages, medical administrative systems, medical expert systems and computer games). Nearly all subjects had some experience with wordprocessing software, usually WordPerfect. Other programs were much less often used. Thus, most subjects had some experience with computer programs but only one or two considered themselves to be expert users.

Attitude of subjects to computer use in general practice

In the questionnaire administered before the start of the training process, there were some questions intended to give an impression of the attitude of our subjects to the use of computers in medicine and general practice. Subjects were asked to express their agreement or disagreement with the following statements:

"I believe the computer can be a useful tool, but just for administrative purposes." (0-disagree, 10-agree)
The average score for this question was 2.73 (sd 2.4, n=20).
In retrospect this was a bad question because logically, disagreement can be construed either to mean that subject did not think that computers could be useful at all or that they could be used for much more than just administration, but we believe the latter opinion to be the correct interpretation: computers can be used for much more than just administrative purposes.
The next statement was: "I believe computers in general practice can contribute to diagnosis" (0-disagree, 10-agree)
The average level of agreement with this sentence was 6.76 (sd 1.84, n=20), indicating a fairly general belief in at least the theoretical ability of computers to aid doctors in diagnosis.
With regard to the statement: "I would accept the advice of the computer if it had been proved that it could do something better than I" (0-disagree,10-agree),
the average score of 5.6 hides a sharp division of opinion: (sd 3.05, n=20)
While some unhesitatingly endorsed this opinion, others expressed strong disagreement, indicating an irrational attitude toward the machine. Even assuming that its judgement had been proved superior to their own under certain circumstances, they would not follow the machine's advice! This is

also interesting in view of the effects of the feedback of the system on the subject's actions, as discussed later on in this chapter.
When asked to express their opinion on the statement: "I do not believe that computers will ever be a competitor to the general practitioner" (0-disagree, 10-agree) the result, an average score of 8.7 (sd 1.2, n=20) indicates almost unanimous agreement.
The subjects were also asked to answer a multiple-choice question: "How do you view the introduction of computers in the daily work of the general practitioner?"

The responses are summarized in table 5. (Some subjects gave more than one answer.)

Table 5. Reactions to introduction of computers in general practice.

as a great promise	7
with hesitant optimism	8
with some suspicion	3
with great suspicion	0
as useless	0
as a threat to the nature of the job	0
otherwise	4

Table 5 shows that our subjects are interested in the introduction of computers in general practice, and have high expectations of this development, although in some this is tempered by a slight mistrust or a wait-and-see attitude. Open opposition to the introduction of computers is absent, which in our opinion indicates that computers have noiselessly found their way into the mental background of the profession; even those doctors who do not themselves expect to work with a computer are not opposed to the machine per se. This is also illustrated by the answers to another question asked before the experiment:

"Could you ever learn to trust the expertise of a computer programme in your area of expertise?"

The answers to this question are shown in table 6.

Table 6. Responses to the question: "could you ever learn to trust the computer in your own area of expertise?"

yes	12
maybe	7
probably not	1
never	0

This is perhaps also an indication of a more general change in attitude in society towards computers. Only a decade ago, it was quite common for computers to be the subject of anecdotes about people not getting paid for several months, or getting paid grossly incorrect amounts, or anecdotes where the inflexibility of computers was illustrated by examples of conditions which could not be handled satisfactorily by a database system, e.g. two unmarried people living at the same adress, or getting increasingly acerbic reminders for a bill for a zero amount. People often had a strong emotional resistance to 'being reduced to a number' in a database and had little faith in the outcome. This seems to grow less as the sophistication of the hard- and software improves, although new concerns are springing up as a consequence of the large number of databases in which data about a person are registered, and the more or less unpleasant possibilities which could result from combining the data in several databases. It is therefore important that the data of our patients are stored in the DDSS in such a way that medical information can be extracted anonymously.
We have seen that the test population consisted of doctors of different age and medical experience, and also with different levels of experience with computers. Their attitude to the use of computers in general practice was interested but only guardedly optimistic. Later on we will see whether the level of medical and computer experience made any difference in the process of learning the system or in the performance of the doctor-system team when confronted with the simulated patients.

7.2 The training process of our test subjects.

Teaching sessions were planned within one week of one another whenever possible, and the experimental session was planned soon after that.

Participants were instructed in two sessions in the use of the system. The duration of each session was one hour, but in more than half the cases participants needed less time to achieve a sufficient degree of fluency in the use of the system. This especially occurred in young participants and in participants who had much experience with personal computers. There was also a strong tendency of subjects to assume that they would be able to navigate the system without trouble, once the general principle had been explained and just a few trial runs had been made. We clearly observed that the logical structure of the system induced confidence in the subjects that they would be able to handle it. Sometimes this confidence developed so quickly that we could not quite yet share it.

Learning to use the hardware.

The keyboard.
The keyboard as a data entry device was familiar to most subjects, although some had previously encountered it only in the form of a typewriter.
Most subjects had some typing skill, although none had advanced beyond the two-finger stage, and some still belonged to the 'hunt and peck' school. This could be a slight handicap in the filling in of screen forms with names and adresses but it did not present difficulties once the support programme was running, as typing is normally completely avoided during the consultation.

The mouse.
Hardly any of the subjects had used a mouse before, as their experience with computers was usually limited to traditional, non - mouse-driven programs running on an IBM compatible machine in text mode.
Nevertheless, due to its intuitive function, using the mouse was not a problem for most users, although two or three older users who had no previous experience with a mouse had some difficulty keeping three different fingers each to a fixed button and tended to push each mouse button with the index finger instead. This forced them to look away from the screen, interrupting their train of thought.

Learning to use the software

The introductory screens
We had thought beforehand that the introductory screens, where subjects were asked to identify themselves or their patient to the system, would present the least difficulty. We had believed this because of the intentional

similarity of these screens with many existing standard applications. Surprisingly, these turned out to be rather harder to learn than the extremely straightforward SCS menu system. (For an example of these screens, see Chapter 4.) Especially the 'user-friendly' additions and niceties calculated to reduce user effort when filling in the screen forms, such as pop-up menus with codes to be selected for certain fields, or automatic repositioning of the cursor on the next field when the previous field is filled in, were often surprising to 'computer-naive' subjects and tended to disorient them. One possible explanation might be that although at least some subjects used software with a similar interface at this point in their own practice, the administration of these data would normally be done by the doctor's assistant, not by himself...
A form where the screen behaves as the paper in a typewriter probably presents the least surprises to the inexperienced user: he always repositions the cursor himself, and the machine makes no attempts to out-guess or correct the user when it is not absolutely necessary.
Despite the above, the screens of the programme did not present many problems to the user. Fields in the database screens sometimes had to be explained by giving an example (e.g. 'marital status' -> married, divorced, etc.).

The menus of the symptom coding system (SCS)
As there were only the three mouse keys to be remembered, navigating these menus was immediately obvious to nearly all users. Browsing through the tree of possible symptoms by using the left and middle mouse buttons came naturally because of the direct feedback of every action. Only for the act of entering a symptom of the patient in the patient record users occasionally had to be reminded to press the right button; confirmation of this action is given in the form of an unobtrusive momentary flash of the relevant text in another colour, easy to miss if the subject had not learned to watch out for it.

The structure of the symptom coding system (SCS)

Learning the structure of the symptom coding system (SCS) was more of a problem, because by virtue of its sheer size it could not be fully explored within the time limit of the training sessions. Here, teaching had to be confined to the construction principles of the system. As already stated in the introductory paragraph to this section, users very quickly picked up on the structure and gained confidence that they could work out for themselves where a specific item should be located. After having done just a few exercises, the task of data entry was perceived to be so easy that subjects

frequently lost interest. We encountered surprisingly little difficulty in teaching the subjects to use the system in this phase.

SCS - Special terminology

A few words of special medical terminology used in some screens of the SCS had to be explained to all users. These were words created for this application, e.g. the new word 'gezwelling' was introduced to describe any tumor or swelling regardless of its cause, as a combination of the two Dutch words 'gezwel' and 'zwelling', because of the malignant overtones of the first and the non-neoplastic overtones of the second. This usage is unique to the SCS and therefore has to be explained to every user. All in all, there are just a few of these special cases where a doctor cannot be expected to choose the correct item just from the context of his medical knowledge and the specific menu structure given. There were no cases where a subject did not understand the terminology used, after the few exceptions to generally accepted medical terminology had been explained. Of course, they might have used different expressions for the same symptom themselves, but the expressions used in the SCS were perfectly acceptable to them. This confirms our earlier hypothesis:

Doctors will be able to handle the standardised terminology of the SCS without much trouble (H11).

Later on, after the experiment, subjects also gave their own opinion about the terminology used (and the reader should note that this was after they had entered, on average, nearly a hundred symptoms using the system, most of which they saw for the first time in the SCS while doing this!)
In response to the question: "Did you find the terminology used in the SCS..." (0-easy, 10-hard) the average score of our 20 subjects was 5.5. This average conceals a wide range of answers, from 2 to 8.6 (sd 2.00). Although from this it appears that at least some subjects found the terminology of the SCS hard, we nevertheless cannot say that from the standpoint of the observer it was badly understood, as nearly all symptoms entered were, in our opinion, an adequate encoding of information received from the patient. Our own observations do indicate a number of mistakes and difficulties which occurred again and again in specific places in the SCS, which particular places therefore should be subject to review in a later version of the SCS. One aspect of the system software seemed to give many users some trouble: when browsing through the symptom tree they sometimes had to be

reminded that seeing a symptom on screen was not the same as entering it into the record for their patient, and that they had to push the third mouse button explicitly for that. Especially those users who needed to look at the mouse for this task tended to miss the short flash of the text on screen indicating that the item had been entered. Subjects regularly made use of the option to inspect the symptoms gathered so far, and reflected om the case while looking at it.
From all these experiences, teaching the use of the keyboard, mouse, database menus, SCS menus and the underlying structure of the SCS, we can fully confirm our hypothesis from ch. 5:

For users with some general computer experience the use of the system is easy to grasp. (H2)

Even for users without previous computer experience it turned out to be quite feasible to achieve an acceptable working knowledge of the system within the allotted two hours of training time.
A young subject, even if he or she had no computer experience at all, could be taught to use the SCS menu system correctly with the help of the mouse within 10 minutes, as we repeatedly demonstrated. Older subjects without previous computer experience needed more time and repetitive instruction, but all subjects reached a sufficient level of fluency to be able to take part in the experiment within the allotted 2 hours of instruction time. We shall also see the good grasp of the system users had when we examine their performance in the experimental sessions.

The system only assumes its users to be medically, not technically, trained. Persons who have such medical knowledge encounter few problems. We believe that this could be achieved mainly because the system was designed for routine medical use by people who know what practical medicine is about.

7.3 The experimental sessions.

The interval between the second teaching session and the experimental session was 12.0 days on average. Before the session was started, the paper forms to be used during the experiment were shown to the subject and explained. The subject was only asked to write down during the consultation: 1) any hypotheses, however vague, he or she formed during the consultation about the cause of the patient's complaints, and 2) the final diagnosis.

Additionally, each subject was asked to provide an estimate of the likelihood that any hypothesis formed might be true. For this estimate three values were obtained, once immediately after the hypothesis had been written down, once after the subject's own diagnosis had been made, and finally after having seen the system differential diagnosis. The estimates were entered on a 'Visual analogue scale', (VAS), an open bar of 10 cm length, representing a subjective probability estimate ranging from 0 to 1, the latter value representing full confidence in the hypothesis.

All in all, 99 consultations were done instead of the intended 100. Only one participant did not have enough time left to start on the last patient. This was caused by a misunderstanding: before the experiment, subjects had been explicitly informed that the patients they were about to see were in no way special, but such patients as they might expect to encounter in their practice. However, this particular subject seemed to regard the experimental set-up as a test situation where the goal was to trap him into making an error by putting especially treacherous cases before him, and so acted with much greater circumspection and caution than usual. This caused him to need much more time than the others. All other subjects easily completed their five consultations within the time frame set for the task.

7.3.1 The experiment: Starting the program.

As this was something most users had only seen done once or twice before, they sometimes had to be prompted which command to type ('DDSS') to start the programme. Otherwise, no specific difficulties were encountered.

Logging in.
The database screen on which the subject was invited to log in as a user of the system was negotiated without trouble, especially since the users had already filled in their forms during the training sessions and needed only to type the first letters of their name to arrive at the correct entry. The computer log files show that on average, subjects spent 29 seconds at this screen. This is not just the time needed to fill in the form, but includes time needed for explanations and for the subject to familiarize himself with the proceedings and the protocol. There was a slight problem with the user interface discovered at this point: some users paused after finding their entry in the database, then hit the <return> key while or after accidentally touching the mouse. Since moving the mouse has an identical effect to pressing the arrow keys within the program, this sometimes caused them to move inadvertently

to another user's record before confirming the selection by hitting the return key. Usually, subjects did not notice this error themselves. In 5 cases (out of 20 logging-in procedures) this mistake occurred, necessitating going back and re-logging in. Therefore, in future versions of the programme the interface should be modified to make this mistake unlikely.

7.3.2 Selecting the patient.

The screen where the patient was to be selected gave rise to no problems, other than the one with the mouse movement mentioned earlier. According to the computer log files, the time the subject spent in this screen was 75 seconds on average, but this is not an indication of trouble with this screen but rather a consequence of the tendency of many doctors to start questioning the patient with this screen still in front of them to provide background information about age, sex and occupation of the patient.
The subjects' own opinion about these first screens was obtained from a few questions of the the post-experiment questionnaire:

"What did you think of the filling in and the overview of the personal data of the doctor and the patient?" (0-easy, 10-hard)

The average score of 1.9 (sd 1.53) shows that subjects perceived the screens for entry of personal data of the patient and the doctor as rather easy.
In our own observation, the most difficult part of the system was the sequence of 5 keystrokes, including two decisions about saving or not saving the data and adding or not adding a text comment to the data, necessary to store the data of one patient and to go on to the screen to select the next one. Nearly every subject needed some hints from the observer at this point. In the course of the session most subjects did tend to become familiar with these keystrokes.

With regard to the selection of patients:
"How easily could you find your way in the menus to select a new patient?" (0-easy, 10-hard) The average score of 2.4 (sd 1.47) leads to a similar conclusion.
Finally, when subjects were asked "How easily could you find your way in the non-symptom menus?" (0-easy,10-hard) The average score was 3.4; (sd 1.69). These answers indicate that finding the way through these menus was not perceived as difficult by our subjects.

Presentation of the patient.
In order to set the subjects off to an equal start, each patient presented himself or herself, by mouth of the actor, in the same words to every doctor. Additionally, the text of the initial complaint was shown to the doctor on a printed note, to reduce the effect of accidentally differing emphasis. The initial presentation was modelled upon a common mode of presentation of a patient in general practice, where a symptom and one piece of additional information are volunteered by the patient as introduction to the consultation, e.g. "I have a bad pain on the chest and I feel dizzy, doctor".
This introduction usually immediately engaged the attention of the subject doctors, who had as a rule had very little trouble adjusting to the experimental situation, although some doctors tended to forget their role in the first few minutes and attempted to query their patient in medical terms, as in a discussion among doctors. As the 'patient' then started to complain about difficult words and pretended not to understand the questions, the subject was quietly and automatically coerced back into the intended role. Many physicians clearly enjoyed their role and played their part so well that they gave their 'patient' advice on diet, smoking, lifestyle etc.
On the whole, subjects perceived the patients as being quite realistic. Every doctor was asked to score the realism of each patient on a sliding scale just after finishing his consultation with that patient. The average perceived realism for all patients together was high: 80.9 on a scale of 0..100. This may not be altogether surprising as all simulated patients were indeed faithfully based on real cases; nevertheless, it is also a tribute to the acting performance of the doctor who played the patients. Perceived realism was slightly higher among experienced doctors than among inexperienced doctors (83% vs 77%). A possible explanation is that inexperienced doctors are less aquainted with the usual wording of complaints in the GP's office.

7.3.3 The consultations.

Although we had intended that every subject should enter each piece of medical information into the system as soon as it became available, it quickly became clear that that many doctors, especially the older, more experienced ones, found this too much an interruption of their usual routine. They sometimes shifted to another modus operandi whereby they would first question the patient, sometimes keeping notes on paper to keep track of their thoughts, and then enter all the symptoms in a batch. This behaviour, which we dubbed 'non-simultaneous data entry' occurred in 54 out of 99 consultations. The other 45 cases were performed in the intended 'simultaneous'

mode. Strikingly, all (young) doctors with less than 3 years' experience exclusively used the simultaneous style of data entry; the experienced doctors used either mode, often trying the simultaneous method for one or two patients and then falling back into what seemed to be for them a more comfortable style.
The difference between these two groups may be a consequence of the assumed greater flexibility of inexperienced, and hence younger, doctors, or alternatively it is possible that the inexperienced participants felt more in need of guidance and structure in their interviews, for which they looked to the SCS. Our impression is that the second explanation certainly played a role in a number of consultations. The experienced physicians tended more to regard the menus offered on screen as a distraction of their own train of thought, while the inexperienced doctors were led to use it as a source of ideas.

The non-simultaneous style of consultation defeats one of the original purposes of the system, i.e. trying to reduce or eliminate distortion of the observed facts through selection and memory recall errors. It was therefore interesting, and unexpected, to note that there was no difference between the two groups of simultaneous-style and non-simultaneous consultation in number of questions asked, number of symptoms entered, fabrication or unwarranted interpretation of symptoms, time needed for the consultation, accuracy of the diagnosis, or any other of the parameters we looked at. (Table 7)

Table 7. Differences between several parameters for groups using the simultaneous or non-simultaneous style of consultation.

	perceived realism of patients	Ease of symptom coding	# times help needed	# symptoms entered per consultation	# questions asked per consultation
Simultaneous	81.7	58.1	2.6	18.9	28.3
Non-sim	79.8	46.9	2.4	19.1	29.6

7.3.3.1 Questions.

The questions our subjects asked of the patients are our next object of scrutiny. Scoring the number of questions was not so straightforward as the reader might suppose, and some subjectivity came into this. Many doctors asked questions in quick succession, without waiting for an answer. In this case each question was scored separately. Sometimes they asked after several items in the same question. (e.g. "Do you have diarrhea and vomiting?") In

this case the items were scored as separate questions if they were sufficiently different. Sometimes the doctor posed the same question in two different forms, right after another (e.g. "Did you vomit? bring up your food?"). In this case only one question was scored. It was remarkable how often questions were asked to which no answer seemed to be expected because the doctor would continue with the next question without waiting for the patient to give an answer.
The full matrix with the number of questions asked by every doctor of the patients seen by that doctor is given in Table 8.

Table 8. Number of questions asked by each doctor to each patient seen

	doctors -> A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	T	U
A	-	19	-	-	-	-	-	17	38	27	28	-	17	18	27	-	28	-	26	40
C	13	-	18	-	-	-	-	-	24	-	19	19	17	42	15	-	17	38	15	11
E	-	23	24	13	41	48	28	-	42	20	29	41	15	-	-	36	33	23	-	24
G	16	19	-	9	37	71	16	16		10	-	-	13	15	19	47	28	-	-	-
B	20	33	-	26	41	-	13	34	37	23	-	23	-	46	-	31	45	34	21	-
D	27	-	75	36	-	76	-	19	-	29	30	28	-	-	37	-	-	33	-	-
F	22	-	33	-	39	58	23	62	-	-	-	-	29	26	-	56	-	36	11	21
H	-	28	33	13	41	-	16	-	49	-	24	38	-	-	10	36	-	-	28	22

Table 9 shows the number of questions asked by each doctor for his first to fifth patients.

Table 9. Number of questions asked by each doctor to the first ... fifth patient

doctors->	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	T	U
1	16	19	75	26	41	76	13	62	49	20	29	38	13	15	10	47	28	36	21	24
2	13	28	24	36	39	58	16	17	24	10	30	41	29	18	19	56	45	34	26	21
3	27	19	33	9	41	48	16	34	38	23	28	19	15	42	37	31	28	23	28	22
4	20	33	33	13	37	71	28	19	37	27	24	28	17	46	15	36	33	33	11	11
5	22	23	18	13	41	-	23	16	42	29	19	23	17	26	27	36	17	38	15	40

On average, each subject asked 29 questions of each of his or her patients. See Table 10.

Table 10. Average number of questions asked to each patient.

Pat	A	C	E	G	B	D	F	H
# quest	25.9	20.7	29.3	24.3	30.5	39.0	34.7	28.2

There were marked differences in the number of questions asked of patients between the group of relatively rare diseases and the group of the more common ones. The rarer diseases needed more questions to diagnose, with the exception of patient H. (A case of iron-deficiency anaemia)
This is in accordance with our model of the medical problem-solving process, because a) a doctor has the patterns for common diseases more at his fingertips than for rarer ones, and is thus able to produce a stock list of questions quickly, and b) he will sooner arrive at the correct diagnostic hypothesis, because the pattern is fresh in his mind, thus wasting less questions on hypotheses which are later discarded. Cf. Balla: "The strength of a physician's initial disposition to regard his patients as having a certain illness is directly proportional to the incidence of that disease in the population he serves." [Bal82] and "The difficulty of the cases was reflected in the percentages of of data per case and cues per case"; "Obviously, the difficulty of a case is reflected in the physician's task of data aquisition" [Rid93].

There was also a difference in the number of questions asked of patients depending on their order in the sequence of five. See table 11.

Table 11. Average number of questions and number of questions asked per minute as a function of the order of presentation of the patient.

Pat	#quest	Q/min
1	32.9	1.52
2	29.2	1.63
3	28.1	1.59
4	28.6	1.74
5	25.5	1.67

Remarkably, by and large the number of questions asked declined with every patient seen in a session. To account for this phenomenon several theories may be advanced, e.g. it is possible that it is an artifact of the experimental

situation: at first the doctors want to perform well in this extraordinary situation where they are being monitored, while for the next patients they are able to relax more and slip into their real routine; alternatively, increasing time pressure may prompt a more economical mode of data gathering. Finally, because the number of questions per minute stays fairly constant for consecutive consultations, we believe that possibly the doctor asks questions just to fill the time: while he is working with the program, he asks a few questions just to keep the patient busy. If the programme becomes familiar and does not need as much time as in the beginning, the extra chatter decreases. As we will see later, the amount of information gathered with the increased number of questions, as expressed in symptoms of the patient entered in the patient record, is zero.

This difference in number of questions to the first and last patients was not caused by an accidentally uneven distribution of hard and easy patients over the sessions:

Table 12. Average position of each patient in the sequence of 5 patients.

Pat	A	C	E	G	B	D	F	H
	3.2	2.9	3.9	2.9	2.9	2.9	2.4	2.8

Table 12 shows that when we average the position in the sequence of every patient for all sessions in which he or she was used, most patients end up close to the ideal average position of 3 (the average position we would expect to see if every patient occurs in position 1, 2, 3, 4, or 5 with equal probability.)

We posed a hypothesis earlier:

By the presentation of the menus doctors are will query their patient wider (more categories of symptoms and signs) than when not guided by the system. (H8)

This hypothesis cannot be strictly confirmed or denied because of the lack of a control group. However, during the experiment we frequently observed subjects scanning the menu and then asking a question, which seemed to originate from seeing a menu item on screen. In a number of cases this was

also explicitly stated or admitted by the subject. "Let me see, what else can we ask..."

After the consultations, we asked our subjects whether they thought they had gathered more information:
"Did you find that the menu-structure invited you to ask less or more questions?" (0-less, 10-more); To which the average answer score was 6.5 (sd 1.80), indicating that subjects assigned a modest increase in the number of questions asked to the system. This is in agreement with our own observations.

However, it is interesting that in our study the number of questions asked to the patients is somewhat lower than in the earlier study of Ridderikhoff, who used a closely similar experimental setup with the same patients, but without a computer providing menus. In this experiment, the number of questions asked to each patient was on average just over 34 for the group of family physicians, which group conforms most closely to the subjects in the present experiment. [Rid85, p. 207]. To account for this difference, several hypotheses may be advanced, none of which can be proved at the present time.
1) The difference may be merely a consequence of a slightly different scoring policy; as explained above, it is not always obvious how many questions a physician asked.
2) It may be that the circumstances of the earlier experiment, which took place with more people in the room because each consultation was also videotaped, made the doctor feel less at ease and so try harder to arrive at a correct diagnosis.
3) It may be that having to cope with the computer and the patient at the same time occupied the attention of the doctor, leading him to focus less on the patient's problem.
4) It may be that the attention-focusing influence of the system enabled the physician to gather the same information with fewer questions. In this context it is interesting to note that the information entered into the system remained nearly constant when averaged over the first, second etc. patient; but as we have seen the number of questions decreased rather strongly with each consecutive patient.

The order of questioning employed by our physicians fitted the design of the SCS; all used the customary medical order of history, physical examination, and lab tests and special examinations, although sometimes during the

physical examination or lab test stage they would fall back into history taking as a new idea crossed their mind. This confirms the second part of our hypothesis 5:

The user will have an intuitive feeling of what is going on in the system. (H5)

All subjects followed the traditional, 'natural' order of of events in their consultations: history, physical examination, laboratory results, diagnosis. This is the order that is also the most natural to follow when using the DDSS. However, one typical error pointing to disorientation of the user did occur from time to time: Sometimes when users were in the middle of physical examination, they would revert to the history-taking, and then attempt to encode history items as physical findings. This indicates a certain confusion, or lack of awareness of the difference between subjective evidence of the patient, and the subject's own observations of the patient. Otherwise, no specific difficulties were noted. On the whole, the users did have a good grasp of what was going on inside the system, which confirms the first part of the hypothesis. When asked after the experiment: "What did you think of searching through the hierarchical structure of the SCS (the actual symptom tree)?" (0-easy, 10-hard) the average result was 5.0, with a considerable disagreement in the answers: (sd 2.42, range (1.2-8.8) This indicates that some users subjectively found the structure of the SCS tree much harder than others. This self-assessment frequently did not coincide with our own, made with the advantage of being able to compare different subjects.
A similar result was obtained in answer to the question: "Did you find the terminology used in the SCS" (0-easy,10-hard)
average 5.5 (sd 2.00, range (2-8.6)).
Users had an accurate sense of the structure of the data involved, and correctly decided for themselves when they had made an error in input, or when to review the data they had gathered for a patient.

7.3.3.2 Time.

The time needed for the consultation of every doctor with every patient is shown in table 13; the time as a function of the first, second etc. consultation of every doctor is shown in table 14. Both tables show the time from the start of the consultation to the point where the doctor has made his diagnosis and filled in the forms up to that point.

Table 13
Time needed for every consultation, until doctor's diagnosis (minutes). (= Phase 1)

	doctors-> A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	T	U
A	-	14	-	-	-	-	-	11	15	13	15	-	11	15	18	-	20	-	18	19
C	10	-	11	-	-	-	-	-	8	-	9	19	11	26	8	-	14	16	16	12
E	-	11	13	9	24	23	8	-	11	13	14	22	14	-	-	14	18	12	-	13
G	12	8	-	11	13	29	6	13	-	7	-	-	14	13	10	24	20	-	-	-
B	12	17	-	23	17	-	10	21	14	14	-	14	-	30	-	16	23	23	20	-
D	14	-	38	21	-	46	-	12	-	14	16	20	-	-	24	-	-	17	-	-
F	9	-	14	-	23	23	10	37	-	-	-	-	20	20	-	31	-	29	15	16
H	-	13	17	12	22	-	8	-	15	-	12	18	-	-	11	19	-	-	24	17

Table 14. Time needed for consultation until diagnosis for first to fifth patient

	doctors-> A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	T	U
1	12	14	38	23	24	46	10	37	15	13	14	18	14	13	11	24	20	29	20	13
2	10	13	13	21	23	23	8	11	8	7	16	22	20	15	10	31	23	23	18	16
3	14	8	17	11	17	23	6	21	15	14	15	19	14	26	24	16	20	12	24	17
4	12	17	14	9	13	29	8	12	14	13	12	20	11	30	8	19	18	17	15	12
5	9	11	11	12	22	-	10	13	11	14	9	14	11	20	18	14	14	16	16	19

The average time needed for the consultation is shown in table 15 for the individual patients and in table 16 for the first to fifth patients seen. These tables show (T. 15) that for the patients with the less prevalent diseases the subjects needed more time, and also (T. 16) that there is a very clear tendency to shorter consultations as familiarity with the system increases.

Table 15. Time needed for the phases before (T1) and after (T2) consulting the diagnostic support function of the DDSS, averaged per patient.

Pat	A	C	E	G	B	D	F	H
T1	15.4	13.3	14.6	13.8	18.1	22.2	20.6	15.7
T2	1.5	1.4	1.3	0.7	1.1	1.7	1.5	1.0

Table 16. Average time needed per patient for the two phases of the consultation, for the first to fifth patients.

Pat	T1	T2
1	20.4	1.3
2	16.6	1.4
3	16.7	1.0
4	15.2	1.3
5	13.9	1.4

The average duration of a consultation, from the first presentation of patient to the conclusion, was 17.9 minutes. This is considerably longer than was found in the experiment of Ridderikhoff with the same patients, but without the use of the system. (11.25 minutes, sd 3.87). [Rid85, p 207]. Even when we subtract the 1.3 minutes the subject spent on average in our experiment on the system's differential diagnosis, an interval not present in the other experiment, the resulting 16.6 minutes is still longer. An encouraging observation is that although the time needed for the first patient was 20.4 minutes on average, by the fifth patient the time had dropped to only 13.9 minutes, which is just slightly longer than the time found by Ridderikhoff.
It should also be noted that in the Ridderikhoff experiment, the total available time per patient was shorter, possibly causing his subjects to hurry more.

The maximum number of symptoms entered into the system in one minute for every consultation is shown in the following figure. It shows the maximum number of symptoms entered in a single minute for each consultation, and hence the greatest speed of data entry reached in every consultation.

Subjects were able to find the items sought for in the menu system in a reasonable time. On average, the subjects asked 28 questions in 16.6 minutes for every consultation. Additionally, in this same time interval, on average 19 symptoms were entered in the database.

This shows that most users could reach a reasonable speed when entering symptom data into the system. Maximum speeds of up to 4 or five symptoms per minute were quite common. This indicates that the speed of the combined system/user team can be quite acceptable in our view.
Overall speed increased significantly from the first to the last experimental session, probably from increasing familiarity with the system. There were

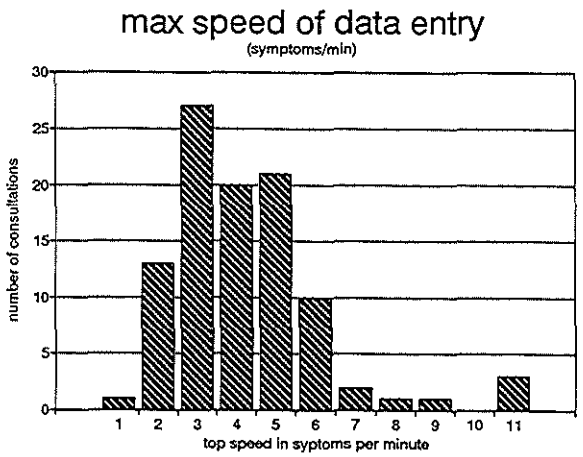


fig. 6 Distribution of the maximum speed of data entry for every consultation.

some differences between the first and last patients, pointing to a clear learning effect while using the system. (Table 17)

Table 17. Averages for several parameters as a function of the order of presentation of the patient: total time needed, number of questions asked, number of symtoms entered, depth of symptom

rank	time needed	# ques asked	#sympt entered	depth
1	21.7	32.9	18.6	6.38
2	18.0	29.2	18.5	6.31
3	17.7	28.1	19.8	6.42
4	16.4	28.6	18.8	6.37
5	15.3	25.5	19.3	6.39

We see that while the number of symptoms entered, and the level of detail with which symptoms are entered do not change with the order of the patient in the sequence, there is a clear trend toward shorter interviews, as experience with the system grows. At the same time, there is a decrease in the number of questions asked.

7.3.3.3 Looking up and entering symptom data.

If the physician had acquired an item of information from the patient, he should enter it into the decision support system. Those physicians who had followed their reasoning on screen by selecting the relevant menus as they asked questions could immediately click on the relevant item; those who had let the system be for the moment while concentrating on the patient's problem had to look the symptoms up first. Neither style presented many problems to the subjects. As we have seen earlier, there were few differences between the group who performed their consultations simultaneously and the one where the entering of data was done separately.

The subject of entering data brings us to our earlier hypothesis 11:

Doctors will be able to encode all their observations within the SCS (H9).

In a number of cases, the subject could not find an item in the tree.
The cause of this could be either that the item was indeed absent; or that its encoding was not sufficiently intuitive to the subject that he might locate it by himself. In such cases, the subject was helped and the offending symptom noted. A few items were consistently hard to find for most subjects, indicating a real problem with the symptom coding tree. Examples of such items are noises of different kinds in the lungs or the heart, or the results of tests of the neuromotor system, such as reflexes, or nerve root irritation. Rarely, a symptom was found to be completely absent from the tree: this occurred about 10 to 20 times in all. Therefore, more than 99% of symptoms could be encoded without trouble, given an accurate knowledge of the tree. Hypothesis 11 can therefore on the whole be confirmed.
One of the problems sometimes encountered was a confusion of the subject between a bodily function and its product: e.g. coughing and phlegm, miction and urine, etc. Once this was pointed out the correct category could always be found without any trouble.

The data entered were divided over the main areas of the SCS: history, physical examination and test results, as shown in table 18.
The medical history accounts for about 60 percent of the data gathered, while the lab and other special tests account for only a small percentage of the information. This is highest in those conditions where a specific test is decisive for the diagnosis, e.g. in hyperthyroidism and anemia. The patient with suspected gall stones was frequently 'sent to the radiologist' for diagnosis. Otherwise, our subjects reached a diagnosis with very little information besides history and physical examination. The idea that doctors are very keen on lab tests cannot be confirmed from these data.

Table 18. Relative proportion of different areas of information (History, Physical examination, lab results) in the patient record. (percent of total information)

Pat	Hist	PE	Lab etc.
A	54	43	4
C	64	35	1
E	65	21	15
G	49	49	2
B	64	34	2
D	52	47	1
F	48	36	16
H	58	25	16

Table 19 shows the average total number of items gathered and entered into the database for each patient. Patient D (ectopic pregnancy) is slightly above average, probably because more hypotheses were tested on average for this patient, but the other patients show a remarkably uniform distribution.

Table 19. Medical data items entered into the system, averaged over all consultations per patient.

Pat	A	C	E	G	B	D	F	H
count	21.1	18.4	18.1	18.2	17.1	25.0	18.3	17.2

Table 20. Number of data items ('symptoms') entered into the system as a function of the order of the patient.

Pat	1	2	3	4	5
Items	18.6	18.5	19.8	18.8	19.3

Table 20 shows that the number of medical data items entered did not vary appreciably with the ordering of the patients within a session, indicating that

the amount of information gathered did not vary with growing familiarity with the system, although the number of questions did decrease markedly, as we have seen before. This means that data gathering became more efficient as the subject gathered experience with the system; to elicit the same amount of information, less time and fewer questions were needed.

By the presentation of the menus doctors are stimulated to probe deeper with their questions, and so to get a more detailed picture of the patient (H10).

This is not a hypothesis which can be strictly judged for lack of a control group. Comparison with the earlier Ridderikhoff experiment [Rid85] [Rid87] shows that questions do seem to be more specific. The average depth of the symptom encodings is about 6.3, corresponding to a considerable level of detail. For example:

patient data/ history/ tr. circ/pain/ localisation/ thorax
patient data/ history/ tr. circ/pain/ since / minutes
patient data/ history/ tr. circ/pain/ irradiation / left arm

would be represented as three symptoms of depth 6 in the SCS.

Level of detail of the symptoms entered.

This is hard to measure objectively; however, because every additional level in the SCS tree represents a refinement of the information stored, we can compare the level of detail in patient records by looking at the average depth of the symptom tree for that patient.

Table 21. Averages of time needed, number of questions asked, number of symptoms entered, and depth of entered symptoms as a function of the patient.

Pat	A	C	E	G	B	D	F	H
time	16.9	14.7	15.9	14.7	19.2	23.9	22.1	16.7
# quest	25.9	20.7	29.3	24.3	30.5	39.0	34.7	28.2
# sympt	21.1	18.4	18.1	18.2	17.1	25.0	18.3	17.3
depth	6.57	6.30	6.34	6.46	6.26	6.51	6.33	6.29

There were some clear differences between individual simulated patients (table 21).

This table shows differences between patients with common and somewhat rarer diseases: the commoner conditions are handled more quickly, and with fewer questions. There is no clear difference between the number of symptoms entered and the level of detail of these symptoms in both groups, however.
Our subjects were of the opinion that with the use of the system there would be more data gathered and stored in the medical record than without it, as the answers to the next two questions show.
"Did you think the number of gathered data items increased because of the use of the system? (0-no effect,10-important effect)"
The average score of 4.9 is indicative of a considerable effect ascribed to the system by our subjects; but opinions differed considerably as the standard deviation shows. (sd 2.74, range (0.3-8))
"Do you think that by using the system more information about the patient will find its way into the doctor's record system?" (0-No,10-Yes)
produced a similar response: average 6.6 (sd 2.42, range (0.2-8.9))

Now, with regard to our earlier hypothesis :

The user will be able to find the shortest paths in searching for specific items in the SCS tree. (H4)

We observed that when exploring a symptom for the first time, subjects often looked under several menus when in doubt how a specific item would be coded. However, they learned very quickly, and because of the internal consistency of the system, were frequently able to extrapolate successfully their previous experience with the coding scheme to new symptoms. After gaining some experience with a particular part of the tree, they could usually find the required item without looking up wrong menus and having to retrace their steps.

The subjects themselves had mixed feelings about the search of the system for the location of symptoms:
"What did you think of searching through the hierarchical structure of the SCS (the actual symptom tree)?" (0-easy, 10-hard)
average 5.0; (sd 2.42, range (1.2-8.8))

Under-reporting
Some doctors did ask a considerable amount of detail but only entered a fraction of their positive findings in the system, causing a loss of sensitivity for the diagnostic module and an incomplete registration of the patient's information in the data record, e.g. in the case of subject G, there was a clear gap between questions asked and symptoms entered (cf. table 17). While on average, for every three questions asked two data items were entered, in 28 out of 99 consultations the ratio was below one item per two questions. This may be just a consequence of a low 'hit rate': since questions to which the answer is normal, i.e. which do not uncover some abnormality in the patient, are not registered, it is possible that few symptoms were found with many questions. But also in absolute numbers, the number of symptoms entered must sometimes be considered insufficient for the diagnostic module to produce a useful differential diagnosis: 33 of the 99 consultations ended with less than 15 symptoms entered, and 19 even ended with less than 10 symptoms. Although 15 data items may seem a lot, the reader should keep in mind that the specification of "dull aching pain on the chest irradiating to the left arm and chin, starting an hour ago" in the SCS would take up about 6 items. Any differential diagnosis based on so little evidence almost inevitably becomes shaky. By comparison, the number of data items present for the diseases in the knowledge base ranged from about 60 to over 200.
In diagnostics, the doctor always has the edge over the machine in the quantity of information he has available about the patient because although the machine can work with the same amount of data as the doctor, in practice it will always be less, because of conscious and unconscious selection on the part of the doctor what to enter and what to leave out, and of course also because of limitations to what is codeable within the system.

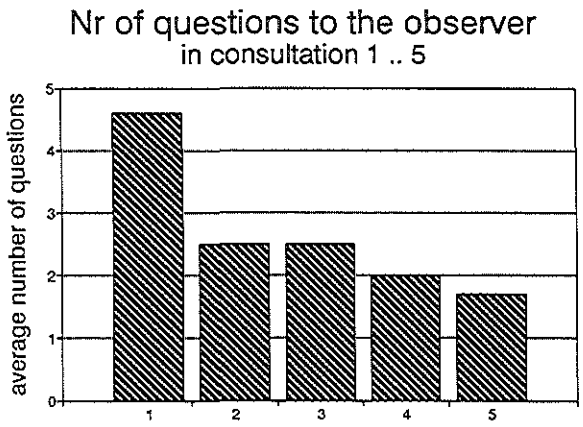
Over-reporting
Conversely, in the records of 13 consultations (13%), involving 5 different doctors, subjects were observed to encode symptoms which they had neither asked for, nor heard from their patients, thus fabricating evidence to match the idea of the patient's disease in their head! This seemed to occur spontaneously, and certainly without ill intentions. Even more frequently, subjects selected an interpretation of something they had heard from their patients which was more specific than they had a right to assume at that point. This illustrates the avidity with which the physician seizes on any point where he can construct a resemblance between the patient and the pattern he is trying to match.

Entering the patient's data in the electronic record system.
In nearly all cases, the standard method of data entry by mouse click sufficed. The subjects were aware of the option to view all gathered symptoms by pressing F2, and frequently made use of this option. Due to an oversight in the experimental design, the exact number of times this occurred has not been recorded, but it is estimated to lie between one and three times per session of 5 consultations. Sometimes the symptom editor was used to delete an item erroneously gathered, either by an accidental push of the wrong mouse button or because of a misunderstanding between doctor and patient. Such deletions were rare: 13 in 99 consultations, and out of 1879 symptoms collected in total.

Need for assistance with the program.
During the experiment, subjects sometimes asked a question to the observer regarding the use of the system, the structure of the symptom tree, or other subjects. These questions were noted down, but answered only if this was necessary for the natural conclusion of the consultation. Most questions were about the structure of the SCS tree: "where can I find such and such a symptom", or "how do I encode this symptom"; a few were about the software, e.g.: "how do I get from here to the next patient?" Usually this latter type of question could be solved by the subject without outside help from the observer; subject felt somewhat uncertain but could perform the desired action without any real trouble. Such software-related questions occurred mostly with the first one or two patients. As stated in the section on training, because of the size of the SCS, the subjects could not be familiarized with its full contents during the training sessions. In a number of cases, unfamiliar situations led to questions to the observer. Of 265 such questions in 99 consultations, 204 were related to coding an observation within the SCS.

Fig. 7 shows that the number of such questions was highest during the first consultation and declined steadily thereafter.

fig. 7



The manual had been given to every subject before the training sessions started, with the instruction that it contained information about how the system worked and why it worked that way, but that the subject would not be tested on the knowledge of its contents. Subsequently it was found, naturally, that practically none of the subjects had read the manual. Nevertheless it proved possible to teach the system without the use of the manual. During the experimental sessions the manual was not used at all although it lay on the subject's desk in a prominent position during the experiment.
Of course, the availability of the observer to answer questions about problems uncovered during use will inevitably result in less use of the manual. But on the whole, users were quite able to find out for themselves what they wanted to do and how to do it.
As assistance of the observer was mainly limited to the structure of the SCS, we are confident that in a longer experiment, after a short initial training period, the subjects could manage the programme perfectly well on their own without reference to outside help or even to the manual.
This largely confirms our earlier hypothesis:

The use of the manual and assistance of the observers during the test will be minimal. (H3)

'Minimal' is perhaps not the correct expression to use to describe a rate of several interventions per consultation; but the rapid decrease of requests for

help after the first experiences with the system had been made, from over 4 to less than 2, shows that this is not a structural problem.

Use of the free-text option.

For those cases in which the SCS did not provide enough detail to record the patient's data faithfully, there existed an escape option: by adding the required information as free text doctors could enter any information they wished to record. This option was used only 15 times for the 1879 symptoms entered in all consultations together (< 1%). This is another clear indication of the relative completeness of the SCS, and it also confirms our hypothesis 15:

It will only rarely be necessary to make specific notes in the free-text field for specifications beyond the levels of specification within the SCS. (H12)

Our observation agrees with this, in that we rarely had the impression that the attainable level of detail was considered insufficient by the subject. In the few cases where a particular item could not be coded satisfactorily, subjects did not seem to mind but happily continued with the next item. Most subjects were quite prepared to adapt their method of questioning to the structure of the SCS tree. Tolerance of the limitations of the system was high.

Use of the keyboard.
The keyboard was not used at all during the consultations beyond those few occasions where a deletion of an entered symptom was necessary or a free text note had to be added to a symptom in the SCS. The completely mouse-driven interface is therefore a realistic option for medical software of this type.

7.3.3.4 Hypotheses during the consultation.

Before the start of the experiment, subjects had been asked to write down any hypotheses they might entertain during the consultation, however vague these might be. Nevertheless, we observed that many subjects were relatively reluctant to write down such ideas. Sometimes they did not seem to be aware that they entertained any, even if to the observer there was clearly a purpose in the questions to the patient, pointing in a specific direction. In some cases, when gently prompted not to forget writing down any hypothesis he might

have, the subject actively denied having a hypothesis, despite just having asked after three symptoms characteristic of only one specific disease in quick succession. We cannot fully explain this behaviour. It is possible that although the subject does have a hypothesis, he or she is not fully conscious of it; or that the subject is reluctant to be pinned down on a specific guess before he is ready to believe in its correctness. After all, it is still quite common for students to be admonished 'to get the facts first' before attempting to fit them into a pattern.

Because of this observation, we must assume that the number of hypotheses entertained by the subject during the consultation is probably somewhat larger than our records show, especially with regard to hypotheses that are formed in the beginning of the consultation, and discarded before they take on much substance, perhaps at the next question.

We did observe that generally questioning certainly seems to be guided by the hypothesis currently at the top of the list. (Hypothesis-driven data collection) This has also been observed by many students of physicians' problem-solving behaviour. [Bar72] [Els78] [Rid89] We also noted that the behaviour of the subject yields strong clues whether any hypothesis is currently being entertained or not: the questioning in the first case is swift and purposeful, with little pause for thought in between. If no clear hypothesis has been formed, the questions are formed slowly, and without obvious connection even to the medically trained observer, seemingly at random. This situation was rare, but it was especially evident in some very inexperienced doctors, who might be expected to have a smaller stock of hypotheses in readiness. Once a clue had been obtained, they could swiftly shift into a purposeful line of questioning; but it often took them somewhat longer to get to this point than an experienced physician who seems to be able to generate a hypothesis on the spot for every patient presented. In the great majority of cases our physicians started a clear and purposeful line of questioning immediately after having heard the presenting complaint (early hypothesis generation).

In the words of Johnson-Laird and Wason: "The distinction between conscious deductions and everyday inference is probably a reflection of a more general contrast that can be drawn between explicit and implicit inferences. The inferences that underlie problem solving are often slow, voluntary and at the forefront of awareness; they are explicit. The inferences that underlie the ordinary processes of perception and comprehension are rapid, involuntary, and outside the consious awareness: they are implicit."

Table 23 shows the number of hypotheses written down for every consultation between doctor and patient. Fig. 8 shows the distribution of the number of hypotheses written down in the first phase of the consultation, i.e. before seeing the system dd.

Table 23. Number of hypotheses written down for each consultation, before seeing the system differential diagnosis.

doctors->	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	T	U
A		1						4	4	2	2		2	1	3		2		1	4
C	1		1						3		2	2	2	2	1		3	2	1	2
E		3	1	2	4	5	1		2	2	1	3	2			1	1	1		2
G	1	1		2	3	1	1	1		1			3	1	2	3	2			
B	2	2		2	2		2	4	4	1		3		3		2	2	2	4	
D	1		6	1		1		2		1	1	4			5			1		
F	1		1		6	5	2	5					2	1		5		2	2	1
H		1	1	2	4		1		4		3	3			2	3			4	1

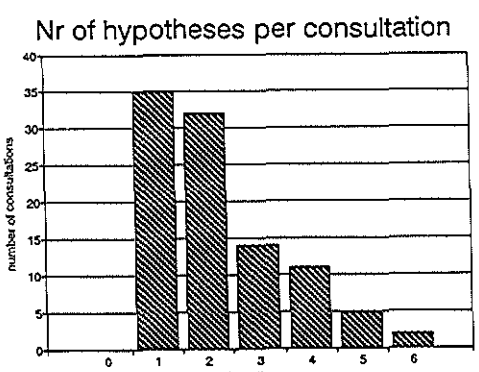


fig. 8

The average number of hypotheses noted down per consultation is only 2.25, which is low even if we take into account that some ideas probably were not registered. Ridderikhoff found an average number of hypotheses of 2.94 for a subject population which besides general practitioners also contained a few general internists, which latter group may be presumed to have increased the average to a degree.

Possibly also the instruction to write down every diagnostic idea was communicated with more force in that experiment.

In phase 2, i.e. after seeing the system dd, only a few subjects added a new hypothesis to their list. See Table 25.

Table 25. Number of hypotheses added by subjects after seeing the system differential diagnosis.

# added	0	1	2
# cases	85	12	2

An added hypothesis was ususally one which appeared high in the system differential diagnosis, and which had not been considered by the subject before.

The level of detail of hypotheses written down differed widely, ranging from the highly specific ("Myocardial infarction"), to the extremely vague ("something with the stomach", "psychosocial").
Confidence of subjects in their own hypotheses was fairly high; the average subjective probability for the first hypothesis at the time of writing it down was 57.0%, rising to 65.6% when the consultation was over. This rise indicates that the first hypothesis written down was very often also the final diagnosis.
If we look specifically at the hypothesis which in retrospect is going to be the diagnosis, the initial subjective probability (at the time of first writing it down) is 62.9%, the final subjective probability for the diagnosis is 82.4%, and after seeing the system dd this goes up another 2 percent to 84.3%.

In undecided cases doctors are expected to ask for more symptoms in order to confirm or reject particular hypotheses.(H15)

In the majority of cases doctors ended their consultation without having explicitly rejected all but one hypothesis. Usually, at least one other hypothesis remained, with an estimated probability which could be from only slightly to very much lower than the diagnosis selected. No additional attempts were made to increase the difference in certainty between competing diagnoses. This hypothesis must therefore be rejected.

7.3.3.5 diagnoses of subjects

At the end of the consultation, all subjects had made a diagnosis. However, usually not all other hypotheses considered along the way had been discarded: they still had a residual subjective probability, as shown by the indications of our subjects, and sometimes this subjective probability for an alternative hypothesis was only very slightly lower than that for the diagnosis. Thus, hypotheses were only rarely rejected. They fell into disuse, they might be forgotten, but they were not struck off the list. This is hard to explain if we assume a deductive model for the diagnostic process, but quite obvious if we assume the model explained in previous chapters, where the process of diagnosis is seen as an attempt to match a pattern in the doctor's head with symptoms of the patient. At any time, only one pattern is being examined; if the matching process does not develop satisfactorily a new pattern is generated and tested against the patient's data. Hardly any critical testing takes place; the doctor is quite able to believe with equal strength in two mutually exclusive hypotheses at the same time. Hypotheses which are no longer current recede to the background without being rejected, and can be re-activated at any time by a cue from the patient. As already mentioned above, the subjective certainty of our subjects of their own diagnosis, reached (with the help of the SCS menu system) before they had seen the system's differential diagnosis, was high: 82.3 %. This increased only minimally after seeing the system dd: 84.3%. This subjective certainty of our subjects was strikingly higher than in the earlier study by Ridderikhoff [Rid85]. He found a final subjective probability of 73.9% in his experiment with the same patients. The initial subjective probability estimate for the first hypothesis was 44.2% in Ridderikhoff's study, which is also lower than the 57% we found. The increase in certainty is similar in both studies, however.

Whether this was an effect of the use of the system or a consequence of the slightly different conditions of the experiment cannot be determined; in the current study doctors had more time available, and although the extra time was mainly needed to interact with the system, it may have had a beneficial effect on the subjective certainty of the subjects.

Quality of subjects' diagnoses.

When we strictly judged the subject's diagnosis for correctness, assigning 1 point for a completely correct diagnosis and 0 points for an incomplete or incorrect one, 43% of consultations resulted in a completely correct diag-

nosis. (Of course many of the diagnoses which we had to discard as incorrect for this measurement were approximately correct).

This did not differ between users of the simultaneous or non-simultaneous consultation style.
Experienced doctors, defined as doctors with more than 3 years' experience in practical patient care did not diagnose better than inexperienced subjects, with less than one year of experience. Likewise the number of actual diagnostic blunders, serious mistakes with possibly dangerous consequences, did not differ between experienced and inexperienced doctors. Experienced doctors asked slightly more questions, and found the paper patients slightly more realistic.

7.3.3.6 system diagnosis

The results of the system diagnosis are summarized in Table 26 below.

Table 26. Diagnostic accuracy of the system for the different patients. Only the cases where the correct diagnosis topped the list are counted.

Pat	Diagnosis	#1,%	N
A	MI	45	11
C	at. eczema	8	12
E	gallstones	53	15
G	asthmatic br	31	13
B	Ischialgia	77	14*
D	ectopic preg	0	10+
F	hyperthyroid	50	12
H	iron def. an.	42	12

*In one case the system dd was not consulted.
+The database representation for this disease contained an error, making it nearly impossible for the system to select this disease.

The table shows the percentages for the correct diagnosis at the top of the system's list. The bad performance for ectopic pregnancy was caused by an error in the database; the bad performance of the eczema patient was caused by the difficulty subjects experienced in describing the symptoms within the SCS.
The knowledge base consisted of descriptions of diseases closely related to the disease of the patient. The differential diagnosis given usually

encompassed a fairly large number of diseases, from three to fifteen. Most of these had very low scores for the number of matching symptoms between the patient and the pattern, except the first two or three. The correct diagnosis was to be found among the first three options in the system's list in the great majority of cases.

7.3.3.6 The second stage: reactions to the system differential diagnosis.

The second stage of the experiment was started when subjects had made a diagnosis on their own and formally confirmed that they were ready. They were then asked to press F4 to set in motion the system's diagnostic module. This they did with enthousiasm, often waiting with an extended finger above the key: "Can I press it now? Can I press it now?". They were very curious to see what the machine would do.
The most striking aspect of the experiment at this point was the extreme shortness of the second stage. Doctors on average spent less than two minutes, and many spent less than one minute, in contemplation of the system differential diagnosis. Perhaps related to their confidence in their own diagnosis, which was remarkably high, they were not inclined to take any other suggestions into consideration. Subjective probability after stage 1 is already so high (82%) that little improvement is possible. After the second stage, subjective probability has improved by two percent. Subject's confidence in his own diagnosis was observed to be high even if the diagnosis was wrong. In all 99 consultations, only 16 new hypotheses were considered in 14 different cases after the subject had seen the system differential diagnosis. In just one case this led to rejection of an incorrect diagnosis and acceptance of the top system diagnosis, which was correct. In only 6 cases, subjective probabilities shifted enough after seeing the system dd to bring another hypothesis to the top of the subjects' list. In three cases, the result was an improvement, in one case, a deterioration, and in two cases, the result was neutral. In several cases where the subject's diagnosis was wrong, and the system differential diagnosis was topped by the correct diagnosis, the subject did not even seem to reconsider his own opinion.
The typical pattern of subject's behaviour after seeing the system DD was a short glance at the screen, and if no surprises occurred there, no change to subjects' own conclusion was made. Our subjects seemed to take reassurance from seeing the system dd even if it was bad, perhaps because they did not see anything they should have looked at in their own decision-making process. We believe that we must therefore on the whole reject our hypothesis 14 from chapter 5:

The doctor will consider more hypotheses in his decision making after having seen the system differential diagnoses. (H14)

Although the option to do so was always clearly indicated to the subject, in only 14 cases, subjects started a new round of data acquisition after having seen the system differential diagnosis. This second round was usually confined to just one or two additional questions. If doctors considered more hypotheses after seeing the system dd, we certainly did not see much evidence of it, nor did subjects remark on it.
We also had entertained another hypothesis beforehand:

After seeing the degree of concordance (number of analogous symptoms) between the symptoms of the patients and symptoms belonging to a disease, the doctor will ask for more symptoms of the patient (Confirmation = hypothesis given the evidence (Carnap). (H16)

This phenomenon was not observed at all. The hypothesis is rejected.

Subjects answered a few questions about the system dd after the experiment:

"What did you think of the diagnoses offered by the system:" (0-superfluous,10-very useful)
The average score of 5.4 (sd 1.81, range (0.6-7.9)) seems to indicate that subjects found the system differential diagnosis rather useful, but this is in complete contradiction to their actual behaviour. Even in cases where a subject reached a wrong diagnosis, and where the system diagnostic support list was topped by the correct diagnosis, subjects did not change their mind about their own diagnosis, or asked extra questions of the patient after seeing the system diagnosis.

Score-numbers
A score-number was given next to every element of the system differential diagnosis, representing the degree of overlap between the patient's symptoms and the correcponding disease profile in the knowledge base.
"What did you think of the score-numbers given with the diagnoses?"
(0-superfluous,10-very useful)
The average score of 3.9 (sd 2.70, range (0.5-8.3)) is rather low; observations made during the experiment confirm that many subjects had hardly noticed these scoring numbers next to the diagnosis in the list. The three non-responders had not seen them at all.

Doctors will use the option of looking up symptoms for suggested hypotheses to assist them in their justification process and to trace missing symptoms.(H17)

This occurred very rarely. In no case did this lead to additional questions to the patient. The hypothesis must be rejected.

7.3.4 System breakdowns: hard- and software errors.

System breakdown ocurred three times, and in every case this was an instance of the same software defect: the mouse stopped working. This situation could be remedied every time by saving the state of the system, using the keyboard equivalents of the mouse buttons, and restarting the system. No data loss ocurred, and only a slight time delay of about two minutes was incurred.

7.3.5 User errors

User errors were scarce; in no case did this lead to any data loss. Sometimes users accidentally entered an incorrect symptom for the patient and decided to erase this. This occurred 13 times, on a total of 1879 symptoms entered. Some symptoms were erroneously entered twice, which has no effect at all but is an indication that the subject may have forgotten that the symptom in question had already been entered. Such duplicate entries occurred 28 times on a total of 1879 different symptoms. We see that users rarely made mistakes in the use of the system, and that these mistakes did not detract much from the overall efficiency.
The online help facility was used only three times in all 99 consultations put together.

7.4 Miscellaneous items

There remain a few points yet to discuss.
At any point, the F1 key, when pressed, brings up a pop-up menu with information about the options open to the user at that point. Meaningful information about the semantics of the SCS cannot currently be given, although the software has been designed with a view to keep this option open. The function of the F1 key is advertised on the bottom line of every screen in the system.

The presence of the observer provided an invitation to the subject to ask questions about the proceedings, rather than attempt to find out for hinself. This tendency was guarded against by only giving assistance when it was really necessary. Most problems could be solved by the subject on his own. The absence of a really context-sensitive help function, sensitive to the position of the user in the SCS tree, was regretted by several subjects.

The system is proof against accidental erasure of data. (H6)

No data were accidentally lost during the entire experiment; within the system it is possible to erase patient data, but if the user selects this option a special 'confirmation dialog box' with a warning pops up on the screen. It is currently not possible to erase other programme data (e.g. doctor information) from within the programme. To do this, maintenance staff can gain access to the database using standard dBase tools. It is assumed the user will not need to do this.

Currently, there is no standard backup procedure from within the system. This is not as it should be. Ideally, the system should automatically make regular backups of all data on a separate device, such as a tape drive, so that the system can be fully restored if a catastrophic failure of the hardware should occur. At the moment, the only way to make backups is by hand, copying all necessary files to another medium.

The use of colours will help the user to orient himself within the menu tree. (H7)

Although this feature is currently implemented in the system, so that e.g. the history-taking occurs against a blue background, and entering the laboratory values against a black one, we feel that the familiarity of the subjects with the system did not become great enough within the current experiment to establish this as a subconscious cue. None of the subjects remarked on it. None seemed to use it consciously, and mistakes where subjects wanted to encode history data in the differently-coloured physical examination part of the tree did still occur. Beyond these general remarks not much can be said about the effect of colour in the user-interface on the subjects.

The doctor should be able to retrace his previous steps of data acquisition at any moment of the process. (H13)

This is certainly possible; by pressing a function key the user can get an instantaneous overview of all symptoms gathered for the patient so far. This option was used fairly regularly by our subjects, indicating that they knew of the possibility. The hypothesis is confirmed.

Data gathered in previous consultations can also be reviewed, but this could not be tested because there were no previous consultations in our experiment.

7.5 An overall assessment of the functionality of the system

Having read all of the above, at this point we may attempt to evaluate one of the main hypotheses of the experiment, posed in chapter 5.

Doctors do not find the use of the system disturbing.(H1)

Design.

The presentation of the system and its parts should intuitively fit the user's needs. This means in our case that the user recognizes the different parts of the system, can use the methods implicitly present in its structure, knows how to select items, has an overview of the menu structure, can orient himself in the tree by means of the background colour of the screen and the path window, and knows how to get from one menu to another and why. Therefore it is essential that systems like this one should be designed by people who know what doctors actually do. We believe that the design of such a system by doctors, troublesome as the process can be when expertise in computer science is lacking, has shown that a new light can be shed on the user interface and the user-friendliness of such a system. The DDSS has sprung from a line of research into the decision-making processes of general practitioners and general internists. Part of its success must be attributed to the insight this has given in the processes of routine practice, as opposed to demonstration cases in journals.
Its standardized terminology is a necessary hurdle which, as it turned out, our subjects were well able to take.

Speed.
As should be clear from the section 'time' above, the system enables the user to work with near-normal speed during a normal consultation after only five

patients have been seen. Further improvement is to be expected with increasing experience. The response of the system is fast enough on modern hardware never to give the user the sense of having to wait; all screens appear instantaneously.

Ease of use.
Especially the almost complete avoidance of typing while using the programme will recommend it to the physician-user. Increasing familiarity with the screens will make operation nearly automatic after only a short while.

Immediate feedback.
The advantage of being able to consult the software while the patient is there, and without having to get up, pick up a book, or sit in another chair should not be minimised; they are prerequisites for any system if it is to be accepted in general practice as the daily companion of the general practitioner.

Learnability.
Doctors as a rule do not have much spare time; it proved difficult sometimes in our experiment to find an open space in the diaries of our subjects. Yet we were able to teach them the use of the system within two hours, to a level where they could almost handle it on their own.

Accomplishing the task at hand.
In the experimental setting, all users were able to proceed satisfactorily with their normal problem-solving process: the initial presentation, initial information-gathering leading to hypothesis generation, leading to more information gathering, sometimes more hypotheses, more data-gathering, and finally a conclusion. All subjects were able to integrate the use of the system with their own problem-solving process. However, especially the older subjects had a strong tendency to tackle the experimental situation by splitting their task: first they would concentrate on the patient, then on entering the data into the machine, after which they would return to the patient again, etc. All subjects reached a normal conclusion of their consultation, using the system.

Opinion of subjects
For any system to be acceptable it must of course be accepted by the people who work with it. What did our freshly introduced users say after having had hands-on experience with the system for less than three hours?

"In what degree did you find using the system disturbing for the consultation?" (0-not disturbing,10-seriously disturbing)
results: average score 6.5 (sd 2.53, range (2.2-9.8))

On average therefore, the subjects did indicate a considerable level of disturbance to the normal doctor-patient contact because of the use of the system. Interestingly, our own observations indicate that there was a rather poor correlation between subjective judgement and objective performance here: some subjects who, in our opinion, were very fluent in the use of the system indicated that they found its use very disturbing, while others who seemed to us to have much more trouble with it did not indicate so in the questionnaire.

Our final conclusion with regard to hypothesis H1 is that although every effort has been made to keep the system as simple as possible, to a beginning user, especially an older beginning user without computer experience, the use of the system while doing a consultation simultaneously is still a strenous task.
On the other hand we do believe that practice makes perfect, and that future users of the system will be able to achieve a fluency through practice which will make use of the system nearly automatic, freeing the attention of the user from the details of the operation of the machine and enabling him or her to use the system to assist in the making of a diagnosis, almost unconsciously supporting the mental processes of the user instead of hindering them. Probably such users will be people who have familiarized themselves from an early age with computers and automated systems, unless an even simpler mode of data entry can be devised - voice selection of menu items on screen?

Chapter 8. Discussion and conclusions

8.1 Introduction

The fallibility of human judgement has always induced people to invent and develop schemes and rules to improve their decision-making, or in other words, to maximize the predicted outcome. "Prediction is difficult, especially of the future" Bohr said cynically. Since the capacities of computers came within range of applying sophisticated mathematical procedures to complicated problems, scientists have tried to develop programmes to support people in optimally predicting outcomes.
It appears very attractive to combine the capabilities of human reasoning (with a fallible memory) and the perfect memory of computers (with rudimentary intelligent reasoning capacities). But the question then remains whether these two elements can be connected and produce useful results, results at least slightly better than the performance of an average human.
This now still remains an illusion, partly because the cognitive processes of human beings are only incompletely known and the processes and systematics of computers simply do not reflect those of creatures of flesh and blood.
Exploration of decision support systems has mainly been focused on the application of statistical and mathematical methods to processes which were supposed to mirror, or at least partially reflect, reality. As a consequence of the use of sequentially operating computers for such models, sequential data processing was also assumed to occur in the formation of judgements in human beings. In this line of thought Ledley (a mathematician) and Lusted (a radiologist) [Led59] formulated 'reasoning foundations of medical diagnosis' for more structured diagnostics. This publication was the start of the development of a large series of computer programs implementing various types of decision-making for the medical profession.
More than 95% of all systems created for diagnostic support did not survive the voyage from laboratory to clinic. [Lun87]
"I know", Feinstein [Fei77] said, "of no published work, or clinical setting, or specific world situations, in which statistical methods have made a prominent contribution that could not have been achieved just as easily without statistical formula."
Croft [Cro72], testing 10 of the most commonly used mathematical diagnostic models on the same large data set, found their diagnostic accuracy to be more sensitive to variations between diseases than to variation between the models. He recommends to future researchers that they should not continue

with increasingly sophisticated mathematical techniques, but instead concentrate on the real obstacles in practical medical diagnosis.
"Instead we should concentrate on what people actually do and develop descriptive models to account for decision processes", said Howell [How82]. Any investigation that does not take into account the real situation in medical practice is likely to fail as explanatory of medical methods. Or, as Biörck [Biö77] stated: "This type of investigation has to be done by people who know what doctoring means". It has to be adapted to the level of health care at which the doctor really makes his decisions. Taylor [Tay76] wrote in 1976: "It is, therefore, more promising to begin projects of this kind with an analysis of the decision made by the physician 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 physician who will use it."
The present study started with an in-depth investigation of decision-making behaviour of general practitioners and general internists [Rid89].

8.2 The philosophy of the DDSS

The majority of developers have often ignored the many human, contextual and cultural factors that determine whether a new system will be accepted by end-users [Lun87]. As many support and expert systems (the difference is irrelevant in this context) have been developed without identifying the real barriers, needs and desires of the target population, the initial enthusiasm for these types of system subsided. Attention shifted towards simpler systems such as computerized administrative systems.
Benefiting and having learned from past failures we developed the Diagnostic Decision Support System (DDSS) with an open eye to the daily work of the general practitioner.
We started from the theory of problem-solving as formulated by Newell *et al.* [New58].

a) it should predict the performance of a problem-solver handling specific tasks;
b) it should explain how human problem-solving takes place;
c) it should indicate what processes are used;
d) it should indicate what mechanisms perform these processes;
e) it should predict the incidental phenomena that accompany problem-solving, and the relation of these to the problem-solving task;

f) 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;
g) it should explain how specific and general problem-solving skills are learned, and what it is that problem-solver "has gained" 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.

From this we could formulate our own conditions for DDSS.

1) Use of the system whilst seeing patients. "Practice makes perfect"; a repetition of activities leads to perfection in performance and subsequently to acceptance of the system as a valued support in decision-making.

2) Acceptability depends also on the (perceived) complexity of the system. Most people seem to be afraid of handling a computer. Quick and easy learning are prominent assets for a system that wants to gain acceptability.

3) A prerequisite to easy learning is a very effective user-interface. We designed the interface with great care, profiting from suggestions of various people, among whom were colleagues and students of the Technical University Delft and lecturers [Rub87].

4) No more typewriting! In order to enter data the physician has often been required to take part in a long and complicated technical dialogue, not to mention the impediment of a keyboard, a tool designed and developed during the 19th century, to accommodate 20th century equipment.

5) The system should be adapted to the usual procedure of data acquisition of the doctor, which means the routine of medical history followed by physical examination, etc. It should also be capable of following every sudden jump in the train of thought the doctor may (and certainly often does) make.

6) The use of a standard terminology. If typewriting is to be avoided and feedback is to be provided, standardization of terms is a logical consequence. Most problems in implementation of this type of information systems stem from a lack of standard medical definitions [Cro72].

7) Completeness of terminology, finding everything in the system the doctor wants to encode. Our design of a standard terminology encompasses a newly developed structure for medical data presentation (symptoms, signs, tests, etc.). Whether the user is able to find everything he wants to encode is one of our main questions in the study.

8) Reliable and valid registration of patient data. As medical items served as entries for data registration we wondered whether the doctor and the patient could easily agree on certain entries. Patients who can check certain interpretations of the doctor will certainly help to validate the patient records and clinical data bases [Gar78]. Patients should actually be encouraged to participate in the communication. Electronic registration also carries the advantages of unbiased and complete recording.

9) Easy recording and support screens may invite the doctor to acquire more and more detailed data, and to explore additional organ systems. In a recent study Ridderikhoff [Rid93a] found that patient-physician interviewing is laborious, time-consuming and ineffective in eliciting specific data. Studies in problem-solving [Bar72] [Blo80] [Els78] [Gro85] [Rid89] have shown that physicians start narrowing their vision very early in the diagnostic process which detracts from optimal data acquisition and decision-making.

10) The system should support the usual reasoning processes of the practising doctor. Whereas Ledley & Lusted [Led59] advocated a sequential process of logical steps, others have found different descriptions of doctors' inference processes. The hypothetico-deductive method was assumed to be "a nearly universal characteristic of human thinking in complex, poorly defined environments" [Els78]. This stance was challenged by McGaghie and others [McG80] [McCo86] [Rid91]. We followed the conclusions, based on direct observations, of the earlier study of decision-making [Rid89], which

established inductive reasoning as the leading method of inference [Rid93b].

11) The diagnostic procedure is mainly a matter of pattern recognition, [Eng63] [Cro74] [Rid91] where the pattern is viewed as a configuration of observable phenomena in the patient. These configurations may vary from patient to patient but also from doctor to doctor and from textbook to textbook. If the diagnosis is primarily viewed as some composite of symptoms and signs it can be determined by pattern matching. The diagnostic judgement can then be quantified by the completeness of the match between the patient's data and the data of comparable diseases in the knowledge base.

12) If some overlap exists between the selected diagnosis and a number of diagnostic alternatives with comparable symptom configurations then the doctor may come to reconsider, revise and rethink his judgement in the light of the additional information. He may decide to collect additional data, to order specific tests or perform special physical examinations. This can all be done whilst seeing the patient. Our maxim therefore is: Rethink - Reconsider - Retry - Recollect - Restart.

13) The system should be highly transparent, transferable and maintainable. It comprises the use of modern and flexible computer languages, a transparent structure, easy updating of shell and medical contents, and a modular construction. It must run smoothly on globally available hardware and must, if necessary, be adaptable to existing information systems in the general practice environment. It must give the user the feeling of being permanently in control.

14) It must ensure the feeling of privacy for the patient as well as for the doctor. On the other hand, there exists an opportunity to profit from the huge amount of empirical data and experience contained in the patient records and the physicians' data bases. With a strict guarantee of anonymity it may be possible to use this information source in the future.

We shall discuss these matters in more detail.

8.3 Acceptability of the DDSS

Many information scientists complain about the low grade of acceptance of decision support and expert systems in the medical world [Sch79].
Underlying this complaint is the tacit assumption that information science has much to offer to the practising physician. Whether this is true remains to be seen. Many systems are built to capture, encode and emulate the knowledge of human experts. But this expertise is primarily focused at the specific domain of the (group of) expert(s), which may widely diverge from the daily practice and the needs of general practitioners. Other systems require an intensive dialogue with the doctor which may hinder him in his finely-tuned observations which may explain the patient's behaviour and illness. Many doctors feel annoyed when the computer is described as an "artificial partner"; a partner, however, that does not exhibit intelligent behaviour, that does not have opinions. The doctor who looks to his metallic companion for an opinion, and who really thinks it is an opinion, might tend to forget that he is reading is a certain input designated to represent data according to rules of transformation which the programmer has chosen and built into the operation. [McM79] How the input data have been gathered is not initially a concern of the designers and programmers of such systems. Only when the main engine of the application is finished do they find that the results are very dependent on the quality of the data to be entered, and that the data gathered for any patient may differ considerably between observers.
Likewise, most systems have been designed for use in an experimental setting, where the doctor comes to the machine in a context in which he has the complete data available, and tests them on the machine to see if it comes up with anything. In other words, the problem has really already been solved before the machine has even entered the proceedings. The experimental and laboratory settings in which most systems have been designed might be an important factor in the explanation of why these systems are so little used in practice.

We have taken the stance that it is the user who decides over materials, procedures, decisions and outcomes. The user is free to enter those data which he thinks appropriate, to follow his own lines of thought, to question the patient in his own customary style, to process data particularly chosen for the occasion, and to choose among various options the one he thinks is most plausible. He decides, he is in control, he is the sorcerer and not the apprentice. Any decision he makes is his responsibility. The DDSS deviates

definitely from systems that dogmatically offer advice which is neither sought nor asked for [Cro72].
The discrepancy which sometimes crops up between participants' opinions (in response to questions in the questionnaire) and their actual performance with the DDSS might be explained by the negative image and bad publicity of such systems in the medical world. These adverse reactions will certainly subside when users get used to the system. The relative fluency with which they handled the machine and the system creates optimistic visions.

Several other factors determine the acceptability. Among these is certainly the factor 'time'. Time is the doctor's most precious commodity. It is therefore of the utmost importance that the time for a consultation using the system does not extend beyond the normal duration of patient-physician encounters. Previous statistics of experimental encounters under very similar conditions [Rid89] showed an average consultation time for initial visits to general practitioners of approximately 12 minutes. In our study the average appeared to be 18 minutes, but the time decreased rapidly with increasing experience with the system. Participants who had some experience with computer use scored within the normal range for routine consultation. Besides, it is not realistic to expect doctors to be fluent in the use of the computer and the structure of the SCS after two hours of training. The results, however, are very encouraging and our expectation is that it is indeed possible to use the system within the routine time schedules of daily practice.

These encouraging figures are partially due to our principle: No more typewriting! Of all the factors which are seen as a barrier to data processing the doctor believes typewriting to be the most serious one. Current administrative systems in general practice are most commonly used by the doctor's secretary. This leaves the medical module, if incorporated, somewhat neglected and certainly not a useful tool for scientific exploration. Writing and processing patient data with one mouse click appeared to be exhilarating to the participants. The sheer rapidity of this option (up to 11 items in a minute!) seduced them to consult the patient record frequently. The easy editing of this record gave the participants a great deal of confidence.

Next in line of acceptability is the quick and easy learning. Underscoring our statement about the doctor's secretaries' use of the administrative system is our observation that participants (of whom 75% possessed their own computer) had most trouble with what we beforehand had assumed to be easiest

part of the system: the introductory pages in which the doctor's and patient's identification and some demographic information is to be filled in. Just those few lines of typing, the use of the keyboard and some notion of the screen design (which, by the way, was very much comparable to that of current administrative systems) appeared to be a greater obstacle than the hundred thousand nodes of the SCS. An important time-saving can be reached in practice when these forms are filled in by the secretary beforehand so that the doctor may fully concentrate on his actual task.

The remarkable ease of learning the various categories, menus and details of SCS will be discussed in the next section.

8.4 Standardization of terminology

"If there is one single change which would assist the development of information science in medicine it would be the adoption of a standard, predefined terminology amongst doctors everywhere" (de Dombal, [Dom78]).

At any point during a process of developing a device specifically meant for the practising physician one has to face decisions between congeniality and practical demands. Congeniality means adaptation to the idiosyncrasies of every individual doctor; practical design, however, demands that the uniformity principle be adhered to that all doctors collect the same information from identical patients and express it in standard terms and formulas. Any support system presupposes generality of meaning and concepts across users, nations and cultures. As was outlined before, this assumption cannot hold for practising physicians. As Galen [Nut81] in the 2nd century pointed out: "Cleavage of opinion comes from the failure to distinguish between the particular and the general. This schism is less one of doctrine than of method; it is one between diagnosis in terms of experience and diagnosis in terms of reasoning."
As a support system is specifically meant to support doctors in their routine work, the system must be moulded to their procedures and not vice versa. But as these processes have, in their view, a unique and individual character, we are kept in a quandary. A generally applicable support system requires a standard and uniform concept and standard and uniform terminology. To achieve this a number of conditions have to be fulfilled such as:
- a strict definition of the medical data;

- the medical data have to be standard or to be standardized in nomenclature and meaning;
- diagnoses have to be defined as configurations of symptoms and signs in their standardized terminology [Blo78].

Our solution to this problem was based on two ideas:
a) the human body has a limited range of expressions for pathological stimuli, and
b) questions to elicit medical data from the patient, can and should be adapted to the circumstances, while the items of medical significance elicited can be standardized. Standardized questionnaires are useless.

By carefully screening medical textbooks and dictionaries, a database, containing most symptoms and signs which the diseased body can produce, was composed. This database was organized according to principles of hierarchical ordering, more or less in conformity with ideas of Gross [Gro77]. The system must contain sufficient detail to allow doctors to register almost all medical information used in practice without typewriting.

Questions convey a special meaning. When we collect data we do so with an implicit or explicit idea in mind [Pop77]. By defining nearly all possible symptoms in a detailed system of description, standardization of registration can be reached without restricting the form in which the questions are asked, as happens in systems which employ a questionnaire model for data entry.
If the answer can only be a specific item in the SCS its recording can be ensured in a standard way. If we once have a patient record consisting of such standard items, a matching procedure with aggregates of likewise standardized items, contained in a medical knowledge base of diseases, can be accomplished, provided that the knowledge base was specifically created for this purpose.

The test with the SCS consisted of three elements.

1) The completeness of the mapping of symptoms and signs.
2) Electronic registration and editing of patient data.
3) Pattern-matching between patient data and the disease-patterns within the knowledge base.

The criteria for evaluation were:
- the number of information items gathered per patient;

- number of items not codeable within the system;
- ease of search for specific symptoms;
- use of the built-in notepad for any particular information of the patient.

The measures taken to facilitate the search for particular items in the SCS have contributed much to the success of the system. The various categories in which the clinical encounter can be distinguished: social and medical background, medical history, physical examination, tests, and treatments, is not only familiar to the doctor but makes his choice very easy. The classification of the category of medical history by organ system, and of the category of physical examination topographically, was one well-known to every participating doctor and guided him/her easily through the tree. As the successive menus at similar levels are consistently analogously and repetitively structured and the various items alphabetically arranged, the doctors learned to find their way so easily that some overconfidence could be observed after the first lessons.
Since the participant had the opportunity to formulate his questions in his/her own habitual way (e.g. participants easily switched from a interview style as between colleagues to one between doctor and patient) he could follow two methods of interviewing the patient: 1) guided by the system and 2) following his routine method. Although we preferred the first, more direct approach, the possibility of adopting the second strategy gave the testing procedure an additional measure of realism.

The Symptom Coding System must be regarded as one of the great successes of the project. What many had viewed as an overwhelmingly difficult or even impossible task [Pau76], the construction of a system to encode the patient's data in a quick, flexible way with a limited vocabulary, turns out to be quite possible with the rather simple tree-based hierarchical approach. For 20 doctors in 99 consultations with 8 different types of patient to enter nearly 2000 symptoms using this system and yet to find less than 20 instances where something they wish to enter cannot be encoded, is an amazing score for a system which has been constructed according to just a few guiding principles. Even if other workers in the medical field, e.g. clinical specialists, would choose not to use this particular configuration of the SCS, which was mainly designed for general practitioners, it is quite possible to design other, but similar, systems which allow for the peculiarities of the field in which they are to be tested.
In addition, for a system with almost a hundred thousand nodes, the accessibility and speed of use were excellent. Any node in the system can be

reached with less than ten or twelve clicks of the mouse button, and the construction of the tree is so logical, being designed with the doctor-user and his way of thinking in mind, that doctors can learn to traverse it correctly and pick out the items they want within a few hours of teaching, as has been convincingly shown.

The lines and notepads that were reserved for otherwise unencodeable particulars of patients were not used. Either the SCS might be so complete that no addition was found necessary, or the typing barrier was so high that the participant refrained from using this possibility of the system.

The SCS is the nucleus and the success of the DDSS. In the combination of a hierarchical system and a standardized vocabulary it is unlike previous attempts at standardization of terminology and therefore cannot be compared with any other system. Because of its internal structure and its detailed standardization of terminology the SCS is easily translatable into any language used in countries where Western medicine is practised.

8.5 The reasoning process

Although Elstein et al. [Els78] believe the hypothetico-deductive method of reasoning to be "a nearly universal characteristic of human thinking", he and his co-workers have only observed one part of the method which consists of a combination of two distinctive sections: an introductory inductive section and a judgemental deductive continuation. This second part is seldom performed in practice and was observed neither in a previous study [Rid89] nor in the present one. Sadegh-Zadeh believes deductive reasoning to be contradictory to the daily task of a practising physician [Sad74]. The matter is important in so far that the DDSS tries to compensate and redress the imperfections that adhere to the inductive way of reasoning. The matter is also important in so far as it explains some of the findings in the experiment. For the benefit of the reader we list here a number of the elements which can be recognized in inductive inference.

- early in the problem-solving process a diagnostic idea ("hypothesis"[1]) jumps to mind;
- this idea is based on only a few symptoms;

[1] The term "hypothesis" should not be taken in a strict scientific sense here.

- the idea stems from the doctor's own experience;
- information acquisition is coloured by the early hypothesis generation (tunnelling one's vision);
- the doctor's purpose is to verify a plausible idea;
- verification, in contrast with falsification, restricts the information acquisition to only a small number of positive (=matching) findings; [Bri87]
- the verification strategy distracts from probing deeper for more details;
- the verification strategy distracts from probing wider around the problem (lateral thinking);
- the verification strategy leads to neglect of alternative explanations of the observed evidence and bars reconsideration of one's judgement. Indeed, discovering some similarity between one's idea and the perceived evidence will inevitably lead to the conviction of being right [Pol58] [Osk65] [Ein78] [Vre93].

The redressing capacities of DDSS concentrate on three elements.
a) to probe deeper for more and more detailed information;
b) to consider possible solutions and ideas in various directions;
c) to reconsider one's judgement by contemplating alternative explanations for the present illness of the patient.

In essence, these elements reflect the learning methods advocated by de Bono [Bon70] and Brandsford & Stein [Bra84]. The idea of "lateral thinking" (thinking in a more panoramic way) has been put best by de Bono: The difference between "funnelled" and "lateral" thinking is that between digging a hole deeper and deeper, and digging a number of holes at different sites.

The 'hypotheses' entertained by our subjects were of a highly personal nature, reflecting personal experience as much as, if not more than, book learning. This is in agreement with observations of others: ideas jump to mind on the basis of just a few symptoms [Bar72] [Els78] [Rid89]. It is the basic process of pattern recognition: a superficial resemblance suffices to recognize similarity; it is like recognizing a well-acquainted person in a crowd [McC86]. Pattern recognition is a process of which doctors are proud; it is his 'flair clinique' which enables him to deliver immediate judgements. Pattern recognition plays a major role in inductive reasoning and prevents retracing of argumentation steps because of the implicit nature of the 'pattern' or 'hypothesis': it is formed without its owner being conscious of any steps leading up to it.

As a consequence of the personal character of the 'hypothesis' little overlap between hypotheses for the same patient by different doctors was found. Besides, hypotheses were often on a rather general level which makes the subsequent arrival at a specific diagnosis unlikely.

The verifying character of inductive reasoning carries another consequence: the so-called hypothesis-driven-data-acquisition. Once an idea has come to mind the doctor tries to find and select evidence in favour of this idea. His questions will be directed by the hypothesis he entertains at that moment. Only when patients' answers are evidently contradictory to the idea the doctor will switch to another idea and the process of hypothesis-driven-data-acquisition starts again.

By asking the participants to indicate the plausibility of the hypothesis (also called the probability statement) which he has just formulated, we obtained some insight in the mental state of the subject. His estimate of subjective, or predictive, probability usually reached far beyond the evidence the hypothesis was based upon. Any comparison with objective (=frequentistic) probabilities such as prevalence or incidence rates is therefore pointless. Subjective, or inductive, probabilities indicate the degree to which the doctor is prepared to gamble on a certain idea which he believes to be the winning one. This high subjective probability lingers on during the diagnostic process, so that at the end the defeated hypotheses are still held dear. In a verification process rejection of hypotheses is indeed not one of the most prominent characteristics.

The verification process leads also to (often unacceptable) selection of patient's answers or manipulation of data, an example of which is the entering of data which were not obtained from the patient but which served as valuable evidence in the light of the entertained hypothesis. These phenomena of fabrication and unwarranted interpretation of symptoms were found both in the simultaneous (guided by the SCS system) and the non-simultaneous (free interviewing and entering data afterwards) consultation style. Obviously, this effect is an innate characteristic of the verifying nature of inductive reasoning. One cannot expect doctors to discard such an implicit characteristic of their usual thought processes within the few hours of the experiment.

We do not intend to change an obviously usual behaviour of clinical problem-solving and decision-making. The DDSS is meant to make clear and explicit what is commonly hidden in the diagnostic process.

8.6 Diagnosis

The main purpose of diagnosis is to identify a disease in order to predict the future course of the illness in the patient. The predictive capability of a diagnosis is based on several premises:

1) a unified concept of the particular disease.
2) diseases are classified in a generally accepted taxonomy (nosology).
3) diseases present themselves in a rather typical phenetic appearance.
4) diseases can be distinguished by the observation of one or two different features (monothetic) or by the combined differences in their complete phenetic appearance (polythetic) [Sne73].
5) diseases follow a more or less pre-determined course.

This conception, however, is not generally accepted. Commonly two trends may be distinguished: an empirical and a conceptual one. These trends stand for, as Boinet said, "les deux grandes idées doctrinales qui reviennent sans cesse à travers les siècles après de long detours et avec des fortunes diverses". The empirical, and oldest, opinion starts from the thought that all suffering is unique and each patient shows his or her particular arrangement of symptoms and signs, and only when cases show considerable similarity a particular disease is suspected. This concept follows the adage: Post hoc ergo propter hoc. In this context terminology is just a way of expressing similarity between cases; it acts as a verbal shorthand for the doctor. For the empirical doctrine the words of Berkeley [Blo76] apply: "In truth there is no such thing as one precise and definite signification annexed to any general name, they all signifying indifferently a great number of particular ideas."

The other doctrine, the conceptual one, stands for ordering and systematization of objects (diseases) into a hierarchical and discernible classification. Such a classification was first performed by Sauvages de Lacroix in 1731. His work found followers, regrettably mostly in differing directions such as etiologic, anatomic, and biochemical classifications. The ideal of one unified concept of disease appears to be illusory so far.
The classes, or taxons, of a classification may contain the knowledge taken from numerous patient cases which represent the average picture and course

of the buzzing, blooming confusion that embodies the real world of illness and discomfort. It is the doctor's task to make a connection between the presentation of symptoms in the patient and the pre-defined picture as described in the taxonomy. Once the connection is, implicitly, established it is supported by the conviction of generality to the doctor. Generality exists not in its nature but in its use. An idea which is considered in itself as particular becomes general by being made to represent or stand for all other particular ideas of the same sort [Blo78].

In daily practice these two conceptions merge into one single conception representing the doctor's wisdom gathered by experience and his knowledge of "official medicine". But such a coupling creates confusion and contributes to the irreproducibility of diagnostics, especially when just diagnostic names are coded and recorded.

Diagnosis only becomes meaningful if its meaning (= underlying composition of phenomena observable in the patient; facts on which a doctor must decide) is clear and accessible to examination. King [Kin67] views diagnoses as congeries of factors selected to form a class or type which acts to organize experience. The selection of those factors which compose the disease entity is essentially arbitrarily except in so far as the disease description as a whole must convey some utility in the handling of experience. It made Dukes sigh: "Plenty of people are still dying of diseases which other people do not believe in." [Duk87]

The result of these notions is far from encouraging.
Taking the official medical nomenclature as used in medical diagnosis and judgement statistics show a gloomy picture. De Dombal [Dom92] found an overall accuracy of 41-53% in the UK and mentions an estimated accuracy of 60-65% globally. Zarling [Zar83] reported an accuracy of 44% for myocardial infarction, a figure which is in the same range as in the study of van der Does & Lubsen [Doe78]. In an earlier study with the same patients as were used in the present test Ridderikhoff [Rid86] found a diagnostic accuracy of 46% and in the present test the accuracy was found to be 43%, independent of the number of years of experience of the diagnostician.

We wanted to follow the basic tenet of our theory for the DDSS, i.e. to trace the doctor's behaviour as closely as possible. To this purpose, we tried to combine the two concepts, the empirical and the conceptual, into one system. The ultimate goal was:

- to allow the doctor to follow his habitual routine, and
- to confront him with the evidence upon which he bases his diagnosis.

With King we regard diagnoses as combinations of symptoms and signs. Matching patients' symptoms with the symptoms of diseases found in the knowledge base may produce a complete match or a list of alternative explanations. Each name is accompanied by the number of matches so that the doctor may check the accuracy of the match. DDSS does not attempt to make any judgement about the results. It is entirely up to the doctor to decide and to take responsibility for his decision.

From the observations mentioned before it is hardly amazing to find that doctors are fully convinced of their diagnoses, even in cases when substantial evidence was scarce. It is part of the empirical conception of medicine as well as inductive reasoning. What is really worrying is the observation that participants hardly considered the presented alternatives as possible incentives for reconsideration and revision.
Although all subjects were very anxious to have a look at the "diagnoses screen", they did not seem to bother about subsequent actions. This might be disappointing in the experiment, but it certainly challenges us to proceed with the DDSS and its implementation in actual practice. We are convinced that it will help in improving health care.

A few words about the knowledge base. The knowledge base was composed of 46 diseases defined with their configuration of symptoms and signs as obtained from 10 - 12 different textbooks and various other sources. A mixture of the often widely varying descriptions of diseases was taken as an average picture. These pictures and their diagnostic names were entered in the knowledge base by means of SCS. Among these 46 diseases were the actual diagnoses of the presented patients and for every patient case 5-6 diseases with similar symptom-patterns. The diagnostic performance must be viewed in this light. On the one hand the limited number of options, on the other the possibility that textbook descriptions hardly match the presentation of illnesses seen in primary health care. A real assessment performance can only be done if a proper knowledge base that complies with the doctor's special domain is constructed. For the experiment of testing the functioning of the system, a restricted knowledge base sufficed, but this precludes meaningful discussion about the diagnostic accuracy of the system.

8.7 System function

The system was primarily designed for two purposes:
a) keeping an administrative record of the 'raw' patient data, i.e. his or her complaints, symptoms, signs, etc, stored in a structured uniform way, and
b) being able to use these structured data to aid the doctor in arriving at a diagnosis.

Once the primary goal of symptom data storage in a highly structured and uniform format is attained, the problem of determining a diagnosis from the patient data may be explored in ways not previously considered. In order to obtain such a diagnostic goal the system should provide
1) an extensive medical knowledge base that complies with the doctor's knowledge domain, and
2) algorithms that make sure that when determining a diagnosis, one or more comparable diagnoses are selected from the knowledge base in which similar symptoms appear.

These requirements led to the construction of the DDSS shell.
By sticking to a method of systematic, structured design and programming, we were able to handle complex C and Clipper related issues.
The use of the two languages together stressed the need for a strictly modular construction of the system. Modules written in one language should communicate seamlessly with modules written in another language. The communication between modules takes place by intermediate data files and by returning different DOS error levels. The separate modules are incorporated into a single working programme system by means of a DOS batch file. This has the advantage that the interface between different modules is easily completely defined, and substitution of one module by an improved version is very easy and guaranteed not to interfere with the working of the rest of the program. The disadvantages are a slight loss in speed and a greater complexity of the final working system, which consists of a number of files, all of which are necessary for the complete system to function.

The Symptom Coding System can be developed completely separately from the DDSS programme code. The Symptom Tree is created with a separate, specially developed compiler from ordinary textfiles, and is then called by C modules. New developments in medicine, new ideas in clinical practice, opening of new domains and specifications for different medical disciplines can be smoothly incorporated without any necessary adaptation of the code for the DDSS programme itself.

Apart from two occasions when a non-fatal, recoverable software error occurred, the system ran smoothly during approximately 100 hours of use, in 99 patient cases and with doctors who treated the device sometimes rather unkindly. For a person with some computer experience, the practical use of the system could be learned within minutes. Finding the correct keys and using the mouse came so naturally that hardly anyone felt the need to use the manual, or the on-line help functions. This is as it should be. The main problem encountered was that one or two subjects found it hard to keep the fingers on the three buttons of the mouse, and tended instead to press each mouse button with their index finger.
Users felt in control of the system and freely explored the SCS tree.

Time lags hardly existed and even in extreme cases remained within annoyance limits.
The template matching algorithm's performance depends on the amount and quality of the data input. Given only a small number of symptoms the system comes up with a large list of differential diagnoses. This possibility might be interesting in the beginning of the consultation when the doctor is still walking in the dark. At any moment during the consultation the doctor may consult the knowledge base for further exploration, especially to compare the acquired data with the symptoms and signs of particular diseases as a quick reference.
The menu-system appeared to be extremely user-friendly and efficient. Utmost attention has been paid to keeping the screen uncluttered, displaying only the information which the user needs at that point.
The system is meant to operate in a stand-alone situation. This option was chosen because of the privacy of the information that is exchanged in the consulting room. However, to profit from all the accumulated empirical knowledge stored in the various records of doctors, exchange of information is worth pursuing. If anonymity can be guaranteed DDSS will open up roads to construct knowledge bases for specific domains as well as scientific research.

The system functioned as it was meant to do. It stood all tests and (mis)uses that were likely to occur in daily practice. We are confident that the system will run smoothly and effortlessly to the satisfaction of doctor and patient.

8.8 Scaling up the system

An interesting question to ponder is whether, and if so, how, the situation will change when the system is scaled up to include more diseases in its knowledge base and all the patients in a practice in its electronic record system.
With regard to storage requirements there is little to fear; modern storage media provide ample space to store the consultation data, which in our study required only a few hundred bytes per consultation on average. A considerable economy of storage would also be possible by compressing the data with a standard data compression algorithm like that of Ziv-Lempel, or by leaving out the (redundant) human-readable form of the data, which is currently included in every record but which if needed can also be looked up in the SCS tree with a slight time penalty. The speed of the template-matching algorithm in the current implementation was sufficient when the knowledge base contained only about fifty diseases; but when this number rises there could be a detoriation in response time. Analysis of the programme execution time (profiling) showed that it is in its current form I/O bound[1], and that a slight modification of the disk access routines would speed up the execution by a factor of four. Further tuning should be able to at least double the speed again; finally, the use of faster processors than the currently used Intel 80386 at 25 MHz will enable us to double or triple the speed again, for a total gain of up to 25 times the current speed. If we regard a response time of 5 seconds for the differential diagnosis as acceptable, the size of the knowledge base which can be handled with such an improved system becomes about 1000 to 2000 disease patterns, and even more with the increasing computer power which will become available in the next few years.
A more difficult problem is which disease patterns to include in the knowledge base. There exist far more diseases than are ever seen by any single doctor; a catalogue of inherited and genetic abnormalities alone lists thousands of diseases, most of which are so rare that most doctors never see one in their entire lifetime as a practitioner of medicine. Some criterium will have to be adhered to. For example, only those diseases which a doctor in the field where the system is to be used can expect to see (statistically) at least once every 5 or 10 years should be included, plus a number of rarer but 'dangerous' diseases which it would be very undesirable to miss.

---

[1] An expression meaning that the execution time is determined by the speed with which the disk can be accessed.

For the general practice, we estimate that 1000 to 2000 diseases would be an ample selection.
There is yet another problem: how will the differential diagnosis look when such a large number of patterns is available in the knowledge base? Clearly, a list of tens or hundreds of possibilities, each with just two or three matching nodes, is not much help when making a diagnosis. It will probably be necessary to establish a cut-off point, beyond which a partial template match will not be shown. This may be done in several ways, with a separate threshold being defined for each pattern, or by looking at the relative quality of the match with regard to other diagnoses. A match with a score of 3 need not be shown if there is also a pattern which matches 10 items; but it may have to be shown if it is the best or only match. It is also possible to set the threshold according to the number of items entered for the patient, and to show a warning that not enough data are available for meaningful support.
What will happen when the data for a patient are entered who is not ill in a medical sense? We must remind ourselves that in general practice, the majority of people who come to consult their doctor do not really suffer from any recognized disease; they are worried, they have complaints, but they cannot be helped by strictly medical means. Is there a danger that such patients will be assigned a diagnostic label, just because the computer comes up with a template match? As we have repeatedly stated before, it is the doctor who is in charge and in control of the system. The responsibility for the diagnosis remains squarely his. Although we strongly recommend that the data of every patient be entered into the system, the system does not make a diagnosis; its role is to give the doctor suggestions to ponder and to accept or reject as he sees fit. Since it has been frequently remarked that 'hysteria has a remarkably high mortality for a neurotic disorder', perhaps even some of the 'non-diseased' patients will benefit from a correct somatic diagnosis.

8.9 Conclusions and recommendations

DDSS has given what we expected of it. In the hands of experienced physicians it may add a new dimension to their skills. But it will also challenge their competence and actions. For the doctor, the responsibility for his decisions remains squarely his own, even if the system's suggestions may not always coincide with his own opinions and judgement. It is entirely up to him how he will respond and behave. He has the option to rethink and retrace his steps.

For every hypothesis he might possibly entertain he can check the acquired information and examine the composition of symptoms and signs as they are represented in the medical knowledge base.

In our experiment the retracing process proved of great value. The successive steps in the form of generated hypotheses could easily be traced by the user and the experimenter. Any (standard) datum was registered in the order in which it was entered. This procedure provided insight in the variability of the acquired data, of the hypotheses, of the subjective probabilities assigned to these, and ultimately, to the diagnoses and the evidence on which the judgement was based.

All these recommendations and conditions have led to what we consider to be our top priority: the immediate use of the system by the doctor himself during the patient contact. Our slogan is: discover and explore the possibilities of DDSS whilst seeing the patient. It may take the practising doctor into the world of advanced informatics. If we indeed consider a computerized medical diagnostic system a valuable asset in improving health care, DDSS gives the individual doctor the opportunity by critiqueing his judgement, his management and his actions. Many doctors feel very uneasy if somebody is "watching over their shoulder". Most doctors are unaccustomed to being criticized. With DDSS every doctor has the possibility to verify (or falsify) his ideas and his plans, unobserved by others, in his own practice. And moreover, within the time limits given for normal consultations.

Doctors are expected to profit from this experience. By it very nature, experience is subjective and therefore not transferable. But to learn only from one's own mistakes would be slow and unnecessarily costly to one's patients [McI83], apart from the notion that such learning does not contribute to accessible medical knowledge. DDSS may contribute to more explicit and testable learning by encouraging an attitude of constant and watchful examination of one's own steps and comparing this with the steps one's colleagues might have taken.

We are encouraged to keep looking for even simpler and more user-friendly methods of input. Perhaps voice recognition will provide an answer here; we can easily imagine how it would be possible to adapt the system to voice-driven menu selection. For the time being, the field of voice recognition is not sufficiently advanced yet to allow its use in such a programme as ours,

but with the current speed of developments this stage may be just a few years away.

Also with regard to user-friendliness, many nodes in the SCS tree are currently taken up by menus to indicate the position of a specific complaint or phenomenon in or on the body. Such information is much more easily codeable by pointing to the location on a map of the body, rather than having to go through three or four layers of menus for every position item. This is an improvement which would be quite feasible to implement with current technology.

DDSS must be viewed as a first and most successful attempt to structure and organize the diagnostic procedure of practising physicians. It is still in its infancy. New developments in computer software may improve further acceptability and user-friendliness so that many practising doctors may want to use this program. This is exactly what we want: many doctors with their great experience and the standardized data from their real patients contributing to the construction of that particular medical knowledge base that complies with their domain. It will benefit medicine as a science, the doctors, and, last but not least, their patients.

Summary

In chapter 1, it is argued that at present the standardization of disease definitions and their classification into a nosology is chaotic at best.
Different views on disease, diagnosis and the current state of nosology are discussed. It is argued that the prime purpose of diagnosis is the identification of the patient's disease on the basis of an existing classification, thus allowing an accurate prediction of the outcome and the identification of methods to influence this outcome favourably. In our view, similarity of patients should be defined as much as possible in terms of easily distinguishable characteristics, using the clinical presentation of patients rather than (often unknown and ill-defined) presumed underlying causes. We try to classify the patient on the basis of similarities and dissimilarities between his or her signs and symptoms and those of a pre-defined pattern called a disease.
In chapter 2, we take a look at the introduction of computers into medicine and examine the obstacles to success in applying computer technology to medical situations. It is argued that important obstacles to progress lie in the lack of standardization of medical terminology, incomplete understanding of the processes involved in human medical decision making, use of incorrect models to simulate these processes, lack of organized knowledge about medical conditions (partly because of the lack of standardization mentioned earlier), lack of structure of medical data records in currently used systems, and emotional barriers resisting the introduction of computers into the essentially human encounter between doctor and patient.
In chapter 3, several approaches to the diagnostic process are discussed, notably prescriptive approaches such as employed in medical decision making, and the descriptive approaches explored in cognitive science. Some philosophical and probabilistic models are reviewed. A model is introduced, based on earlier research, according to which under normal, daily circumstances a doctor's diagnostic process is characterized by 1) early generation of a diagnostic hypothesis, and 2) an inductive, verification-seeking mode of reasoning, rather than the deductive or hypothetico-deductive and falsifying mode of reasoning often postulated in literature. It is argued that this leads to a number of characteristic weak points in medical diagnosis, notably oversight of other likely hypotheses and incomplete verification of diagnostic hypotheses, while the occurrence of hypothesis-driven data collection leads to wrong interpretations or unconscious fabrication of evidence to support the hypothesis currently held.
In chapter 4, we explore the requirements to be met by the DDSS in the light of the model developed in the previous chapter. The goal of the DDSS

is to improve the diagnostic process at many points: better observation, both in depth and in scope of observation, and better recording of the observed facts, leading to better hypothesis generation and evaluation, and ultimately to a better knowledge base from which to derive diagnoses. We have also specified a large number of constraints for the system, notably that it should be compatible with existing standards for medical database design, that the user-interface should be extremely easy to handle and that the system should be used *during* the consultation. Finally, the doctor should at all times be in control of the consultation, with the system following his thought rather than prescribing it.

In chapter 5, the design of the DDSS is explained and the mathematical methods and algorithms used in the programme are illustrated by examples.

The DDSS is based on a systematic coding of the patient's symptoms by means of a menu-driven system called the SCS. The support the DDSS offers to the practitioner of medicine consists of two types of feedback. During the data-gathering phase, by showing the structured symptom menus. After the doctor has made a diagnosis, by offering on demand a list of differential diagnoses produced by template matching with patterns stored in a pre-defined database. This database should be constructed specifically for every user group to contain the typical patterns seen by doctors belonging to such a group. The time and space requirements of a system using such data structures and algorithms are analysed.

Chapter 6 contains a description of the experiment to be performed with the finished system, consisting of 100 consultations of 20 subject doctors, using 8 different simulated patients presenting fairly common somatic complaints. A number of hypotheses, introduced in chapter 5, is further worked out to be tested by the experiment to follow. The ease with which subjects can learn to use the system, their experiences with it, the time needed for consultations using the system, the quality of the diagnostic support offered and the doctor-subjects' reactions to it are scrutinized with special care.

In Chapter 7, the results of the experiment are presented and discussed.

The main results are that the system is very easy to learn, that it can indeed be used during the consultation, that the interface is simple to learn and to use, that the time needed to use the system during the consultation is not prohibitive, that an extensive number of patient data end up in the medical patient database, and that the speed with which the diagnostic support is offered is sufficient. However, doctors do not appear to make much use of the diagnostic suggestions even in cases in which these suggestions are better than their own diagnoses.

In Chapter 8, the results of the experiment are discussed in the light of previous publications dealing with expert and decision support systems, and decision-making by doctors. It is argued that a system like the DDSS, which has been designed and built with a constant eye to the needs of the working general practitioner, is better suited to clinical use than many other systems which have been developed in the laboratory and were later transported to a clinical setting.

Samenvatting

In hoofdstuk 1 wordt beargumenteerd dat de standaardisatie van ziektedefinities en hun ordening in een nosologisch systeem tegenwoordig nog op zijn zachtst gezegd in een chaotische toestand verkeert. Verschillende benaderingen van ziekte, diagnose, en de huidige stand van de nosologie worden besproken. De stelling wordt verdedigd dat het voornaamste doel van diagnose de identificatie van de ziekte van de patient in een bestaand classificatiesysteem is, teneinde daarmee een nauwkeurige voorspelling te kunnen doen van de afloop van het ziekteproces en methoden te kunnen identificeren om dit verloop gunstig te kunnen beinvloeden. De overeenkomsten tussen patienten zouden volgens ons zoveel mogelijk moeten worden gedefinieerd in termen van gemakkelijk vast te stellen kenmerken, waarbij meer nadruk wordt gelegd op de klinische presentatie dan op de (vaak onbekende of slecht gedefinieerde) veronderstelde onderliggende oorzaken. We proberen de patient te classificeren op basis van de overeenkomsten en verschillen tussen zijn of haar symptomen en verschijnselen, en die van een pre-gedefinieerd patroon dat 'ziekte' genoemd wordt.
In hoofdstuk 2 wordt de introductie van computers in de geneeskunde besproken en de belemmeringen die een succes bij de toepassing van computers in de geneeskunde in de weg staan onderzocht.
Er wordt bepleit dat belangrijke obstakels voor vooruitgang liggen in de gebrekkige standaardisatie van medische terminologie, onvolledig begrip van de processen die plaatsvinden bij menselijke besluitvorming, gebrek aan georganiseerde kennis van medische aandoeningen (gedeeltelijk ook door het eerder genoemde gebrek aan standaardisatie), gebrek aan structuur in de medische status zoals gebruikt in huidige computersystemen, en emotionele barrieres tegen de introductie van de computer in het essentieel-menselijke contact tussen arts en patient.
In hoofdstuk 3 worden verschillende benaderingen van het diagnostisch proces besproken, vooral prescriptieve benaderingen zoals die worden gebruikt in de medische besliskunde, en de descriptieve modellen die door de cognitieve wetenschappen worden verkend. Verder worden enige filosofische benaderingen besproken. Gebaseerd op eerdere research wordt er een model geïntroduceerd waarvan de voornaamste punten zijn dat het diagnostisch proces van de arts wordt gekenmerkt door 1) vroege generatie van een diagnostische hypothese en 2) een inductieve, op verificatie gerichte redeneertrant, in plaats van de in de literatuur vaak gepostuleerde deductieve of hypothetico-deductieve redeneerwijze. Er wordt bepleit dat dit aanleiding geeft tot een aantal typische zwakke punten in de medische diagnostiek, met name het over het hoofd zien van andere mogelijk hypothesen en incomplete

verificatie van diagnostische hypothesen, terwijl het gebruik van hypothese-gestuurde gegevensverzameling tot verkeerde interpretatie of zelfs onbewuste vervalsing van gegevens van de patient kan leiden om de op dat moment gehanteerde hypothese te ondersteunen.
In hoofdstuk 4 verkenen we de eisen die moeten worden gesteld aan het DDSS in het licht van het in het vorige hoofstuk ontwikkelde model. Het doel van het DDSS is om het diagnostisch proces op veel punten te verbeteren: beter observeren, zowel in de breedte als in de details van de waarneming, en betere verslaglegging van het waargenomene, wat moet leiden tot een betere hypothesegeneratie en -evaluatie, en uiteindelijk tot een betere kennisbank waaruit de diagnosen kunnen worden afgeleid. Tevens hebben we een aantal randvoorwaarden voor het systeem opgesteld, met name dat het systeem verenigbaar moet zijn met bestaande normen voor medische databases, dat de gebruikers-interface bijzonder eenvoudig moet zijn, dat het systeem gebruikt moet worden *tijdens* het consult, en dat de arts op op ieder moment het consult leidt, waarbij het systeem hem volgt in plaats van zijn handelen te bepalen.
In hoofdstuk 5 wordt het ontwerp van het DDSS uitgelegd en de wiskundige methoden en algoritmen die in het programma worden gebruikt met voorbeelden geillustreerd. Het DDSS is gebaseerd op een systematische codering van van de symptomen van de patient door middel van een menugestuurd systeem dat SCS heet. De ondersteuning die het DDSS de uitoefenaar der geneeskunde biedt bestaat uit het geven van terugkoppeling zowel gedurende de fase van het informatieverzamelen, door het tonen van symptoom-menu's, en nadat de arts een diagnose heeft gesteld, door op verzoek een lijst met mogelijke diagnosen te tonen die door vergelijking van de symptomen van de patient met patronen uit een vooraf samengestelde kennisbank is opgesteld. Deze kennisbank moet eigenlijk voor ieder gebied waarop het systeem wordt ingezet speciaal op maat gemaakt worden om de typische patronen te omvatten die door artsen op dat gebied worden gezien. De rekentijd- en geheugeneisen die een systeem met deze gegevensstructuren nodig heeft worden geanalyseerd.
Hoofdstuk 6 bevat een beschrijving van het experiment dat met het zo ontstane systeem wordt uitgevoerd, bestaande uit 100 consulten, door 20 arts-vrijwilligers uitgevoerd op 8 verschillende gesimuleerde patienten die geenszins ongewone somatische klachten presenteren. Een aantal hypothesen, geïntroduceerd in hoofdstuk 5, wordt verder uitgewerkt om te worden getoetst in het uit te voeren experiment. Het gemak waarmee proefpersonen leren het systeem te bedienen en hun ervaringen ermee, de tijd die ze voor consulten met het systeem nodig hebben, de kwaliteit van de verleende

diagnostische ondersteuning en de reactie van de arts-proefpersonen daarop worden met extra zorg bekeken.
In hoofdstuk 7 worden de resultaten van het experiment gepresenteerd en besproken. De belangrijkste resultaten zijn dat de bediening van het systeem zeer eenvoudig te leren is, dat het inderdaad goed tijdens het consult kan worden gebruikt, dat de extra tijd die nodig is voor een consult met het systeem het gebruik niet onmogelijk maakt, dat een groot aantal gegevens van de patient ook inderdaad in het electronisch dossier terechtkomt, en dat de snelheid waarmee de diagnostische ondersteuning wordt geboden acceptabel is. De artsen leken echter weinig gebruik te maken van de diagnostische suggesties van het systeem, zelfs in gevallen waar die suggesties beter waren dan hun eigen diagnosen.
In hoofdstuk 8 worden de resultaten van het experiment besproken in het licht van eerdere publikaties over expert- en beslissingsondersteunende systemen, en beslissingsprocessen bij artsen. Er wordt betoogd dat een systeem als het DDSS, dat is ontworpen en gebouwd met voortdurende aandacht voor de behoeften van de praktizerende huisarts, meer geschikt is voor klinisch gebruik dan veel andere systemen die zijn ontworpen in het laboratorium en naderhand pas overgezet naar de klinische situatie.

References

Aho78 Aho AV, Kernighan BW, Weinberger PJ. The AWK programming language. Addison-Wesley, Reading, Mass. 1988.

And68 Anderson J. The computer: medical vocabulary and information. br med bull 24 (1968) 194-8.

And68a Anderson JA, Boyle JA. Computer diagnosis: statistical aspects. br med bull 24 (1968) 230-35.

Ano79 Anon. The concept of disease. (editorial) BMJ 1979, 751-2.

Ash72 Asher R. Talking sense. Pitman medical, Tunbridge Wells 1972.

Bal80 Balla JI. Logical thinking and the diagnostic process. Meth Inf Med 19 (1980) 88-92.

Bar72 Barrows HS, Bennett K. The diagnostic skill of the neurologist. Arch Neurol 26 (1972) 273-277.

Bar87 Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain, an evolving diagnostic decision-support system. JAMA 258 (1987) 67-74.

Biö77 Biörck G. The essence of the clinician's art. Acta Med Scand 1977;201:145-147

Blo76 Bloor M. Bishop Berkeley and the adenotonsillectomy enigma: an exploration of variation in the social construction of medical disposals. Sociology 10 (1976) 44-61.

Blo78 Bloor M. On the routinised nature of work in people processing agencies.in: Relationschips between doctors and patients (A. Davis, ed.) Saxon House, London, 1978.

Blo80 Blois MS. Clinical judgement and computers. NEJM 303,4 (1980) 192-97

Blo81 Blois MS, Tuttle MS, Sherertz DD. Reconsider: a program for generating differential diagnoses. proc IEEE (1981), 263-68.

Blo83 Blois MS. Conceptual issues in computer-aided diagnosis and the hierarchical nature of medical knowledge. J med phil 8, (1983) 29-50.

Boh75 Bohlinger RE, Ahlers P. The science of "pattern recognition". JAMA 233 (1975) 1289-90.

Bon70 Bono de E. Lateral Thinking. New York, Harper & Row, 1970

Boo77 Boorse C. Health as a theoretical concept. Phil sci 44 (1977) 542-73.

Boy68 Boyle JA, Anderson JA. Computer diagnosis: clinical aspects. br med bull 24 (1968) 224-9.

Bra68 Braithwaite RB. Scientific explanation. A study of the function of theory, probability and law in science. Cambridge University Press, 1968

Bra76 Bradshaw-Smith JH. A computer record-keeping system for general practice. BMJ (1976) 1395-7

Bra84 Brandsford JD, Stein BS. The ideal problem solver. New York, WH Freeman & Co, 1984

Bre80 Brehmer B. In one word: not from experience. Acta Psychologica 45 (1980) 223-241.

Bri87 Briskman L. Doctors and witchdoctors: Which doctors are which? I. BMJ 295 (1987) 1033-6.

Bri87a Briskman L. Doctors and witchdoctors: Which doctors are which? II. BMJ 295 (1987) 1108-10.

Bro59 Brodman K, Woerkom AJ van, Erdmann AJ, Goldstein LS. Interpretation of symptoms with a data-processing machine. arch intern med 103 (1959) 776-82.

Cam79 Campbell EJM, Scadding JG, Roberts RS. The concept of disease. BMJ 2 (1979) 757-762.

Car60 Carnap R. Statistical and inductive probability. in: Madden EH.(ed.): The structure of scientific thought and introduction to philosophy of science. London, Routledge & Kegan Paul (1960) 269-279

Cha69 Chapman LJ, Chapman JP. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. J. Abnormal Psychology 74, (1969) 271-80.

Cle47 Clendening L, Hashinger EH. Methods of diagnosis. St Louis, Mo. Mosby 1947.

Cro72 Croft DJ. Is computerized diagnosis possible? Comp biomed res 5 (1972) 351-67.

Cro74 Croft DJ, Machol RE. Mathematical methods in medical diagnosis. Ann.Biomed.Eng., 1974, 2, 69-89

Cum76 Cumberbatch J, Heaps HS. A disease conscious method for sequential diagnosis by use of disease probabilities without assumption of symptom independence. Int J Biomed Comp 7 (1976) 61-78.

Day85 Day WHE. Optimal algorithms for comparing trees with labeled leaves. J classification 2 (1985) 7-28.

Dif84 Difford F. Reducing prescribing costs through computer controlled repeat prescribing. J roy coll gen pract 34 (1984) 658-60.

Doe78 Does vd E, Lubsen J. Imminent Myocardial Infarction Rotterdam. Dissertation, Rotterdam, 1978.

Dom72 Dombal FT de, Horrocks JC, Staniland JR, Guillou PJ. Pattern-recognition: A comparison of the performance of clinicians and non-clinicians - with a note on the performance of a computer-based system. meth inform med 11 (1972) 32-7.

Dom72a Dombal FT de, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. BMJ 1972,2, 9-13.

Dom78 Dombal FT de. Medical diagnosis from a clinician's point of view. Meth inform med 17 (1978) 28-35.

Dom78a Dombal de FT, Horrocks JC. Use of Receiver Operating Characteristics (ROC) curves to evaluate computer confidence threshold and clinical performance in the diagnosis of appendicitis. Meth.Inf.Med. 1978;17:157-161

Dom87 Dombal FT de. Computer-aided decision support- the obstacles to progress. meth inform med 26 (1987) 183-4.

Dom91 Dombal de FT, Dallos V, McAdam WAF. Can computer aided teaching packages improve clinical care in patients with acute abdominal pain? BMJ 302 (1991) 1495-7.

Dor80 Doroszewski J. Hypothetico-nomological aspects of medical diagnosis, part I: General Structure of the Diagnostic Process and its hypothesis-directed stage. Metamedicine 1 (1980) 177-94.

Edd82 Eddy DM, Clanton CH. The art of diagnosis. Solving the clinicopathological exercise. NEJM 306 (1982) 1263-8.

Ein78 Einhorn HJ, Hogarth RM. Confidence in Judgment: Persistence of the illusion of validity. Psychol. Rev. 1978;85:395-416.

Els78 Elstein AS, Shulman LS, Sprafka S. Medical Problem Solving. Cambridge (Mass), Harvard University Press 1978.

Eng63 Engle RL, Davis BJ. Medical diagnosis: present, past, and future. Arch int med 112 (1963) 512-19.

Eng76 Engelhardt HT. Ideology and etiology. j med phil 1 (1976) 257-68.

Eng79 Engelhardt HT, Spicker SF, Towers B Clinical judgment, a critical appraisal. D.Reidel Publ.Co., Dordrecht, 1979

Fei73 Feinstein AR. An analysis of diagnostic reasoning I. The domains and disorders of clinical macrobiology. Yale J biol med 46 (1973) 212-32.

Fei73a Feinstein AR. An analysis of diagnostic reasoning II. The strategy of intermediate decisions. Yale J biol med 46 (1973) 264-83.

Fei73b Feinstein AR. The problems of the "Problem-Oriented Medical Record". Ann intern med 78 (1973) 751-62.

Fei77 Feinstein AR. The haze of Bayes, the aerial palaces of decision analysis and the computerized Ouija Board. Clin Pharm Ther 21 (1977) 482-96.

Fri77 Friedman RB, Gustafson DH. Computers in clinical medicine, a critical review. Comp biomed res. 10 (1977) 199-204.

Gar60 Garland LH. The problem of observer error. Bull NY Acad Med 36 (1960) 570-584.

Gar78 Gardiner RC. Quality Considerations in medical records abstracting systems. J Med Systems 1978;2:31-43

Gel89 Gelsema ES. Pattern recognition and artificial intelligence in medical research and clinical practice. meth inform med 28 (1989) 63-5.

Gil73 Gill PW, Leaper DJ, Guillou PJ, Staniland JR, Horrocks JC, de Dombal FT. Observer variation in clinical diagnosis. A computer aided assessment ofits magnitude and importance in 552 patients with abdominal pain. Meth Inf Med 12 (1973) 108-113.

Gol68 Goldberg

Gri81 Griner PF, Mayewski RJ, Mushlin, AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. Ann Int Med 94 (1981) 553-600.

Gro77 Gross R. Was ist eine Krankheit? Gibt es stabile Krankheitsbilder? Metamed 1 (1977) 115-22.

Gro85 Groen GJ, Patel VL. Medical problem-solving: some questionable assumptions. Med Educ 19 (1985) 95-100.

Gru72 Gruer KT. Livingston new town- using a computer for general practice record. J roy coll gen practit. 22 (1972) 100-7.

Har72 Harre R. The philosophies of science. An introductory survey. Oxford University Press, London, Oxford, New York (1972).

Has92 Hasman A. Huisarts informatie systemen. Tijdschrift voor medische informatica 21,2 (1992) 59.

Höp90 Höppener P. Automatisering en wetenschappelijk onderzoek in de huisartsenpraktijk. thesis, Maastricht 1990.

How82 Howell WC, Fleishman EA. (eds.) Human performance and productivity. Information processing and decision making. Hillsdale, Erlbaum Associates (1982).

How82a Howell WC. An overview of models, methods, and problems. in: Howell WC, Fleishman EA. Human performance and productivity: Information processing and decision making. Hillsdale, Lawrence Erlbaum, 1982.

Jac92 Jacobs JA, Stobberingh EE. Naamsverandering van bacterien. I. Classificatie. NTVG 1992; 136, 1799-802.

Jac92a Jacobs JA, Stobberingh EE. Naamsverandering van bacterien. II. Nomenclatuur en identificatie. NTVG 1992; 136, 1803-7.

Joh77 Johnson-Laird PN, Wason PC. Thinking, readings in cognitive science. Cambridge University Press,1977.

Joh81 Johnson PE. Expertise and error in diagnostic reasoning. Cognitive science 5 (1981) 235-83.

Joh83 Johnson PE. What kind of expert should a system be? J med phil 8 (1983) 77-97.

Kah72 Kahnemann D, Tversky A. Subjective probability: a judgment of representativeness. Cogn Psych 3 (1972) 430-454.

Kan68 Kanal LN. (ed) Pattern recognition. Washington, DC: Thompson book comp 1968.

Kas78 Kassirer JP, Gorry GA. Clinical problem solving: a behavioral analysis. ann intern med 89 (1978) 245-55.

Kel76 Kellert SR. A sociostructural concept of health and ilness. j med phil 1 (1976), 222 etc.

Kin67 King, L.S.: What is a diagnosis? JAMA, 1967, 202, 154-157

Kle68 Kleinmuntz B. The Processing of clinical information by man and machine,in: Formal representation of human judgment. New York, John Wiley & Sons (1968) 149-86.

Koc83 Kochen M. How clinicians recall experiences. Meth inf med 22 (1983) 83-6.

Kom79 Komaroff AL. The Variability and inaccuracy of Medical Data. Proc IEEE 67 (1979) 1196-1207.

Krä80 Kräupl Taylor F. The concepts of disease. Psychological medicine 10 (1980) 419-24.

Lah78 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 17 (1978) 103-105.

Lam86 Lamberts H. (reason for encounter classification)

Led59 Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. Science 130 (1959) 9-21.

Lei91 Lei J van der, Musen MA, Does E van der, Man in't Veld AJ, van Bemmel JH. Comparison of computer-aided and human review of general practitioners' management of hypertension. Lancet 338 (1991) 1504-8.

Lei91a Lei J van der, Musen MA. A model for critiquing based on automated medical records. Comp and biomed research 24 (1991) 344-78.

Lic78 Lichtenstein S, Slovic P, Fischhoff B, Layman M, Combs B. Judged frequency of lethal events. J Exp Psych 4 (1978) 551-78.

Lun87 Lundsgaarde HP. Evaluating medical expert systems. Soc Sci Med. 24 (1987) 805-19.

Mar76 Margolis J. The concept of disease. J med phil 1 (1976) 238-55.

Mar91 Markus A. Is it permissible to edit medical records? BMJ 303 (1991) 349-51.

McC86 McCormick JS. Diagnosis: the need for demystification. Lancet, (1986), ii, 1434-5

McG80 McGaghie, W.C.: Medical problem solving : a reanalysis. J.Med.Educ., 1980, 55, 912-921

McI83 McIntyre N, Popper K. The critical attitude in medicine: the need for a new ethics. BMJ 287 (1983) 1919-23.

McM79 McMullin, E.: A clinician's quest for certainty. in: Engelhardt HT, Spicker SF, Towers B. Clinical judgment: a critical appraisal. Dordrecht, D. Reidel Publ. Co 1979.

McM83 Mcmullin E. Diagnosis by computer. J med. philos. 8 (1983), 5-27.

McM83a Mcmullin E. Introduction. J med. philos. 8 (1983), 3-4

Mec78 Mechanic D. Medical sociology. 2nd ed. New York, Free Press, (1978).

Med69 Medawar P. Induction and intuition in scientific thought. London, Methuen 1969.

Mer86 Merskey H. Variable meanings for the definition of disease. J Med Phil 11 (1986) 215-32.

Met80 Mettes CTCW, Pilot A. Over het leren oplossen van natuurwetenschappelijke problemen;een methode voor ontwikkeling en evaluatie van onderwijs. T.H.Twente, Thesis, 1980.

Met82 Metcalfe DHH, Reid DAT. General practice and technological promise. in: Malcolm A, Poyser J. (eds) Computers and the general practitioner. Oxford, Pergamon press 1982.

Met84 Metcalfe DH. The computer in general practice. J med eng technol 8, 2 (1984) 53-5.

Mil85 Miller Brown W. On defining 'disease'. J med phil 10 (1985) 311-28.

Mis79 Mischel W. On the interface of cognition and personality: beyond the person-situation debate. Amer Psychologist 34 (1979) 740-54.

Mos75 Moser RH. The first five minutes. (editorial) JAMA 231 (1975) 1169-70.

Mul84 Mullins P. BASIC or Pascal: which is the language for the generalist? J fam practice (1984) 393-5.

Mus89 Musen MA, van der Lei J. Knowledge engineering for clinical consultation programs: modeling the application area. meth inf med 29 (1989) 28-35.

New58 Newell A, Shaw JC, Simon HA. Elements of a theory of human problem solving. Psycholog. Review, 1958;65:151 - 166.

Nob77 Nobrega FT, Morrow GW, Smoldt RK, Offords KP. Quality assesment in the hypertension: analysis of process and outcome methods. NEJM 296 (1977) 145-148.

Nor88 Norman GR. Problem-solving skills, solving problems and problem-based learning. Med Educ 22 (1988) 279-86.

Nut81 Nutton, V. (ed): Galen: Problems & Prospects. London, The Wellcome Institute for the history of medicine, 1981

Osk65 Oskamp S. Over-confidence in case-study judgments. J Consult Psychol 1965;29:261-265.

Pat87 Patil, Ramesh S. Causal reasoning in computer programs for medical diagnosis. Comp methods and programs in biomedicine 25 (1987) 117-24.

Pau76 Pauker SG, Gorry GA, Kassirer JP, Schwartz WB. Towards the simulation of clinical cognition. Taking a present illness by computer. Am J Med 60 (1976) 981-996.

Pay88 Payer L. Medicine and culture. New York, Holt (1988).

Pol58 Polanyi M. Personal knowledge. Towards a post-critical philosophy. London, Routledge & Kegan Paul, 1958

Pop72 Popper KR. Objective Knowledge. Oxford, Clarendon (1972).

Pop77 Popper KR. On hypotheses. in: Thinking, readings in cognitive science. Cambridge University Press, 1977

Pop83 Popper KR. Realism and the aim of science. Ed. by W.W.Bartley III, Rowman & Littlefield, Totowa, New Jersey, (1983).

Pre70 Preece JF, Gillings DB, Lippmann EO, Pearson NG. An on-line record maintenance and retrieval system in general practice. Int j biomed comp 1 (1970) 329-37.

Pre84 Preece JF, Ashford JR, Hunt RG. Writing all prescriptions by computer. J roy coll gen pract 34 (1984) 655-7.

Qua86 Quaak MJ, Westerman RF, Schouten JA, Hasman A, van Bemmel JH. Computerization of patient history- patient answers compared with medical records. meth inform med 25 (1986) 222-28

Rei78 Reiser, SJ The decline of the clinical dialogue. J Med Philos 3 (1978) 305-313.

Ret88 Rethans JJ, Höppener P, Wolfs G, Diederiks J. Do personal computers make doctors less personal? BMJ 296 (1988) 1446-8.

Rid81 Ridderikhoff J. Functiebepaling huisarts. Medisch contact #2, (1981) 37-40.

Rid85 Ridderikhoff J. Models of decision-making in the general practice: a design for a descriptive research. Med Inform 10 (1985) 323-337.

Rid85a Ridderikhoff J. The paper patient - a device for investigating into general practice. in: Sheldon M, Brooke J, Rector A. (eds) Decision making in general practice. Macmillan, Houndmills 1985. 237-53.

Rid86 Ridderikhoff J. Decision-making strategies in the general practice. Thesis, Rotterdam 1986.

Rid86a Ridderikhoff J. Besluitvorming in de geneeskunde. In: Lebelle P, Muller Sj. (red) Besluitvorming, wat weten wij ervan? Kluwer, Deventer 1986 119-30.

Rid87 Ridderikhoff J. Obstacles in computer-aided diagnosis. In: Reichertz PL, Lindberg PAB. (eds) Lecture notes in medical informatics. Springer, Berlin 1987. pp 86-98.

Rid89 Ridderikhoff J. Methods in medicine. Dordrecht, Kluwer Acad. Pub., 1989

Rid91 Ridderikhoff J. Medical problem-solving: an exploration of strategies. Medical education 25 (1991) 196-207.

Rid92 Ridderikhoff J. The artificial and elusive demarcation between health and disease. Proc ??? , Budapest 1992.

Rid93 Ridderikhoff J. Information exchange in patient-physician encounter. A quantitative approach. Meth Inf Med 32 (1993) 73-8.

Rid93a Ridderikhoff J. Problem solving in general practice. Theor. Med., 1993;14:in press

Rub87 Rubinstein R, Hersh H. The Human Factor. Designing computer systems for people. Bedford (Mass), Digital press, 1987

Rus56 Russell B. Logic and knowledge. London 1956.

Sad74 Sadegh-Zadeh K. Subjektive Wahrscheinlichkeit und Diagnose. Meth inf med 13 (1974) 97-102.

Sag81 Sage AP. Behavioral and organizational considerations in the design of information systems and processes for planning and decision support. IEEE transactions on systems, man and cybernetics 11 (1981) 640-678.

Sca80 Scadding JG. The concepts of disease: a response. Psychological medicine 10 (1980) 425-7.

Sch63 Scheerer M. Problem-solving. Scientific american 208 (1963 #4) 118-31.

Sch88 Schaik FJ van. Effectiveness of decision support systems. Thesis, Delft 1988.

Sch79 Schw 1979

Sho79 Shortliffe EH, Buchanan BG, Feigenbaum EA. Knowledge engineering for medical decision making: a review of computer-based clinical decision aids. proc IEEE, 67 (1979) 1207-1224.

Sho87 Shortliffe EH. Computer programs to support medical decision making. JAMA 258, 1 (1987) 61-6.

Sho89 Shortliffe EH. Testing reality: the introduction of decision-support technologies for physicians. Meth inform med 28 (1989) 1-5.

Sim71 Simon HA, Newell A. Human problem solving. The state of the theory in 1970. Am. psych. 26 (1971) 145-59.

Sne73 Sneath PHA, Sokal RR. Numerical Taxonomy. The principles and practice of numerical classifications. San Francisco, W.H. Freeman & Cy, 1973.

Sob79 Sober E. The art and science of clinical judgment : an informational approach. in: Engelhardt HT, Spicker SF, Towers B. Clinical judgment, a critical appraisal. D. Reidel Publ.Co., Dordrecht (1979).

Sok74 Sokal RR. Classification: purposes, principles, progress, prospects. Science, 1974, 185, 1115 - 1123.

Ste89 Stefan MD, McManus IC. The concept of disease: its evolution in medical students. Soc Sci Med 29:7 (1989) 791-92.

Sut89 Sut (chapter 2: diagnostic accuracy of de dombal)

Tay76 Taylor TR. Clinical decision analysis. Meth Inf Med 15 (1976) 216-224.

Tay90 Taylor TR. The computer and clinical decision-support systems in primary care. J family practice 30,2 (1990) 137-40

Too81 Toon PD. defining 'disease' - classification must be distinguished from evaluation. J medical ethics 7 (1981), 197-201

Tve74 Tversky A, Kahnemann D. Judgment under uncertainty: heuristics and biases. Biases in judgment reveal some heuristics of thinking under uncertainty. Science 185 (1974) 1124-1131.

Vre93 Vreugdenhil H. Confidence in Knowledge. Thesis, Amsterdam, 1993

Wag64 Wagner G. Fehlerforschung als Aufgabe der Medizinischen Dokumentation. Meth Inf Med 3 (1964) 93-94.

Wag78 Wagner G, Tautu P, Wolber U. Problems of medical diagnosis. A bibliography. Meth Inf Med 17 (1978) 55-74.

War78 Wardle A, Wardle L. Computer-aided diagnosis - A review of research. Meth inform med 17 (1978) 15-28.

Was77 Wason PC. On the failure to eliminate hypotheses - a second look. in: Thinking, readings in cognitive science. Cambridge University Press (1977)

Wes89 Westerman, 1989, (chap 2)

Whe37 Whewell W. History of the Inductive Sciences. 3 volumes. London, J.W. Parker, 1837.

Whi81 Whitbeck C. What is diagnosis? Some critical reflections. Metamedicine 2 (1981) 319-29.

Whi81a Whitbeck C. A theory of health. In: Caplan AL, Engelhardt HT, McCartney JJ (eds) Concepts of health and disease: interdisciplinary perspectives. Addison-Wesly, Reading, Mass. (1981) 611-26.

Whi82 While R, Emrys-Jones GJH, Sprackling PD, Fell PJ. The experience of some pioneers. in: Malcolm A, Poyser J. (eds) Computers and the general practitioner. Oxford, Pergamon press 1982.

Wil76 Williams PW. Criteria for choosing subsets to obtain maximum relative entropy. The computer journal (1976) 21; 1: 57-65. (1977?)

Wor72 Wortman PM. Medical diagnosis: an information processing approach. Comp Biomed Research 5 (1972) 315-328.

You82 Young, D.W.: A survey of decision aids for clinicians. Br.Med.J., 1982, 285, 1332-1336

Zar83 Zarling EJ, Sexton H, Milnor P. Failure to diagnose acute myocardial infarction. JAMA 259 (1983) 1177-81.

Acknowledgements

The publication of this book marks the completion of a project which turned out to be of greater proportions than we had assumed before we tackled it. Many people have helped me with their work, advice and moral support over these past years. I want to thank the Institute for Family Medicine of the Erasmus University of Rotterdam for giving me the opportunity to work on a subject which must sometimes have seemed overly abstract and far removed from the practical reality of the general practitioner to them. I hope that the contents of this book will show them that this is not the case, and that we are dealing with very fundamental matters with potentially great practical consequences for every doctor. In particular, I want to thank my mentor, Co Ridderikhoff, who initiated the project, has shown unflagging support throughout its course, and without whom it could not have been finished. I thank professors H.J. Dokter and E.S. Gelsema for their assistance in the process of writing this thesis, and professors J.D.H. Wilson and J.D.F. Habbema for consenting to be on the committee. Furthermore I want to thank Mr. W. v.d. Berg, and Mark Smit, Sita Harkisoon, Melvin Jacobson, Prem Chitaroe, and Chi Woe for helping me out with the programming problems, and Mrs S. van der Made-Potuyt and Tonny Weijn for their analysis of the database structure. Mrs. Joyce Doets-de Wit checked the manuscript with me for errors in the use of the English language. I also wish to thank all the doctors associated with the Institute who made room in their busy schedules and freely gave their time to take part in the experiment as subjects. Finally, without Mr. W. Roelofsen who was ready to listen when there seemed little chance of success the work would not have been completed at all.

About the author

Egbert van Herk was born on the 8th of June, 1958 in Amsterdam. In 1976 he graduated from the Atheneum "Koninklijke Scholengemeenschap Apeldoorn", and went on to study chemistry for 3 years at the Rijksuniversiteit Groningen. After military service in 1979-1980 he switched to studying medicine at the Universiteit van Amsterdam. In 1988 he graduated as a doctor and started work as an AIO at the Institute for Family Medicine, Erasmus university Rotterdam. He is married and has a son.

Appendices

Appendix 1: Pre-enquete

Codenaam _____ systeemcode ___
(deze regel door onderzoeker in te vullen)

a) Algemene gegevens:

Datum ☐
Naam ☐ geslacht ☐
Geboortedatum ☐

b) Een paar vragen over uw medische achtergrond:

Wat is uw medische status?

co-assistent ☐
basisarts ☐
(huis)arts-in-opleiding ☐
huisarts ☐
anders, ☐ nl: _____

Hoelang bent u al arts? ☐ jaar

Wat is uw praktijkervaring als arts?

	jaar
Geen	☐
Waarneming/assistentie	☐
Opleiding	☐
Eigen praktijk	☐

c) Een paar vragen over uw ervaring met computers:

Hebt u zelf een computer?

Ja ☐

Nee ☐

Hebt u anderszins ervaring met computers?

geen ☐

gering ☐

redelijk ☐

veel ☐

Als u enige ervaring met computergebruik hebt, hebt u dan misschien ervaring met een van de volgende typen programma's:

wordprocessor (bv. Wordperfect)

geen ☐

gering ☐

redelijk ☐

veel ☐

database (bv. Dbase)

geen ☐

gering ☐

redelijk ☐

veel ☐

medische administratieve programma's

geen ☐

gering ☐

redelijk ☐

veel ☐

medische expertsystemen

geen ☐
gering ☐
redelijk ☐
veel ☐

computerspelletjes

geen ☐
gering ☐
redelijk ☐
veel ☐

een programmeertaal (bv. BASIC)

geen ☐
gering ☐
redelijk ☐
veel ☐

d) Een paar vragen over uw houding tegenover computers in de geneeskunde.

Geeft u nu alstublieft uw mening over de volgende stellingen, door een dwarsstreepje te zetten op de schaal die door het vakje wordt aangegeven.
Voorbeeld: Pit-bull terriers zouden verboden moeten worden.
oneens ═══════ eens

Ik denk dat een computer een nuttig hulpmiddel in de praktijk kan zijn, maar alleen voor het administratieve werk.

oneens ═══════ eens

Ik denk dat computers in de huisartsenpraktijk een bijdrage kunnen leveren aan de diagnostiek.

oneens ═══════ eens

Appendix 2: Post-enquete

Codenaam pp _____ systeemcode _____

Datum _____

Gebruik van de computer tijdens het consult.

In hoeverre stoorde het gebruik van de computer het verloop van de consulten?
niet _____ ernstig

In hoeverre vindt u dat het aantal door u verzamelde gegevens door het gebruik van het systeem toenam?
geen effect _____ belangrijk effect

Denkt u dat er door het gebruik van het systeem meer gegevens over de patiënt in het registratiesysteem van de arts terecht komen?
oneens _____ eens

Gebruik van het DDSS programma.

Wat vond u van het invullen en het overzicht van de personalia van arts en patiënt?
makkelijk ⊏═══════════⊐ moeilijk

Hoe gemakkelijk kon u de weg vinden in de menu's om een nieuwe patient te selecteren?
makkelijk ⊏═══════════⊐ moeilijk

Hoe gemakkelijk kon u de weg vinden in de niet-symptoommenus: personalia, diagnose, behandeling?
makkelijk ⊏═══════════⊐ moeilijk

Wat vond u van het zoeken in de hiërarchische structuur van het SCS ?(de eigenlijke symptoomboom)
makkelijk ⊏═══════════⊐ moeilijk

Vond u de in het SCS gebruikte terminologie
makkelijk ⊏═══════════⊐ moeilijk

Vond u dat de menustructuur u uitnodigde tot minder of meer vragen?
minder ⊏═══════════⊐ meer

Gebruik van de diagnostische support.

Vond u de door het systeem geboden diagnosen
overbodig ⊏═══════════⊐ zeer nuttig

Wat vond u van de scoregetallen bij de diagnosen?
overbodig ⊏═══════════⊐ zeer nuttig

Vindt u dat het vermelden van alternatieve diagnosen uitnodigt tot extra informatieinwinning van uw kant?
oneens ⊏═══════════⊐ eens

Als u de diagnostische hulp beschouwt als feedback op uw eigen handelen, vindt u dat dan
storend ⊏═══════════⊐ een goede hulp

Toekomst.

We willen graag nog uw verwachting weten over vier facetten van het systeem:

Het ddss kan in de toekomst een waardevolle bijdrage leveren aan de praktijkuitoefening van de huisarts, door standaardisatie van de terminologie.
oneens ☐ eens

Het ddss kan in de toekomst een waardevolle bijdrage leveren aan de praktijkuitoefening van de huisarts, door verbeterde patientenregistratie.
oneens ☐ eens

Het ddss kan in de toekomst een waardevolle bijdrage leveren aan de praktijkuitoefening van de huisarts, door hulp bij de diagnostiek.
oneens ☐ eens

Het ddss kan in de toekomst een waardevolle bijdrage leveren aan de praktijkuitoefening van de huisarts, door vergelijking van het gedrag van huisartsen.
oneens ☐ eens

Hebt u nog commentaar of suggesties?