Modelling medical diagnostic processes

Thesis

How to cite:

Alpay, Laurence Lea (1991). Modelling medical diagnostic processes. PhD thesis The Open University.

Modelling Medical Diagnostic Processes

Laurence Lea Alpay

MSc. in Knowledge Based Systems, University of Sussex, U.K, 1987
D.U.T d'Informatique, Universite Rene Descartes, France, 1981

Thesis submitted for the degree of Doctor of Philosophy in Cognitive
Science, The Open University, Milton Keynes
December 1990

Author's number: M7025239
Date of submission: 2 January 1991
Date of award: 6 March 1991

VOLUME 1

Abstract

The thesis investigates the development of medical reasoning processes and how student modelling of such processes can be achieved in intelligent tutoring systems. The domain of orthopaedics was chosen for the research. Literature has shown that medical reasoning has been modelled mainly from an expert point of view. The research problem addressed is to model explicitly various levels of medical expertise in terms of reasoning strategies. The thesis reports on a system, DEMEREST (DEvelopment of MEdical REasoning STrategies), a *developmental* user model component which describes successive stages of medical reasoning and which could ultimately be part of a medical tutor. The system diagnoses physicians' reasoning strategies, determines the level of expertise and produces a plan corresponding to the application of these strategies. As a basis of doing so, a set of seven reasoning strategies was identified in the medical problem solving literature. These strategies are based on generalisation, specialisation, confirmation, elimination, problem refinement, hypothesis generation and anatomy. An empirical study was carried out to examine the development of these strategies. Protocols of ten physicians at various levels of expertise were collected and analysed. A number of interactions of strategies at different levels of expertise was identified in half of these protocols and this information was used to construct a model of changes of strategies over time. Planning in artificial intelligence was used as a means of decomposing medical problem solving into a set of goals; the goals being associated with the reasoning strategies. By taking this approach, medical reasoning is viewed as a planning process. The remaining protocols from the empirical study were used to evaluate DEMEREST. The system was tested for its ability to determine a level of expertise for each protocol, model the reasoning strategies applied and their interactions, and generate a plan for each protocol. The assessment of the overall performance of the system showed that it was successful. This assessment also helped to identify conceptual as well as implementation constraints of the prototype system. The main result of the research undertaken in this thesis is that the design of the system DEMEREST demonstrates the feasibility of modelling the development of medical reasoning strategies and its usefulness for student modelling.

Dedication

To the memory of my friend Yvette Stone who died
in April 1990

Acknowledgements

I would like to thank the following people who helped me to make this thesis possible:

My supervisors Dr. Eileen Scanlon and Professor Tim O'Shea for their continuous encouragement and good advice throughout my Ph.D studies.

My external supervisor Dr Mike O'Neil from Imperial Cancer Research Fund (London) for his constructive criticisms. I also thank him for his active help in conducting the experiments and for providing me with necesary medical materials.

Dr Mark Elsom-Cook for his guidance and help in discussing and shaping my ideas during the early stages of my research.

The medical students and physicians from Milton Keynes General Hospital and Hanslope surgery who volunteered as subjects for the empirical section of this thesis.

All the research members of CITE for their help and encouragement, especially Pat Fung, Ann Jones and Maria Yannissi. who gave their time to read drafts of the thesis and provided me with useful feedback. Ann Blandford kindly accepted to become an independent assessor for my empirical data. Fiona Spensley and Josie Taylor who helped me to clarify some of the ideas of my research. Yibing Li, Royston Sellman and Magnus Moar provided technical help and allowed me to use their computers at much needed times. I am grateful to all of them.

Roshni Devi and Maria Yannissi for their invaluable friendship.

Mohammed-Ifti, whose love, support and help made this thesis possible.

My parents who have always encouraged me in my studies, and who I love and respect deeply.

CONTENTS

A. Empirical study

B. Listing of program

C. Output of the program

D. Testing

FIGURES and TABLES

Chapter Three

Chapter Five

Chapter Six

Chapter Seven

Chapter Eight

Chapter Ten

Chapter One

INTRODUCTION

The research problem

In this thesis, the development of medical reasoning processes and how such processes can be modelled in intelligent tutoring systems (ITS) is investigated. Modelling medical diagnostic processes has been expert-based. That is, the reasoning processes of medical students have usually been compared with those of the expert physicians and little attention has been given to modelling students' medical reasoning from the students' point of view. Moreover, the underlying assumption regarding the medical diagnostic processes used by students is that while students and more experienced physicians differ in the medical knowledge they possess, they tend to use similar reasoning processes to carry out medical diagnoses. A possible alternative to the expert-based approach that is being proposed in this thesis is to model medical diagnostic processes from a developmental perspective. In other words, to model different stages that one goes through; from starting as a medical student and eventually becoming expert physician. In particular, this is modelled in regards to the reasoning strategies applied. The thesis explores how this modelling could be applied in a student model component of an intelligent medical tutor. One of the aims is to identify the student's medical diagnostic strategies and determine her level of expertise. This information would then be passed on to the teaching module of a medical tutoring system.

The central research problem addressed in this thesis is to design the student modelling component of a medical tutor that would model explicitly various levels of medical expertise in terms of reasoning strategies.

Approaches to the research problem

Four complementary approaches to the research work are explored and applied:

- Literature review

The medical problem solving literature yields a number of results and hypotheses regarding the kinds of reasoning strategies medical students apply during the diagnostic process. A set of seven reasoning strategies was identified in the literature. The strategies are based on generalisation, specialisation, confirmation, elimination, problem refinement, hypothesis generation and anatomy.

- Empirical data

An empirical study was carried out to examine the development of these strategies. Protocols of ten physicians at various levels of expertise - from medical students to a specialist in the domain of orthopaedics - were collected and analysed. Subjects were put into a simulated consultation with a patient suffering from low back pain and were asked to diagnose the patient. Half of the protocols were used for the analysis. Reasoning strategies that emerged from the literature review were observed and a number of interactions of strategies at different levels of expertise were identified. From these interactions, a model of changes of strategies over time was constructed.

- Computational implementation

Planning in artificial intelligence was used as a way of decomposing medical problem solving into a set of goals, each of which corresponds to a diagnostic decision (e.g. take the patient history) and is associated with one or more strategies. By taking this approach, medical reasoning is viewed as a planning process.

A system called DEMEREST (DEvelopment of MEdical REasoning STrategies) was designed to illustrate the modelling and implementation of development of medical reasoning strategies. DEMEREST is a developmental user model component which describes successive stages of medical reasoning and could ultimately be part of a medical tutor. The system can perform the following tasks: diagnose a physician's reasoning strategies and their interactions, determine her level of expertise, and produce a plan corresponding to the application of these strategies.

- Testing

DEMEREST has been evaluated using the other half of the protocols from the empirical study. The system criteria for assessment were the three tasks mentioned above.

Structure of the thesis

The research undertaken is interdisciplinary because it brings together three research areas, namely, intelligent tutoring systems (ITS) in medicine, medical problem solving, and the development of expertise. It was therefore necessary to undertake a literature review for each of these research areas. These reviews presented in chapters 2, 3 and 4 provide perspective to the thesis, i.e. student modelling in medical tutors, medical reasoning of the student and expert physicians and developmental models of expertise in medical diagnosis.

Chapter two reviews ITS in medicine. The chapter examines approaches to the design of intelligent medical tutors, concentrating on the student models of these tutors. The conclusions that emerge from this review are that the main approach to developing medical tutors is based on expert systems and that student modelling in medical tutors is also expert-based and in general neglected. An alternative to the expert based approach would be to have a student model that would maintain a representation of the current state of student from the student's point of view rather than from the expert's.

Chapter three reviews the literature on medical problem solving and the teaching of medical diagnosis. This review provides the basis for a discussion on the features of a student model for a medical tutoring system. This chapter reviews models of medical reasoning focussing on students' reasoning. It also examines the use of these models in teaching medical reasoning. Findings from the review show that medical reasoning can be decomposed into its contents (i.e. the knowledge used) and its form (i.e. the reasoning itself that supports that knowledge), and that differences between medical students and experienced physicians are to be found in terms of the content of the diagnostic process rather than on its form.

The research direction taken in this thesis is to investigate further the form of the diagnostic process. Specifically, the research aims to show that novice and experienced physicians may not necessarily apply the same reasoning strategies. The form of the medical diagnostic process constitutes one of the features of the student model which is to be designed.

Another finding from this review is that students' reasoning has been examined from a developmental perspective. Since the research work aims to design a model that takes into account students' reasoning, this approach seems worth pursuing. Hence, the development of medical reasoning constitutes a second feature of the student model.

Chapter four is a literature review of developmental models, their application to medicine and their application in a tutoring context. Conclusions which can be drawn from this review are that limited research has been carried out with regards to the development of medical expertise and that it is confined to a theoretical level. Moreover, research on the development of medical expertise reported in the literature has focussed on the role of knowledge in developing expertise rather than on the role of reasoning strategies.

Chapter five synthesises important findings from the literature reviews. The interdisciplinary approach then leads to a specification of the design considerations necessary for a developmental student model for medical diagnosis called DEMEREST. The chapter describes DEMEREST and discusses the planning approach to building such a system. The chapter also examines how DEMEREST would be integrated into a medical tutoring system and reports on the medical domain of orthopaedics around which the system is implemented and subsequently tested.

Before implementing DEMEREST, two features of the system need to be researched: 1) the reasoning strategies which the system will model (reported in chapter six) and 2) the development of these reasoning strategies (reported in chapters seven and eight). **Chapter six** discusses the concept of reasoning strategy in the context of medical diagnosis and then reports on a number of reasoning strategies applied by medical students which were identified in the medical problem solving literature. The strategies need to be described in detail if the aim is to develop a system that recognises and diagnoses the reasoning strategies used by a physician (novice or experienced) during a consultation. The descriptions of the strategies stem from an investigation in the medical problem solving literature and from discussions with a medical doctor.

Chapters seven and **eight** report on an empirical study in which physicians at various levels of expertise were asked to diagnose a patient suffering from back pain. This study was undertaken to investigate the development of medical reasoning strategies. **Chapter seven** presents the methodology of the study, while **chapter eight** reports on the results and describes the modelling of the development of reasoning strategies. Results of the study showed no evidence of monotonic development of these reasoning strategies. However, some interactions of strategies at different levels of expertise were identified, which corresponds to changes of the medical reasoning. Using half of the data, a model of changes of strategies over time was constructed.

Chapter nine details the implementation of DEMEREST in LPA Prolog on a Macintosh SE. Implementation of the system and of the medical knowledge are described. Lessons learned from developing a prototype system that models changes of strategies over time are discussed.

Chapter ten is concerned with the testing of the system using the other half of the data from the empirical study. The system was tested for determining a level of expertise for each protocol, modelling the reasoning strategies applied and their interactions and generating a plan for each protocol. Given these criteria, the assessment showed that the overall performance of the system was successful and also helped in identifying conceptual as well as implementation constraints of the prototype system.

The thesis concludes with **chapter eleven** which summarises the contributions of this research to the areas of ITS in medicine, medical problem solving, and the development of expertise. As a prototype system, DEMEREST has demonstrated the feasibility and desirability of modelling the development of medical reasoning strategies for student modelling. The chapter also examines the limitations of the research undertaken and examines possibilities for further research and development.

Finally, a **glossary** of selected medical terms which appears in the thesis is provided (see appendix E).

Chapter Two

INTELLIGENT TUTORING SYSTEMS IN MEDICINE

This chapter provides a review of the literature on the area of intelligent tutoring systems in medicine (ITS) and drawn on the work reported in (Alpay 1988a). This review serves the following purposes: 1) to examine various approaches adopted in the design of ITS in medicine and 2) to establish which area of intelligent medical tutor research this thesis work should address. In the process of reviewing, two types of medical tutoring systems were found; systems which had been specifically designed for teaching and those which had been designed primarily for another purpose but adapted later for teaching. For the latter systems, educational principles have been secondary. Medical tutors and adapted medical tutors are discussed in the first two sections. Some conclusions from these discussions are drawn in the last section.

2.1 Medical tutors

Literature on ITS in medicine shows that medical tutoring systems are limited in number. This section presents these tutors which are also the main ones found in the literature.

GUIDON and instructional systems derived from GUIDON

GUIDON (Clancey 1979) is the most important medical tutor known up to date. The system teaches diagnostic problem solving in the domain of infectious diseases and uses a mixed-initiative dialogue i.e. engages a student in a dialogue about a patient suspected of having an infection and helps the student consider relevant clinical and laboratory data for reaching a diagnosis.

The design of GUIDON raised an unanticipated epistemological issue regarding the organisation of expert systems. It was found that MYCIN's rules (Shortliffe 1976) embody implicit knowledge essential for tutoring. As a result, a complete reconfiguration of MYCIN's rules led to the design of NEOMYCIN which makes explicit the structural, strategic and support knowledge that was compiled in MYCIN. A collection of instructional programs was developed using NEOMYCIN's knowledge base. These are reported below.

GUIDON2 (London and Clancey 1982) is a tutoring system that uses the case method approach to teach medical diagnosis (as in GUIDON). The student modeller component of GUIDON2 is the most interesting feature of the system as it combines two complementary searches (i.e. a top down model driven simulation of the expert and a bottom up data driven search) to understand the student's behaviour.

GUIDON-WATCH (Clancey and Richer 1987) is an advanced interface which makes extensive use of graphics to allow the student to browse through and study NEOMYCIN's knowledge base and view reasoning processes during diagnostic problem solving. GUIDON-WATCH provide facilities to the student which should enhance and improve teaching in the medical field.

GUIDON-MANAGE (Rodolitz and Clancey 1989) introduces students to a set of tasks (in this case NEOMYCIN's tasks e.g. test a hypothesis) that will help them structure and articulate a strategy for diagnosis and provides an environment in which to experiment with these tasks.

Additional medical tutors

Besides GUIDON and its derived systems, two tutors which are on a smaller scale than GUIDON, have been recently developed. A radiology tutor (Sharples and duBoulay 1987) is a system for developing the skills of interpreting cardiac X-rays. The system is designed to provide a 'refresher course' rather than initial teaching. The main contribution of this tutor can be seen in the representation of the domain knowledge where the relationships between pathologies and anatomical features are described.

Another recent tutor is a primary care tutor (McGregor et al 1988) which is based on using a computer assisted learning (CAL) for teaching undergraduate students in General Practice. The tutor provides medical students with a tool to improve and extend their knowledge in patient management and case diagnosis. While some of the features of the primary care tutor are based on GUIDON (e.g. student modelling, mixed initiative dialogue) the interesting aspect of the system is its combination of a CAL system and an ITS system.

2.1.1 Discussion of medical tutors

The focus of the following discussion is on the approaches adopted in developing the tutors and specifically to examine the issue of student modelling, in the context of designing ITS.

Approaches to the development of medical tutors

The main approach in developing medical tutors has been *expert systems oriented.* Expert systems have been the main resource in the development of ITS in medicine (Clancey and Shortliffe 1984). GUIDON is based on the expert system MYCIN while the instructional tutors derived from GUIDON such as GUIDON2, GUIDON-WATCH and GUIDON-MANAGE are based on the medical consultation system NEOMYCIN. The principal component of the expert system in the design of these tutors is the domain of expertise; that is the knowledge base of the expert system. GUIDON makes use of MYCIN's knowledge base for its teaching contents, while the instructional tutors derived from GUIDON contain the knowledge base of NEOMYCIN which makes explicit different kinds of knowledge. In contrast to GUIDON and its subsequent tutors, the knowledge base of the radiology tutor is not based upon an already existing expert system. However, its knowledge base corresponds to the expertise of a human expert radiologist. Therefore, the radiology tutor can also be thought of as expert based.

The second approach in developing medical tutoring has been *CAL based.* The primary care tutor contains a CAL system along with an intelligent tutor. CAL systems have been used widely in medical education (Chard 1988). They incorporate well prepared course material in lessons. In a CAL system, the student is usually given some instructional text and is asked a question which requires a brief answer. After giving her answer, the student is told whether she is right or wrong. In some cases, the student's response may be used to determine her path through the curriculum. When the student makes an error, the program branches to remedial material. In contrast to branching CAL systems, an ITS is characterised by having the subject matter represented independently from teaching knowledge and a dialogue is carried out with the student and the student's mistakes are used to diagnose her misunderstanding. Compared to the expert system approach, the contribution of the CAL system to the design of the tutor is not related to the domain knowledge. In the primary care tutor, the CAL system is used for carrying out such tasks as the generation of a patient case, generation of the tutor's hypothesis list, whereas the intelligent tutor takes care of the tutoring environment, in particular, the dialogue management and student modelling.

Design of medical tutors

There are a number of issues involved when designing an intelligent (medical) tutor such as representation of the domain knowledge, teaching strategies and student modelling.

- **The representation of the domain knowledge** to be taught has been investigated in detail with the reconfiguration of MYCIN knowledge base into NEOMYCIN. Lessons learned from using MYCIN knowledge base for tutoring had implications not only for the design of tutors derived from GUIDON but also for other systems aiming to use an expert system knowledge base for tutoring. Clancey (1987) claims that other representational notations used in medical expert systems (e.g. frames, semantic networks) would bring the same type of problems encountered in GUIDON, and consequently it is important to establish an understanding of the knowledge contained in the system within the epistemological framework. In other words, this means that not only MYCIN but also other medical expert systems would not be effective when used in a teaching mode. The knowledge base of the radiology tutor partially fits into the epistemological framework put forward in NEOMYCIN since its knowledge base contains pathologies and anatomical features (represented as frames) which correspond to structural and support knowledge of the epistemological framework. The knowledge base of the radiology tutor has been built with the tutoring task of the system in mind. In contrast, the medical database of the primary care tutor, which is a medical database originally built to be used by the CAL system, has remained unchanged in terms of its application with the primary care tutor.

- **Teaching strategies** incorporate knowledge about how to teach. Ideally, teaching strategies should integrate knowledge about natural-language dialogues as well as knowledge about teaching methods. Their goals should be to communicate with the student, select problems for her to solve, monitor and criticise her performance, provide assistance, and select remedial material. They should also take into consideration issues such as when it is appropriate to offer a hint or how far the student should be allowed to follow a wrong path of reasoning. Given the above tasks that teaching strategies should perform, the tutorial rules in GUIDON are quite complete and successful. They cover a wide range of tutorial methods. These include, for instance, ways of discussing domain rules, responding when student requests case data and ending the discussion of a topic.

Other tutorial rules are used to maintain the student model or to select valid questions when quizzing about a rule. GUIDON does not follow a socratic or a coaching style of tutoring but adopts a mixed initiative dialogue. In turn, the student or the system can be active in the tutoring session. However, the system still keeps control of the interaction and does not allow total freedom to the student. The discourse and teaching strategies in GUIDON are also used in GUIDON2. Furthermore, the teaching approach in the primary care tutor is based upon GUIDON's - a mixed initiative dialogue between the tutor and the student - whereas the teaching strategies in the radiology tutor are based upon educational psychology.

- A third issue in designing intelligent tutoring systems is **student modelling.** The purpose of the student model is to represent the student's understanding of the domain being taught and to maintain a representation of the current knowledge state of the student. The role and importance of student models have been stressed in the literature. For example, Hartley (1973) has proposed a framework for teaching systems which contains a representation of the student (student model) as one of the components, required for designing these systems. More recently, Self (1988) has argued that any ITS needs a student model and Laurillard (1988) has examined how well student models meet students' likely needs. In the context of medical diagnosis, the student model should be able to represent, for instance, what the student knows about the patient case, what medical knowledge the student used, and what reasoning processes the student has applied to diagnose the patient case. As discussed above, the representation of the domain knowledge and the teaching strategies are two areas of ITS which have been tackled in different ways in the medical tutors (e.g. teaching strategies and representation of the domain knowledge differ in GUIDON and the radiology tutor).

In contrast, it was found that, where it exists, student modelling in the medical tutors followed a single line of approach. Given that the development of medical tutors has been mainly based on expert systems, student models in these tutors have also been constructed from this approach. So GUIDON, the radiology tutor and the primary care tutor contain a subset model (also referred to as an overlay model). In a subset model, the student's knowledge is seen as a subset of the expert's knowledge. The student's understanding is represented completely in terms of the expert component. The student model is built by comparing the student's behaviour to that of a computer-based expert in the same environment. In the case of the radiology tutor, the student model is a subset of a pathology, that is, the student's feature values for an image is seen as a subset of the feature values for a pathology. In the case of the primary care tutor, the student's management plan for the patient case is viewed as a subset of the management plans of the expert general practitioner.

The exception to the above is the student model of GUIDON2 called IMAGE. This is not strictly speaking a subset model. However, the multiple predictions that the model can produce, using a top down search, correspond to a simulation of the expert. Therefore, IMAGE nevertheless takes into account the expert's behaviour in order to represent the student's state of knowledge.

The medical tutors derived from GUIDON such as GUIDON-WATCH and GUIDON-MANAGE do not contain student models. Both systems are part of a larger project to develop a series of tutoring systems that combine to give a student a comprehensive introduction to the process of diagnosis. Each system concentrates on a specific aspect of teaching and so does not contain a user model but assumes that one would be present. For instance, the problem addressed in GUIDON-WATCH is that of the user interface, in particular of browsing and viewing a knowledge based system. In GUIDON-MANAGE, the focus is the language of diagnosis. Specifically, GUIDON-MANAGE lets the student manage the diagnosis by explicitly applying strategic tasks of NEOMYCIN.

Apart from the above susbet model, other kinds of student models for intelligent tutors have been proposed such as the perturbation and bounded user models though they have not found application in medical tutors. A possible explanation for this is that student modelling has not been the main focus in the design of these tutors, and rather, the aim has been to develop workable medical tutors. In the perturbation model, the student's knowledge is seen as perturbation from the expert's knowledge. It is assumed that the student has mislearned skills in the domain. Hence, there is a second set of skills called bugs which are not possessed by the expert but which correspond to erroneous versions of the correct skills. The perturbation model has been used in a number of tutoring systems in other domains than medicine (e.g. DEBUGGY, Burton 1982).

In the bounded user model, the idea is that instead of building an exact model of the student's knowledge, upper and lower bounds are put on the knowledge and an attempt is made to draw these bounds together. This is achieved by using a model of the learning process to infer these bounds from the same observations that the student can make. These bounds are then regarded as hypotheses which can be tested by using them to make predictions about the concrete observations of the student. A bounded user model was part of IMPART, a system intended to guide a discovery learning environment (Elsom-Cook 1987).

2.2 Adapted medical systems for teaching

This section reports on three medical systems which have *not* been designed as tutors but which have been adapted to provide some tutoring environment. These tutors are the major ones found in the literature.

ATTENDING

The main feature of ATTENDING (Miller 1984) is its critiquing approach, a form of explanation used in expert systems. ATTENDING is designed to critique a physician's plan for a patient's anaesthetic management. In critiquing, the system discusses the pros and cons of the proposed approach as compared to alternatives which might be reasonable or preferred. In order to promote the teaching activity, a "teaching interface" has been developed. ATTENDING illustrates how easily a system can be tuned to a limited teaching mode using the critiquing approach: the teaching interface does not involve any modification of ATTENDING itself. It includes different modes where the student is asked to i) propose a reasonable plan for ATTENDING to critique, ii) critique that plan herself and then to compare her critique with ATTENDING's or iii) propose a deliberately poor plan.

QMR

QMR (Miller, Massarie and Myers 1986) is a microcomputer based decision support system designed to provide diagnostic assistance in the field of internal medicine. The QMR knowledge base represents reworking, extension and expansion of the concepts and contents of the knowledge base of INTERNIST-1 - a medical expert system in the domain of internal medicine (Miller, R. et al 1984).

QMR does not have an additional module for teaching. However, QMR demonstrates how the three levels of functionality available to the users coupled with the sound knowledge base of internal medicine can be used for teaching purposes. For instance, by using QMR as an electronic textbook, students can learn a great deal about diseases in internal medicine and how to evaluate the medical literature. In the diagnostic spreadsheet mode of QMR, students have the opportunity to spend selective time constructing disease profiles. In the expert consultant mode (Miller, Massarie 1989), students can learn how to approach medical diagnosis. Another educational potential of QMR is in the generation of simulated patient cases by appropriate manipulation of information in the knowledge base (Parker, Miller 1989).

SPHINX

SPHINX (Fieschi, M. 1984) is a medical expert system which aims to assist physicians in the diagnosis and treatment of patients. The domain of expertise of the system is the diagnosis of diabetes. In order to use SPHINX as a teaching tool (Fieschi, D. 1984), an additional knowledge base was added. This complementary knowledge base contains explanations related to the pathophysiological knowledge that are embodied in the expert knowledge. In addition, the teaching aid of SPHINX contains a teaching module with metarules for i) managing the consultation and generating required explanations and ii) directing the teaching. In a tutoring mode, the student is placed in a simulated consultation with a patient similar to a real situation. In other words, the student is expected to play the role of the physician. The student can, for example, propose findings to look for or request investigations.

2.2.1 Discussion of adapted medical tutors

The focus of the following discussion is on the approaches used in developing the tutors and on student modelling. This discussion examines whether the conclusions drawn in section 2.1.1 regarding the medical tutors still hold for the adapted medical tutors.

Approach to the development of adapted medical tutors

All the adapted medical tutors discussed are *expert systems based.* This approach was also adopted in developing the medical tutors (section 2.1.1). As mentioned in the case of ATTENDING, the system is centered around a form of explanation used in expert systems called critiquing. In the critiquing approach, ATTENDING assumes that the physician has already evaluated a patient, and has already thought about possible management for that patient. Critiquing also implies the precondition that the user is competent in the field being critiqued. This is not a restriction on the critiquing approach but rather reflects the medical reality of the physician having the basic competence to evaluate the advice given. This approach may not therefore be well suited to tutoring novices. QMR is based on the expert system INTERNIST, while the teaching aid system of SPHINX is directly built around the expert system SPHINX. In the case of QMR, its knowledge base is vast and can be viewed in various ways in particular for tutoring purposes, whereas in the case of SPHINX a knowledge base has been added to the existing one for more explanation.

Design of adapted medical tutors

Issues of representation of the domain knowledge, teaching strategies and student modelling are examined.

- **Representation of the domain knowledge.**

In all the three systems, the construction of the knowledge based has been important. In ATTENDING, the teaching feature of ATTENDING is also used as a tool for addressing the problems of maintaining a complete and consistent knowledge base. That is, the tutorial component of ATTENDING is to help test and debug the system knowledge. Moreover teaching is seen as mode of expert system validation and evaluation of the knowledge base. In QMR, the knowledge base is the main strength of the system as every activity turns around it. In the teaching of SPHINX, the additional knowledge base can be viewed as the support knowledge found in the epistemological framework proposed by Clancey.

- **Teaching strategies.**

Though to a certain extent, one might classify the different modes of critiquing as teaching strategies, these are a poorly developed feature in ATTENDING. The way ATTENDING communicates with the student is more like a helper than a teacher. Although the educational value of ATTENDING is limited however, its teaching facility demonstrates that (1) hypothetical patients used for teaching can be selected in areas where the system's knowledge is strong; (2) an extensive knowledge base is a very useful attribute of a tutoring system; and finally (3) it can give practical feedback to the designers of the system. QMR does not have any teaching module attached to it and hence does not incorporate any teaching strategies. QMR is not a tutor as such, but offers a self teaching environment where the student can explore medical knowledge in a flexible way and strengthen her medical knowledge. In the case of SPHINX the teaching aid has a teaching module which manages the dialogue between the student and the system and therefore controls the teaching process by providing the student with the information or pathophysiological explanations requested.

- **Student Modelling in adapted medical tutors.**

ATTENDING, QMR and the teaching aid of SPHINX do not contain any explicit student model. In other words, these systems do not keep track of what the student does or of her progress. In the case of ATTENDING and QMR, there is no reference to further modifications of the systems to incorporate student models. In contrast, in the case of the teaching aid of SPHINX, there is a mention in the architecture of the teaching aid making use of a student model. This eventual student model has not been implemented in the current version of the system but could be incorporated in a later version. The fact that these systems have not been built as tutors in the first place means that they do not incorporate all components (such as student models) commonly found in an intelligent tutoring system.

This discussion of adapted medical tutors has shown that similarly to the development of the medical systems built primarily for teaching, the prevalent approach has been to develop expert systems. Moreover, student modelling in adapted medical tutors was also found to be an issue which has been relatively neglected and calls for further investigation.

2.3 Summary

This chapter has provided a literature review on intelligent tutoring systems in medicine. Medical tutors which were built primarily as tutors as well as medical systems which were adapted for teaching were reviewed. The aims of this review were to examine the various approaches to designing ITS in medicine, and establish one issue of an intelligent medical tutor towards which the research should be directed. The conclusions that emerge from this review are summarised as follow:

i) medical tutors and adapted medical tutors are largely expert based

and

ii) a neglected area of development of medical tutors is student modelling.

Student models used in ITS in medicine have been expert based (i.e. subset models). Moreover, student models not only have been restricted in their kind, but also they have been omitted from many medical tutors. From these conclusions, it is proposed that a student model for an intelligent medical tutor should maintain a representation of the current state of the student from the *student's point of view,* rather than from the expert's perspective.

The next chapter reviews the literature on medical problem solving and teaching of medical diagnosis to establish the essential features of a student model that takes into account the student's medical reasoning.

Chapter Three

MEDICAL PROBLEM SOLVING and TEACHING OF MEDICAL DIAGNOSIS

The previous chapter was a review of intelligent tutoring systems in medicine. The main conclusion from the review was to establish that there was a need for further research on student models for intelligent medical tutors, based on students' medical reasoning. This chapter is a review of the literature on medical problem solving and teaching of medical diagnosis drawn on the work reported in (Alpay 1988b). The aim of this review is to discuss the features of such a student model. In particular, this chapter reviews:

i) Models of clinical reasoning with particular reference to students' reasoning. In order to investigate how medical students make diagnoses one starting point has been to examine studies of medical reasoning of medical students as well as of expert physicians, and hence to focus on differences between novices and expert physicians.

ii) The teaching of models of clinical reasoning. In order to work on a tutoring environment, or a component of a tutoring environment for medical students, one should examine the teaching methods which are used in medical schools and the models of clinical reasoning students are expected to learn.

The structure of the chapter is as follows: firstly, numerical and psychological approaches to medical reasoning are reported and discussed, followed by a review and discussion of the teaching of medical reasoning. Finally, conclusions of this literature review are summarised.

3.1 Numerical approaches to medical diagnosis

Some of the early work to characterise medical diagnosis concentrated on models in which the clinician's thinking was centered around an "input" (i.e. medical information) and "output" (i.e. diagnosis) relationship without regards to what happens in between i.e. the process of diagnosis. Since these models consider neither the medical reasoning process of the physician nor the medical student per se, they cannot provide an insight to that process. There is still some support for numerical approach to medical diagnosis. For instance, Lindley (1985) favoured the usefulness of statistics in decision making. In his book, he looked at the problems involved in decision-making (e.g. medical diagnosis) and argued that there is only one logical way to make a decision. He proposed the use of three basic principles - (1) assigning probabilities to uncertain events; (2) assigning utilities to the possible consequences; and (3) choosing that decision that maximises expected utility. By using these principles, decisions can be reached more efficiently and with less disagreement. These basic principles show that

> "...any deviation from the precepts [principles] is liable to lead the decision-maker into procedures which are demonstrably absurd - or as we shall say, incoherent" (p.vii).

Numerical approaches to medical diagnosis have come under strong criticism (e.g. Gale 1983, Fox 1988). Lindley's approach to decision making seems too rigid and erroneous for most cases. In many instances, there may be more than one strategy to adopt, and it may not be possible to formulate an optimal choice in advance. A counter-example is game theory (Manfred and Rutheld 1981) which was developed to deal with problems of this kind. Game theory is concerned with the general analysis of strategic interactions. It helps to describe a game by indicating the payoffs to each of the players for each configuration of strategic choice to make. Hence, making a decision by using the tool of game theory demonstrates the vital role of strategic interactions in arriving at a decision.

3.1.1 Statistical Models

The most popular statistical models have been those based on Bayes' theorem. Bayes' theorem provides a mechanism

> "to calculate the probability of a disease, in the light of specified evidence, from the a priori probability of the disease and the conditional probabilities relating the observations to the diseases in which they may occur" (Shortliffe et al 1979, p.1214).

Yet, the Bayesian approach has several limitations. Firstly, it assumes conditional independence of symptoms which usually does not apply and can lead to substantial errors. Secondly, it assumes mutual exclusiveness and exhaustiveness of disease categories (i.e. the patient is assumed to have exactly one of the n diseases) which is often false (Shortliffe et al 1979). Finally, in many domains it may be inaccurate to assume that relevant conditional probabilities are stable over time. Nevertheless, Bayesian models of the diagnostic process have been widely used. One example is the work of deDombal (1988) who developed a computer-based decision aid using Bayes' theorem for determining the cause of acute abdominal pain.

3.1.2 Models Based on Decision Theory

In addition to statistical models, other models based on decision theory (such as decision trees, decision analysis) have been used. Decision trees provide a way in which the decision making process can be seen as a sequence of steps in which the clinician selects a path through a network of plausible events and actions. Decision analysis can be seen as an attempt to consider values associated with choices as well as probabilities, in order to analyse the processes by which decisions are made or should be made. The program developed by Gorry et al (1973) for the management of acute renal failure is an illustration of applying decision theory approaches as an aid to diagnosis rather than as a representation of the diagnostic thinking process.

3.1.3 Discussion of numerical models

The main critique of numerical models is that they do not describe the process of medical diagnosis physicians use, but rather prescribe ways in which medical information may be manipulated to reach the most likely diagnosis. This has important implications in the context of intelligent tutoring systems. Using numerical models within an intelligent tutor to teach medical diagnosis seems inappropriate. In order to teach medical diagnosis, one needs to have a description of it. Moreover, one aims at teaching how to make the proper diagnosis and not just how to reach the most likely one.

Numerical models work well in certain situations where a finite number of diagnoses can be known in advance. However, this is not always possible and as Gale and Marsden points out (1983)

> "...they show a clear lack of congruence with the realities of clinical practice." (p.7).

These models can be viewed as "black box" type because they only replicate the final diagnosis of the clinician without her mental steps used to reach that diagnosis. Furthermore, the comparison between a clinician's diagnostic thinking processes and statistical methods is not possible because by nature they are not compared on equal terms (Gale and Marsden 1983). The real comparison is between a statistical model and a clinician who has been given clinical data in an unfamiliar form in an unfamiliar context and is thus attempting to behave as a statistical model.

In summary, models based on statistics and decision theory are inappropriate for describing the medical diagnostic processes and hence for facilitating a better understanding of those processes. In contrast, research which has attempted to describe the diagnostic thinking process is based on psychological studies of the diagnostic thinking processes of medical students and of expert physicians. In the next section some of the work which has been carried out in that direction is reviewed and discussed.

3.2 Psychological approaches to medical reasoning

Empirical studies on the medical reasoning process can be divided into four areas (generic, contents, process and development), each of which corresponds to a specific perspective from which medical reasoning has been studied. This survey starts with an examination and evaluation of early, classic studies which provide a generic form of clinical reasoning. Then, recent views which emphasise medical contents that support a generic form of clinical reasoning are reported. Thirdly, research which aims to reformulate clinical reasoning as a process is examined. The last part discusses studies which intend to describe the development of clinical reasoning. In the scope of this survey, particular attention has been paid to the studies that attempt to understand clinical reasoning from a novice-expert perspective. The following format has been used to report each study: 1) the aim of the study, 2) the subjects taking part, 3) the methodology, 4) the results of the study and the proposed model of medical diagnosis and 5) a critique of the study, in particular regarding models of reasoning of the novices versus of the expert physicians.

3.2.1 Generic form of medical reasoning

In this section two empirical studies of the medical diagnostic process are reported and discussed, those of the Michigan group (Elstein et al 1978) and those of the McMaster group (Barrows and Tamblyn 1980).

Aim of the Michigan group's research and subjects taking part

The aim of this research conducted by Elstein et al (1978) was to understand medical practice better and thereby to improve the instruction and performance of present and future practitioners. Subjects were experienced physicians and 2nd year medical students.

Methodology for the Michigan group's research

The physician was to interact with a simulated patient. The interview was videotaped and transcribed. During the interview, the physician was asked to think aloud. The interview was recalled, and hence additional comments from the physician were obtained.

Model of diagnosis proposed by the Michigan group

Elstein et al (1978) characterised the diagnostic thinking process as one of "hypothesis generation and testing". They suggested that clinicians use early case cues to generate sets of tentative hypotheses for the patient's condition. These hypotheses are then used to structure and guide further interrogation of the case. Hypotheses provide expectations for additional clinical manifestations that should be present if a hypothesis is true for the patient case, and the findings of the patient are compared to expectations to select among the alternatives. Hypotheses and hypothesis sets can be restructured or changed as the diagnosis progresses.

Aim and methodology of the research of the McMaster group

Barrows and Tamblyn (1980) work within the frame of reference of the problem-based learning approach applied to medical education. Problem-based learning is defined as

> "...the learning that results from the process of working towards the understanding or resolution of a problem." (p.18).

This approach has some advantages. Firstly, learning through problem-solving is much more effective than traditional memory-based learning to create in the student's mind a body of information usable in the future. Secondly, the most important skills of the physician for the patients are problem-solving skills not memory skills (Barrows and Tamblyn 1980).

Subjects taking part in the research of the McMaster group

Unlike the Michigan group study which carried out empirical studies to built a model of medical reasoning, the McMaster group used the problem-based learning approach and tailored it to medical education which was to be used by medical students.

Model of medical diagnosis proposed by the McMaster group

The work of Barrows and Tamblyn supports the general description of the diagnostic thinking process given by Elstein et al (1978). Barrows and Tamblyn summarise the significant aspect of the process of clinical reasoning as a five-step process which is claimed to be in sequence: Firstly the physician perceives initial cues from the patient and environment. She then generates multiple hypotheses. Thirdly, she applies an inquiry strategy (e.g. questions, examinations, tests) to refine, rank, verify and eliminate hypotheses. In a fourth step, the physician enlarges the formulation of the patient's problem from significant hypothesis-related data obtained from ongoing inquiry. Finally, the physician reaches diagnostic and therapeutic decisions about the patient's case.

Critique of the generic form of medical diagnosis

The studies of Michigan and McMaster groups have been criticised. Gale (1980) criticised both studies for

> "...spurious quantification and over-interpretation of inexact data." (p.70).

She argued that an emphasis on quantitative method limits greatly the studies to quantify elements such as counts and weightings of cues and hypotheses that are essential but not necessarily sufficient. In other words, these two studies can provide, for instance, information about how much data is collected and how it relates to hypotheses. However, they do not give any indication, for instance, of the clinician's cognitive manipulation of cues, how these are cognitively structured as clusters, and what thinking process generates hypotheses.

Another assessment of the work of Elstein et al (1978) and Barrows and Tamblyn (1980) comes from Feltovitch and Patel (1984b). Their evaluation and analysis of these studies support Gale's view. One criticism is that even though these studies establish important characteristics of clinical reasoning, they also bring unexpected results. For instance, most of the measures of the clinical process which were studied do not differ among clinicians with greater or less experience, or of different externally judged expertise. These measures are for instance, of timing (e.g. percentage of cues to first hypotheses) or of number (e.g. number of hypotheses generated and maintained in active consideration). The problem with these measures is that instead of taking into account of what is being considered within the clinical process, they take into account the general form of that process.

The models of the clinical reasoning process proposed by the two groups do not offer enough flexibility. They force the physician into a framework of clinical reasoning which she may not want to be put into. Even if the hypothesis generation and testing method is the one physicians use in many cases, there may be other instances in which a physician will follow another path of reasoning. This issue is related to the nature of the differences in reasoning between expert and less expert physicians. Elstein et al (1978) studied mainly experienced physicians. Although their work was not primarily concerned with comparing novices and experts, they suggested that the main difference between subjects was to be found in the repertoire of their experiences organised in long term memory rather than in the planning and problem-solving heuristics. In other words, they did not find any differences in the reasoning between novices and experts, and since they argued that expert doctors use hypothetico-deductive reasoning, it also suggested that medical students use that kind of reasoning. Barrows and Tamblyn do not explicitly characterise the reasoning of medical students. However, they expect students to apply the clinical reasoning process at a certain phase of their problem-based learning process. Section 6.2 will discuss further the kinds of strategies that students apply.

3.2.2 Contents of medical reasoning

Rather than focussing on the general form of the clinical reasoning process, recent studies on clinical reasoning have focussed on the medical content which supports that reasoning. This section reviews the work of Feltovitch which demonstrates the importance of medical knowledge in pediatric cardiology.

Aim of Feltovitch's study

The work of Feltovitch et al (1984a, 1984b) illustrates a clear change in the nature of psychological investigations. As was mentioned in the previous section, Feltovitch criticises the work of the Michigan group because they only offer a general form of the process of clinical diagnostic reasoning and neglect its contents, that is, the knowledge of the medical domain involved in the diagnostic process. In fact, unlike Gale, Feltovitch et al do not reject the generating and testing hypotheses process suggested by Elstein et al. However, in the view of Feltovitch et al, the hypothesis generation and testing method is a general problem solving procedure for diagnosis and is a weak method which is applied widely but does not guarantee success. It is the knowledge base which plays the central role in the diagnosis. Feltovitch et al carried out an experimental study of the knowledge base of expert and novice physicians, in which the effects of medical knowledge on the clinical reasoning process were investigated as were the changes in such knowledge as individuals gain experience of the task of medical diagnosis and with the subject matter of a domain of medicine.

Subjects taking part in Feltovitch's study

Subjects were students, residents[1] and specialists varying in their training and clinical experience in pediatric cardiology.

Methodology of Feltovitch's study

Subjects were to diagnose four cases of congenital heart disease while thinking aloud. Each case was designed to assess a different aspect of the subjects' medical knowledge (e.g. assess subjects' differentation of diseases into subtypes). Starting from the concept of frames, Feltovitch et al speculated on a knowledge structure for the organisation of disease models in memory. Basically it was hypothesised that the disease frames are organised in three different levels. At the most global level, individual disease frames were organised within frames representing general disease categories. At the intermediate level were classical disease instances of the category. At the last level for each disease were a set of frames specifying variations and subtypes

---

[1] In the U.K. system, a resident is equivalent to a house officer (see section 7.1. for scale of medical expertise).

of the diseases. Feltovitch et al (1984a) put forward several hypotheses about the nature of the knowledge base differences between experts and novices: the knowledge base of the expert being dense and precise, and the knowledge base of the novice being sparse, classical and imprecise.

Since the study focussed on the quality of the reasoning, Feltovitch defined for each case sets of diseases which correspond to reasonable interpretations for the case and which are easily confused with each other. This was done through expert consultations and reviews of medical literature. These predefined disease sets are called *logical competitor sets* (LCS). These LCS constitute good hypotheses to be considered for the case, and represent a criteria for differentiating novice from expert subjects.

Results of Feltovitch's study

Reports of the study showed that

> "Results were generally consistent with predictions." (Feltovitch et al 1984b, p.5).

That is, some of the hypothesised features of the knowledge base for novices and experts (such as density, precision and classicality) were proven. For instance, results demonstrated that when experts generate one of the LCS diseases at either the disease or variant level, they usually generate them all. This exemplifies the knowledge base of the expert as being dense since it means that when one disease in a category is activated, other diseases will be as well.

For the cases of the study, Feltovitch et al (1984a) identified an *expert form* and an *expert substance* for diagnosis. Expert form involves the full, active use of a LCS for each case. The LCS corresponds to diseases that have similar physiological structure and clinical presentation. Expert substance refers to correct data evaluations, within the LCS of diseases, which are necessary to isolate the correct member of the set. The results indicated that, unlike experts, medical students with little training and clinical practice in the field showed neither expert form nor expert substance. That is, regarding the expert form, students almost never considered the full LCS and focussed on the classic members in cases that encouraged this. This means that LCS members when they exist at all may be represented in a more isolated form in memory. While the results indicated a clear distinction between the knowledge base of experts and of medical students, experiments show that sometimes residents behave like experts and at other times like medical students. It was concluded that the main problems of the non expert physicians were a lack of connections in memory among LCS members or imprecision in the knowledge necessary for discriminating LCS members correctly.

Critique of Feltovitch's study

Although the Feltovitch's study may have been too specific, it has shown that what is crucial to successful diagnosis and does discriminate expert from less expert performance are diagnosticians' disease knowledge, a memory store of the disease models and the memory organisation among them. They have characterised some differences of the medical knowledge base between medical students and experienced physicians from a *knowledge base perspective*. However they have assumed that all the subjects (novices, intermediates and experts) used the same general approach to medical problem-solving, that is, the hypothesis generation and testing method.

The work of Feltovitch et al has led other researchers to investigate further the content and structure of the medical knowledge in respect to its role in diagnosis. Bordage and Zacks (1984) have studied the structure of medical knowledge in the memories of pre-clinical medical students and practitioners. They have given evidence that medical categories are organised around prototypes (e.g. key factors, clear examples), and hence questioned the relation of the memory structure to the clinical problem-solving process. Moreover, they have shown that the prototype view may help facilitate the understanding of learning and problem-solving in medicine.

3.2.3 Medical reasoning as a process

In response to Elstein's description of clinical reasoning as one of hypothesis generation and testing, more recent research has tended to reconsider this account and instead view clinical reasoning as a process. This is illustrated by the work of Gale (Gale 1980, Gale and Marsden 1983) which provides a detailed description and understanding of the diagnostic thinking process in endocrinology and neurology and by the work of Patel (Patel and Frederiksen 1984) which focuses on clinical understanding of patient cases.

Aim of Gale's study

One aim of her research was to provide additional clarity and specificity to current descriptions of the diagnostic thinking process (such as that of the

Michigan and McMaster groups) in undergraduate medical education and clinical practice.

Subjects and methodology of Gale's study

Gale (1980) conducted her research using two methods: 1) a quantitative study using questionnaires in endocrinology and neurology and 2) a qualitative study using videotape-simulated recall of the clinical interview. The first study was used with subjects at the end of undergraduate medical education (i.e. final-year medical students) and after some years of clinical practice (i.e. medical registrars). The second study was used with final-year medical students, pre-registration house officers and medical registrars.

Model of diagnosis proposed by Gale

Gale rejected completely the hypothesis generation and testing process suggested by the Michigan group and proposed as a replacement a psychological framework within which the diagnostic thinking process could be viewed. This framework incorporates the cognitive processes of diagnostic thinking, its stages and development. She argued that the whole interpretive activity of the diagnostic thinking process can be explained in terms of two mutual and simultaneous cognitive processes of structuring and extrapolation. The clinician actively organises the array of clinical information which is being elicited by structuring it in some way. The clinician does so by referring to or extrapolating from the information already organised in her memory in certain structured ways. Gale (1980) identified three stages of the diagnostic thinking process referred to as *initiation, progress, and resolution.* She suggested that these stages occur in the thinking process of any clinician trying to resolve a diagnostic problem.

The first stage is concerned with how interpretations of clinical information are initiated. The second stage is concerned with what happens to the interpretations made during stage one, how they are cognitively manipulated while still current. This step is characterised by the cognitive operations of restructuring and assessment of interpretations by working from the extrapolated context. The last stage is concerned the final rejection or otherwise of the interpretations of the first two stages.

In addition to conceptualising the diagnostic thinking process in terms of psychological processes of structuring and extrapolation, and to identifying its dynamics in three stages, Gale also suggested some differences between students and registrars which may influence the development of the diagnostic thinking process. Some of the differences include: 1) interviewing and examining patients and making diagnoses in order to learn versus in order to treat and cure the patient; 2) recency of initial knowledge and skill acquisition.

Further, Gale indicated that students and physicians share the same general processes of diagnostic thinking. That is, the development of the diagnostic thinking process is not a matter primarily of qualitative changes but of quantitative changes. Both knowledge and thinking processes develop but they do so structurally or quantitatively rather than qualitatively.

Critique of Gale's study

The Michigan and McMaster groups like Gale postulated that experienced and less experienced clinicians share the same general processes of diagnostic thinking. However, Gale went a step further by reconsidering these processes in more depth and by comparing novice and expert clinicians. As Berner (1984) points out the process formulations of the Michigan and McMaster groups are a sequential set of steps applicable for all problems, whereas Gale's formulation is that there are various subprocesses from which to choose depending on the problem. That is, these three stages are not mutually exclusive. At any point during the diagnostic thinking process, the physician may simultaneously be at the stage of initiation of one interpretation, of progress of another and of resolution of a third one. The stages are only in consecutive order in the process of one interpretation. Since the physician may process more than one interpretation at the same time, this means that at any one time the physician may display the characteristics of one or two of the three stages.

In a more recent study, Grant and Marsden (1987) showed evidence of a consistent difference in the memory structures of novice and expert clinicians. They demonstrated that there was no difference between groups of differing clinical experience in the *breadth of thought* but there were differences in the *precise content and structure of thought.* To a certain extent one might say that the results of Gale support the work of Feltovitch et al who emphasise the role of knowledge in the clinical diagnostic process. However, while Feltovitch endorsed Elstein and Barrows's explanation of clinical reasoning, Gale excluded it. Although each group of researchers (Feltovitch et al, and Gale) proposed a different model of the reasoning process, they reached the same conclusion. That is, the form of the clinical reasoning is the same for medical students (novices) and doctors (experts), while the content is different.

Aim of Patel's study

Patel and Frederiksen (1984) focus directly on what it means to understand or comprehend a clinical case. This research has its root in the field of discourse processing which studies how people comprehend textual or spoken material. Comprehension is treated as a process of building an internal cognitive model of the message contained in the written or spoken material. Thus, comprehending a medical case is seen as an interactive process of constructing an interpretive frame (case model) for the case which reflects data and text properties and prior conceptual knowledge of the physician. The model is built under the guidance of grammars that direct and constrain the form the model can take. This work emphasises the case (or problem) model that physicians creates during comprehension, that is, emphasises how this model is assembled from data, knowledge and interactions among these two.

Subjects taking part in Patel's study

Subjects were 1st and 2nd year students and specialists in internal medicine.

Methodology of Patel's study

In a typical experiment, Patel and Frederiksen used some medical text material (e.g case descriptions or disease descriptions) and asked the subjects to read the text and recall or summarise what they have read. They also used probes and questions to help to determine the subjects' knowledge base. The analysis is directed at comparing the semantic structure of the medical text presented with the text produced in recall and/or summary. The medical text presented is formalised as a semantic network. Nodes are concepts (e.g infection, fever) and arcs are conceptual links among them (e.g produces). At a higher level of organisation are propositions which correspond to arcs and nodes together and can have a truth value (nodes + arcs = infection produces fever, is either true or false). Propositions can also be nodes linked together ((infection produces fever) and (fever may last for weeks)).

The text produced in recall and/or summary is analysed through techniques of discourse analysis. These include several steps (i.e. text level clausal analysis, propositional analysis, analysis of inferences a subject makes, and analysis of the conceptual frames the subject employs) whereby the text is decomposed into segments and hence the cognitive processes a subject uses in constructing a text model are identified.

Model of diagnosis proposed by Patel

Patel and Frederiksen proposed a model of the diagnostic process based on theories of cognitive processes in text comprehension. In this model, the diagnostic process is viewed as an interactive process of case comprehension. They do not entirely reject the hypothetico-deductive process suggested by the Michigan group, but rather suggest that this process may play a role in the interactive process of case understanding. During the interactive process, the physician constructs a case representation (through an inductive process) from patient data. The physician uses the case representation to access disease frames, narrowing the range of possibilities and thus generating hypotheses. Through a deductive process, the physician interprets the case representation, that is, the disease frames are tested against it. The case representation is modified and re-interpreted as the interaction develops and the diagnostic activity progresses. The interactive model of case comprehension is analysed using the techniques of discourse analysis mentioned earlier.

Results of Patel's study

Some of the important results suggests that experts are selective in their choice of components of the text, using a high density of causal and conditional relations. In contrast, intermediate students show non selectivity. Although this category of students have a reasonably extensive knowledge, however, it is not well enough consolidated to provide selectivity in constructing a mental model for medical material. In addition, intermediate students and experts make more inferences than beginners. This suggests that intermediate students can recognise the medical components of a patient case and augment them through inference in constructing a case model. However, they cannot yet distinguish the "essence of the patient case" (Patel and Frederiksen 1984).

Critique of Patel's study

The diagnostic reasoning presented as a case comprehension is that of the expert's. Patel and Frederiksen have shown that this approach using of an interactive process in which a representation of case information is constructed can also be applied with medical students. The main difference between the groups lie in the way interactive process of case comprehension is built and modified, that is, in the kinds of inferences generated by the physician. They have been able to demonstrate some differences between novices and experts in processing information in typical and atypical cases, and these differences reflect different processes for case comprehension.

Often, the physician does not diagnose a patient by analysing the medical text containing the patient's case. In most cases, the physician is in a consultation with the patient and hence deals with the patient directly and collects patient data verbally. Therefore, it seems that in producing a model of the diagnostic process, Patel and Frederiksen only took into account one aspect of the diagnostic task, that is, making a diagnosis via written medical material and excluded other factors to carry out a diagnostic task (such as the verbal collection of data).

3.2.4 Development of medical reasoning

The studies discussed in this section attempt to describe the development of medical reasoning: the work of Lesgold investigates the acquisition of medical reasoning in radiology; and the work of Ramsden, Whelan and Cooper examines some phenomena of medical students' diagnostic problem-solving.

Aim of Lesgold's study

Lesgold's work (1984) involved the study of medical diagnosis in radiology. It attempts to explain the way medical diagnosis develops and how the representation of the domain i.e. radiology may influence, and interact with the diagnosis.

Subjects taking part in Lesgold's study and methodology

Subjects were radiology students, from first year to fourth year and experts who were to interpret X-rays while thinking aloud as they were viewing the film.

Results of Lesgold's study

Results showed that the differences between novices (first and second year), intermediates (third and fourth year) and experts, lie in the ways a film is interpreted (i.e. from classicality to flexibility). Lesgold suggests that the process of development of medical skill from novice to expert is as follows:

- Novices learn a set of text-book medical conditions and rules of interpretation connecting film features to some interpretive models. Hence, they interpret films using direct, classic interpretation rules. They are limited to recognising gross visual properties of a film, and assigning to each local feature only one anatomical structure.

- As they view more films, intermediates learn that these simple, direct rules lead sometimes to errors because of, for instance, contextual factors, peculiarities of presentation in individual patients, or interdependencies in film features. They then try to understand the underlying basis of interpretive rules in the principles of anatomy and pathophysiology responsible for the appearance of a particular film. Even though intermediate residents had more knowledge than new residents, they were largely unsuccessful in correctly referring X-ray shadows to anatomic structures.

- Experts, like novices, use direct performance rules. However, these rules are directly structurally embedded. That is, contexual and deep issues are incorporated within the interpretation rules. Experts rely on a mental representation of specific anatomic structures to separate abnormalities from other structures in the area. The passage from intermediate to expert is achieved because the contexual considerations that the intermediate is focussed on, are compiled within efficient direct associational rules of interpretations.

Critique of Lesgold's study

The above explanation of the reasoning process in radiological diagnosis suggests that this field by its nature requires a different approach to diagnosis than other medical fields such as neurology for instance. Radiological diagnosis is a representational and perceptual skill. The radiologist needs to "see a patient" when looking at a film and not just complex visual features. Hence, she needs to have some sort of mental representation of that patient whereby she can make her diagnosis. One might say that there is a similarity between the radiologist and the neurologist since the neurologist also possesses a mental representation of the patient (in this case the patient's nervous system). However, the neurologist does not apply her mental representation directly into a visual picture, that is, the X-ray. Moreover, the environment in which the physician starts and pursues her diagnosis differs from the radiologist's. For instance, the physician sees the patient and asks her questions which correspond to initial cues in Elstein and Barrows models, while the radiologist will usually "see" the patient by viewing the X-ray film.

It should also be pointed out that the radiology tutor (reviewed in chapter two) for tutoring the skill of interpreting cardiac X-rays contains only knowledge of pathologies and anatomical features of the expert radiologist and not from other levels of expertise.

Aim of Ramsden's study

Some phenomena of medical students' reasoning processes

The study reported here takes a student's view rather than an expert's view of the medical diagnostic process by directly focussing on what medical students do. The work of Ramsden et al (1988) concentrates on one key aspect of the clinical reasoning process called problem synthesis. They study the differences in students' methods of handling the patient data. The purpose of that research is to consider whether the kinds of differences that have been discovered in phenomenographic research (concerning students' methods of handling science tasks and reading academic texts) are discernible in students' approaches to organising and explaining patient data. The aim of phenomenographic research (Marton and Saljo 1984) is to describe how students think about specific problems and phenomena from their viewpoint. Results of this research applied to subject areas such as physics, and engineering have suggested two relational phenomena (i.e. holistic and atomistic) which describe relations between learners and what they learn, and not student characteristics. In the holistic approach, students give their attention to the underlying structure and meaning of the task, attempting to derive its intended or implicit message. In the atomistic approach, students focus on the separate parts of the task, distorting its structure and failing to preserve the meaning of each part in relation to the whole.

Subjects taking part in Ramsden's study and methodology

Ramsden et al conducted an experiment with fourth-year medical students. Students were presented with a problem synthesis of patient cases in a summary form. Then, students were asked questions about the case they had just read. The interviews were audio-recorded and transcribed.

Results of Ramsden's study

Results showed two categories of description that portray the different ways in which the information in diagnostic problems is handled by students. These categories are structuring and ordering, and correspond respectively to the holistic and atomistic phenomena found in other subject areas.

Within the ordering approach, two strategies were observed. The first strategy is called *exclusion*. The student selects a clinical feature to be explored. She generates a diagnosis (or more than one) by association (e.g. 'this clinical feature fits with this symptom...'). Then, the student uses a ruling out procedure which involves discarding a diagnosis because of the absence of other associated clinical features. Finally, the student is left with one diagnosis, and tries to explain other symptoms and signs based on that diagnosis, thus forcing them to fit into the diagnosis. The second ordering strategy is called *pattern matching*. The student selects a diagnosis which is associated with one or more selected clinical features. The student ignores the rest of the clinical features, or simply believes that they fit the selected diagnosis. The selected diagnosis matches a disease already known to the student i.e. the pattern, and thus the diagnosis is accepted. The student does not propose other diagnoses and hence no method of ruling out of possibilities is used.

Within the structuring approach, two strategies were also observed. The first one is called *stepwise pathophysiological* strategy. It resembles the description of hypothesis generation and testing proposed by the Michigan group. The student moves from clinical features to diagnosis via a logical, sequential, pathophysiological explanation. The second structuring strategy is called *diagnostic integration*. The student moves from clinical feature to diagnosis and in the reverse direction. As a hypothesis is formulated, it is based on pathophysiological mechanisms that explain the symptoms and signs. In both structuring strategies, the ruling out of diagnoses is used. However, the ruling out is made with an emphasis on information which supports the diagnosis as well as discussion of why the absence of certain information favoured the exclusion of other diagnoses.

Critique of Ramsden's study

Ramsden et al have suggested a number of strategies that students used. However, they do not propose any interactions between these strategies. Moreover, some interpretations and conclusions reached in this study need further clarification. For instance, Ramsden et al argued that the structuring approach contains elements that most clinical teachers regard as desirable, and is closer to what they expect of medical students. In contrast, the ordering approach is less satisfactory. These findings, of course, have important implications for improving medical instruction and assessment. However, there is no evidence to support the claim that expert physicians used a structured approach in their clinical reasoning. In addition, the difference between the pattern matching and the stepwise pathophysiological strategies is not so clear. When examining the protocols of students' interview about patient cases, one can see that both strategies involve references to pathophysiology.

3.2.5 Discussion of the psychological models

Several conclusions can be drawn from the review of models of medical problem solving.

- Firstly, there is a wide variety of research, each using different specialities of medicine, subjects and methodologies to study medical problem solving.

Medical domains

Clinical reasoning has been investigated in the sub-medical fields of pediatric cardiology, radiology, neurology, endocrinology.

Subjects

Subjects of these studies have different levels of training and experience. In particular, three categories of subjects have been used: the expert, the intermediate and the novice. Ideally, the first category should include physicians, specialists or registrars with several years of medical practice; the second category should include residents, house officers; the last category should have medical students (first to fifth year). However, not all the studies follow this categorisation. For example, in Lesgold's research third and fourth year medical students are considered intermediate, whereas Feltovitch et al considered residents as intermediate subjects. In addition, other studies such as Feltovich's are not clear about which kind of student has been considered (first, second year, etc...).

Methodologies

Similarly to the medical domains and subjects, psychological methodologies are also diverse. These are verbal and written protocols. Verbal protocols incorporate think-aloud used by Elstein et al, Feltovitch et al and Lesgold; retrospection (e.g videotape-simulated recall of the clinical interview) by Elstein et al, Barrows and Tamblyn, and Gale; probing by Patel and Frederisken and Ramsden et al. Written protocols include written questionnaires used by Gale; a "read text and summarise in writing" method by Patel and Frederiksen.

Although some researchers may share the same methodology for their studies, some variations occur in the use of that methodology. For example, in the recall of the interview between doctor and patient, Elstein et al used simulated patients whereas Gale used real patients. In Barrows and Tamblyn's experiments, videotape recalls reported interviews between medical students and simulated patients. In addition, some researchers are sometimes unclear in some aspects of the method used. For instance, Elstein et al's experiments (1978) do not specify whether recall of interviews are from the doctor's long term memory (LTM) or short term memory (STM). The long length of each interview suggests that recalls are from LTM. Likewise, in Patel and Frederiksen's study (1984) it is unclear whether probing of the subject occurs before or after reading and summarising the text.

The validity and reliability of the psychological studies reviewed in this chapter and hence of the models of clinical reasoning proposed depend partly on the methodologies used. These methodologies have given rise to objections and doubts. One objection against using verbal protocols as data has been that the reporting process might alter task performance (e.g. reaching the correct medical diagnosis). That is, the effect of verbalisation may change the performance of the subject and thus of the cognitive processes. Another objection is that reports may yield an incomplete record of cognitive processes. For instance, when asked to think-aloud while performing a task, the subject (e.g. experienced doctor or medical student) may fail to verbalise a considerable part of information that passed through short term memory

(STM) and was used in the task performance. As a result, part of the reasoning process may not be identified, and hence the model of cognitive processes proposed may be too general. Indeed, the model of clinical reasoning proposed by Elstein et al is of a generic form.

In retrospect, the retrieval of information from LTM causes problems since the other memory structures may be accessed instead of those created by the just-completed cognitive activity. In contrast, retrieval from STM is more desirable since the subject still holds in her STM information about the completed task. This last form of retrospective verbal report is used in Gale's study (1980). Probing a subject may also produce problems. The subject may be questioned for information that she does not have directly accessible and thus this forces the subject to produce verbal reports that are not close to the actual thought process. For instance, Patel and Frederiksen probed subjects directly to determine their knowledge base. They were interested in obtaining a certain kind of information (e.g. medical facts). On the other hand, Ramsden et al used another form of probing. The medical student was first asked to report what she has learned about the patient case. Then, followed a series of non-directive questions designed to elicit information about the case such as a diagnosis or a set of diagnoses and the reasons for choosing them.

In addition to verbal protocols, written protocols also have limitations that may alter the analysis of cognitive processes. For example, questionnaires (multiple choice questions) are a good means of finding out how much a subject knows. However, the questions can only be presented to the subject if the investigator knows what the right answer is (Welbank 1983). Each questionnaire used in Gale's study includes multiple-choice questions (i.e.

mastery of factual knowledge, interpretation of symptoms and signs, selecting and testing diagnostic possibilities), and a formulation of the diagnosis written by the subject. In contrast, the reading of a technical expository text (e.g. medical text) is a more demanding task in terms of the cognitive processes involved. The analysis of a text (e.g Patel's experiment) can produce difficulties. A common difficulty is ambiguity when analysing propositions of a text (Britton and Black 1985). One type of ambiguity is when a word can have different functions. The word can be either general (e.g the word "when") or specific (e.g. medical term). Another type of ambiguity is when the text itself is ambiguous in meaning. In that case, the recall by the subject may be ambiguous. Moreover, the subject herself may be careless about the use of words and thus create ambiguity.

An additional problem in analysing text is the scoring of recall protocols. That is, a proposition is either recalled or not; no partial credit is given. Subjects rarely produce protocols that are word for word exactly like the original text. The problem for the investigator is to decide how to give credit for a recall. Patel and Frederiksen used strict scoring in their experiments, that is, credit is given for a proposition only when it is closely reproduced in the protocol. In medicine, knowledge is an important factor. Hence, when analysing recall protocols, the investigator should also take the above factor into account. Experienced physicians are high-knowledge individuals. Consequently they tend to use more general and abstract statements and may also add more inferences and details. The performance of a subject does not depend only on the knowledge she has, but also on the text presented. The choice of the text is an important factor which may alter the results obtained.

- The second conclusion that one can draw from this discussion is that there is a lack of uniformity in terms of methodologies as well as subjects and sub-medical domains. This implies that there is no existing 'formal framework' to which medical researchers agree to work. This also means that there is no 'formal paradigm' of the medical problem-solving. Most of these researchers have investigated clinical reasoning from different orientations. While the Michigan and McMaster groups have studied the clinical reasoning process from an information processing perspective, other researchers examined that process from other standpoints such as problem-solving (e.g Feltovitch et al and Lesgold) and propositional analysis (e.g. Patel and Frederiksen).

- The third conclusion from the review of the psychological models is that the clinical reasoning has been studied from an *expert perspective.* All the studies, (except for the work of Ramsden et al and to a certain extent the work of Lesgold), have investigated what the expert physician does, and compared it with less experienced subjects such as medical students. Therefore, it is not surprising that the models of clinical reasoning which have been suggested are solely those of the expert. Laurillard (1989) has come to a similar conclusion about the expert based approach adopted in the studies, adding that there is a methodological problem related to it: most of the studies begin with a theoretical framework into which empirical analyses are fitted. Hence experts and students have been studied in the light of pre-existing descriptions of their performance. In neglecting the way medical students reason and formulate diagnoses, medical problem-solving researchers have assumed that it is enough to know what the expert physician does, and then try to duplicate it to the students. However it seems, as Gale and Marsden (1983) have already pointed out, that they have omitted the important fact that when medical students enter medical schools they already have a certain way of thinking and solving problems which has to be taken into consideration when teaching them medicine.

- The fourth conclusion is that despite the diversity of studies on clinical reasoning, there has been an evolution in the research area of medical problem solving. Studies have tended to emphasise primarily the form of clinical reasoning, and then its contents, in both cases from the expert standpoint. This evolution is shown in figure 3.1.

Numerical models, and psychological models are marked on the diagram. Within the psychological studies, the models of Elstein et al and of Barrows and Tamblyn are placed first since they only looked at the general form of the clinical reasoning. The work of Feltovitch et al succeeds the classical models since it focuses on the contents of the clinical reasoning i.e. knowledge base. The work of Patel and Frederiksen is put at the same level as Feltovitch et al since their work takes into account the importance of the contents of the reasoning process. The work of Lesgold is placed afterwards. Even though the studies of Lesgold are very domain specific, the process of the development of the medical skill which he proposed demonstrates that the form and the contents of the clinical reasoning (in radiology) are closely tied. Last is the work of Gale. She provided a new description of the clinical reasoning process with greater complexity and refinement of the initial theoretical formulation (i.e. Elstein et al). Intentionally, Ramsden et al's work has been omitted in this figure because it only takes a 'student perspective'.

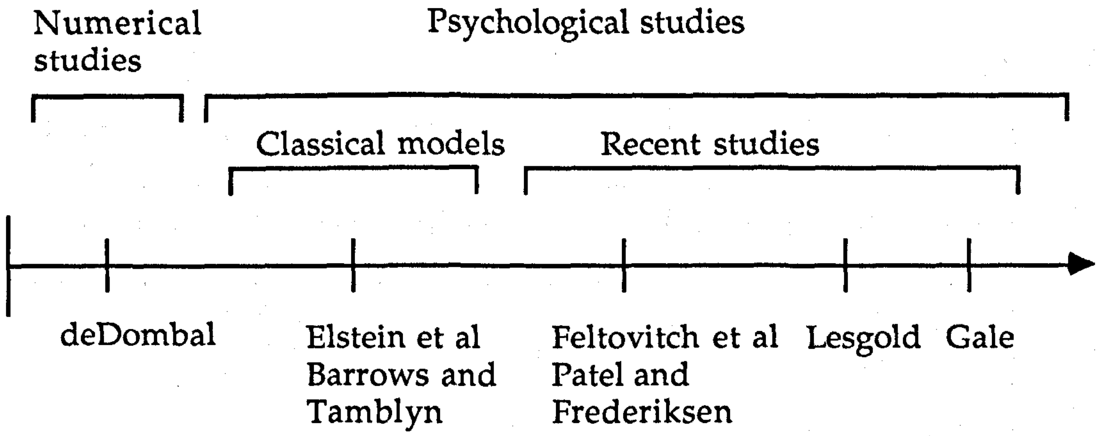


Figure 3.1: The evolution of the clinical reasoning process (at the expert level) from a focus on the form to a focus on the contents

- The fifth conclusion which combines the two previous points is that differences between novices and expert physicians have been found to be more on the content rather than on the form of medical problem solving. Table 3.1 shows medical problem solving of the student as suggested by the different studies. One can see that the content is usually different between novices and experts while the form is not.

In the context of this discussion and in the view of designing a model of medical reasoning that takes into account the student's reasoning, it seems important to mention whether any model of physician's clinical reasoning in particular any model of student's clinical reasoning reported in the medical problem solving literature has been implemented into a computer system. In fact, it was found that few attempts have been made to implement computer program models of the medical diagnostic process reported in this chapter. Following are some examples of medical AI systems which embody some of the theories of medical reasoning.

Clinical Reasoning Process of the Student

Studies	Form	Contents
1. Elstein et al	Same as the expert's : Hypothesis generation and testing	Different from the expert's
2. Barrows et al	Same as the expert's: Variant of hypothesis generation and testing withing problem-based learning applied to medical education	Different from the expert's
3. Feltovitch et al	Same as the expert's: Hypothesis generation and testing	Different from the expert's Sparse, classical and imprecise
4. Gale	Psychological framework - 2 cognitive processes (structuring & extrapolation) 3 stages (initiation, progress resolution)	Evidence that memory structures are organized differently between experts and students
5. Patel	Same as the expert's Interactive process of the case understanding	Different from the expert's
6. Lesgold	Intermediate different from the expert's. Novice and expert similar Model of the development of diagnostic skill - way rules of interpretation are used	Different from the expert's
7. Ramsden et al	Different from the expert's: ordering strategies Similar to the expert's: structuring strategies	Different from the expert's

Table 3.1: Different kinds of clinical reasoning process used by medical students in terms of the form and the contents

- PIP (Pauker and Szolovits 1984) simulates the behaviour of an expert nephrologist in taking the history of the present illness of a patient underlying renal disease. The clinical cognition theory that PIP embodied is, to a certain extent, based on the model of clinical reasoning proposed by the Michigan group. For instance, Elstein et al (1978) suggested that the expert physician needs very little initial information (early case cues) to have an initial guess on the nature of the patient's problem and thus to generate a set of hypotheses. Likewise, in PIP the early generation of hypotheses is done by the triggering process. Moreover, according to Elstein et al (1978), the hypotheses generated serve to explore the diagnostic process further and ask for more data. Similarly, in PIP, the current working hypotheses provide the basis for asking for additional information which is then evaluated with respect to these hypotheses.

- DIAGNOSER (Johnson et al 1981) is a computer simulation program that represents the knowledge required to diagnose patients suspected of congenital heart diseases. This computer simulation model was developed to test theoretical work on the expertise in medical diagnosis in terms of the organisation and manipulation of knowledge. The way the knowledge base in DIAGNOSER is built reflects Feltovitch's findings on the characteristics of the knowledge base. For instance, recall that the knowledge base of the expert was identified as being precise, meaning that experts have more precise feature expectations within disease frames which help them to discriminate among diseases. The knowledge base in DIAGNOSER can be said to be precise as well because the type of knowledge i.e. the disease knowledge, is realised as a hierarchy of disease schemata where each schema consists of a structure of expectations for the patient data of that disease, as well as a structure which explains the expectations for the disease in terms of underlying pathophysiology. Moreover, the reasoning process in DIAGNOSER is similar to the hypothesis generation and testing model proposed by the Michigan group.

- NEOMYCIN (discussed in chapter two) is a reconfiguration of a rule-based expert system MYCIN for application to teaching. NEOMYCIN incorporates a theoretical model of medical diagnosis and an epistemological theory of knowledge. The methodology that was followed to design NEOMYCIN was (apart from interviews and protocol analysis) to review extensively the medical problem solving literature (Clancey 1984) including work of Elstein et al (1978), Feltovitch et al (1984a, 1984b), Johnson et al (1981) and Pauker and Szolovits (1984). Thus, the theory of medical diagnosis developed by Clancey was influenced by studies on medical problem solving. For instance, one of the key features of NEOMYCIN is that as patient data is received, hypotheses are generated and placed on the differential. This feature is reminicent of the hypothesis generation phase in the model of clinical reasoning proposed by Elstein et al. Likewise, the importance of the knowledge as suggested by Feltovitch et al is recaptured in this epistemological theory of knowledge which made explicit the structural, strategic and support knowledge.

The medical systems mentioned above incorporate psychological models of the clinical reasoning mostly based upon the work of Elstein et al, and Feltovitch et al. One might see an exception in NEOMYCIN since its design has not only been influenced by these psychological models but also by the other A.I systems for clinical reasoning. However, no attempt has been made to incorporate into an A.I. system other theoretical work on clinical reasoning such as of Lesgold, Gale, Patel and Frederiksen, and Ramsden et al. In particular, this means that no attempt has been made to implement students' clinical reasoning as reported in the work of Ramsden et al and of Lesgold. The medical systems mentioned above do not take into account a view of the medical student.

3.3 Teaching of medical diagnosis

In this section, teaching of medical diagnosis, in particular of any of the models of clinical reasoning proposed in section 3.2, is examined. Problems that medical students face in their learning of medical diagnosis, and various teaching methods to help them are also reported and discussed. This review is by no means exhaustive; the concern has been to provide a discussion of teaching of medical diagnosis in regards to the models of clinical diagnosis proposed in the literature and to the students' needs in their learning of medical diagnosis.

Some problems in medical education

There are a number of problems in medical education which have implications for the students' learning. Balla et al (1989) reported three of these problems:

1) The split between the pre-clinical and the clinical. This is a problem of lack of integration of knowledge. In the preclinical years, the student does not have an overview of her subject matter and has not been able yet to put it all together. The issue for medical education is to assist the student to integrate her knowledge, teach her when to rely on the preclinical and when to turn to other sources of knowledge.

2) Coping with uncertainty: students need to learn to understand subject matters in depth so as to understand the impact of uncertainty on the process of clinical reasoning. How to deal with uncertainty is not well understood by students and is rarely taught. Students learn routine protocols to follow to carry out a diagnosis. This is a surface approach and reflects a lack of deep understanding of the underlying processes of diagnosis. The issue here as Balla points out is to have the teaching faculty reflect on and understand their diagnostic processes so they will then be able to impart their understanding to their students. One can also add that students should also be able to reflect on their own processes.

3) Availability of factual information: this is a problem of learning how to evaluate data. Students have to learn a vast amount of medical knowledge. The issue here is to provide students with easy availability to large data bases, computerised literature searches and other such features that would help them with the overload of factual knowledge.

Buchanan (in a personal communication 1990) stressed similar problems of medical education and the needs students have in their learning process. Specifically, students need to strengthen their medical knowledge (related to problem no.3 mentioned by Balla) and students also need to get feedback on their reasoning processes (related to problem no.2). Irby (1986) pointed out additional problems in medical education with particular reference to clinical teaching (thus excluding preclinical teaching). These include:

1) Limited emphasis upon problem solving: overwhelming work load place demands upon students which leaves little time for thinking and reflecting.

Students rarely have an opportunity to reflect on their learning, make connections to basic science information (also mentioned by Balla as problem no.1), restructure the knowledge they already have and engage in real problem solving on patients under their care. Furthermore very few questions required students to discuss their reasoning.

2) Lack of clear expectations for students' performance and inadequate feedback to students: as a result students encounter differing and sometimes conflicting expectations for their behaviour.
As Irby (1986) points out students complain about the lack of feedback on their learning and performance. The problem of inadequate feedback to students has also been indicated by Laurillard (1989). She argued for an approach to teaching that highlights the need for students to reflect on the problem solving process itself.

3) Inappropriate role models and clinical settings.
The role models and clinical settings to which students are exposed are not always appropriate for the general professional education of the physician. For example, by failing to attend to psychosocial needs of patients, faculty members and residents do not show examplary models to their students.

Teaching of the models of clinical reasoning

In examining the current state of teaching medical diagnosis, one notices that there is not always a clear distinction between the approaches to teaching and the models of clinical reasoning proposed in the literature. Let us take for example the model of clinical reasoning proposed by Elstein et al. Hypothesis generation and testing is taken as a cognitive model of clinical reasoning, but is usually taught as a method in which medical students are taught to reason in a hypothesis guided mode. This approach has been widely accepted and used in medical schools. Although it has been strongly criticised (see section 3.2.1), the hypothetico-deductive approach represents one attempt to provide an adequate structure to the problem-solving situation, and can be taught in a way that students can easily comprehend.

The model of medical reasoning suggested by Barrows and Tamblyn has been used within the context of problem-based learning as a teaching method. The approach has also been widely recognised (see for example Schmidt 1983) and successfully used. One recent study (Newble and Clarke 1986) showed that the performance of medical students attending a problem-based medical school was higher than students from a traditional medical school.

Teaching methods

A number of teaching methods other than problem-based learning and hypothesis generation have been used in medical schools. These include integrated curriculum and cognitive skills training. An integrated curriculum includes a collection of courses e.g. biochemistry, which are independent from each other. However, as Gale and Mardsen (1983) point out this kind of teaching does not intend to treat the diagnostic thinking process per se, but rather to structure knowledge so as to facilitate its retrieval in the diagnosis process. The aim of cognitive skills training has been to train students in general problem-solving skills and awareness of their own cognitive processes, with the expectation that the results of such training would generalise to their clinical task. For instance, Culter (1985) in his book *Problem solving in clinical medicine* provides students with a tool to learn about problem solving and how to bridge the gap between basic sciences and the bedside, between pathophysiology and the patient, between knowledge and the application of knowledge, and finally between the collection of data and their synthesis into defined problems. The book can be used either by the student alone with a self evaluation section or as part of the curriculum.

New approaches to teaching have been formulated but not applied extensively in medical schools. For instance, Gale (1980) has proposed a new framework for teaching the diagnostic thinking process. First, she put forward three basic principles (structure of learning, transfer of learning, problem-solving and learning) which should be applied to any teaching or learning strategy of clinical reasoning. In relation to these principles, teaching strategies are defined. These include a teaching strategy for structure and a teaching strategy for process. The first teaching strategy involves learning the structural characteristics of stored knowledge as well as enhancing the development and use of such knowledge, whereas the second teaching strategy is to facilitate analysis of cognitive skills. Gale concluded that these teaching strategies both for structure and process are central to any appropriate pedagogy of the diagnostic thinking process. Recently some of these ideas put forward by Gale have been incorporated into the curriculum of a course whose aim was to improve students' understanding of their own diagnostic thinking process (Gale and Marsden 1986). Evaluation of the course curriculum was successful with very few students (2%) who found that they did not achieve this aim.

Some medical educationalists have carried out experiments to try to change the curriculum and the presentation of textbooks used in medical schools. For instance, Hewson (1986) indicates that there is a need to deal explicitly with the topic of appropriate knowledge organisation in medical training. In this study, a teaching intervention consisting of a 6-week seminar was used with sixth-year medical students in pediatric cardiology. The objective was to determine whether the reorganisation of knowledge structures acquired from pre-clinical lectures and from textbooks can be mediated by medical school instruction. Students were asked to diagnose patient cases by reporting aloud any thoughts they had in formulating the diagnosis. They were instructed to think about medicine as experts do, that is, to conceptualise knowledge as consisting of parts that are related in a variety of ways, and which can be regarded as chunks. It was concluded that the teaching intervention introduced through this pilot study showed an improvement in the students' thinking.

Another study (Norman 1984) suggests curriculum innovations which could provide an environment rich in clinical experience and the learning of concepts in the context of clinical problems. This direction of curriculum changes as proposed by Norman showed that the experienced clinician is a better problem-solver by virtue of her accumulated experience and not as a result of any innate or learned problem-solving skills.

While Hewson and Norman tried to change aspects of medical education, Balla (1988) argued for a total change in the way clinical education is viewed. Two main problems in medical education are a) an insufficient emphasis on teaching (assuming that it comes naturally to the physician), and b) an insufficient emphasis on the process of learning. These problems suggest that there is a need to change the perception of the practice of clinical medicine, and hence of clinical education. Balla favoured a scientific practice of clinical medicine as well as scientific teaching methods relevant to this practice.

In Norman's study (1984), the use of concept learning was mentioned. Other studies have examined further the use of concept learning in a clinical context. For example, Friedlander and Gillespie (1987) report how students at all levels learn more effectively using concept mapping. Concept maps show how ideas in each lesson relate to one another. Ideas grow in significance as they become connected to wider arrays of concepts. For example, one might connect *pulmonary hypertension* to *high pulmonary arterial resistance* with the link *may be caused by.* Bordage (1987) has also reported on the importance of concepts, and of a prototype view of categorisations of medical disorders. Categories are better learned when the initial exposure is through representative example i.e. the prototype, as opposed to the whole range of instances. Secondly concepts are initially learned at an intermediate level of abstraction (e.g. angina pectoris) corresponding to the prototype as opposed to more general levels (e.g. coronary disorders). This means that to encourage the formation of prototypes in the memory of the students, the initial exposure should be limited to the most representative examples and should be based on intermediate level concepts which will act as a reference point for future learning.

Computer tools for teaching

Along with teaching methods, electronic aids have been also used. For instance, the development of computers has seen the integration of computer programs in teaching medicine. The utilisation of computers in teaching in general and in medicine in particular has evolved greatly. Some of this includes computer aided instruction systems, and more recently intelligent tutoring systems which use artificial intelligence techniques (as was discussed in chapter two).

3.3.1 Discussion of teaching of medical diagnosis

Several conclusions can be drawn from this review of teaching of medical diagnosis:

- Firstly, the literature does not provide evidence that the models of clinical reasoning suggested in the medical problem solving literature are taught explicitly in medical schools, unless these models are used as teaching methods as in the case of hypothesis generation and testing. Models of clinical reasoning such as Patel and Frederiksen have stayed at the research stage. They are neither taught nor applied as teaching methods in medical schools.

- Secondly, medical students learn models of clinical reasoning (of the expert) in an implicit way. For example, a student who attempts to make a diagnosis receives comments from the teacher but is not told explicitly any model that she should follow. Rather, from these comments and questions to the teacher, the student is to build her own model. This does not mean that the student is aware of the model that she is using. Likewise, the teacher is most probably not aware of the model(s) used by the student. This issue is also related to the problem mentioned by Irby of role models that physicians play for the students.

- Thirdly, teaching methods have been usually more oriented towards helping the students in the learning of factual knowledge rather than providing feedback on their reasoning processes. Means to help learning of factual knowledge are found:

i) In the method of integrated curriculum that helps students to structure the knowledge they learn.
ii) In new approaches such as learning the structural characteristics of stored knowledge (e.g. Gale 1980), learning to reorganise knowledge structures acquired from pre-clinical lectures and from textbooks (e.g. Hewson 1986), and learning to form prototypes (e.g. Bordage 1987).
iii) In computer tools such as QMR (reported in chapter two).

In contrast there are fewer methods to help students get feedback on their reasoning processes. These include the hypothesis generation and testing approach (Elstein at al 1978) which is in any case very generic and the teaching approach suggested by Gale (1986), which is to facilitate the analysis of cognitive skills.

3.4 Summary

This chapter has reviewed medical problem solving and the teaching of medical diagnosis. The aim of the review was to provide a discussion on the features of a student model that would be based on the student's medical reasoning. The main conclusions of this review which have implications for the design of such models can be summed up as follows.

- From the review of medical problem solving, it was found that:

i) There is no formal paradigm of medical problem solving within which researchers agree to work, and hence no consensus exists on the exact nature of medical reasoning.

ii) Models of medical reasoning have been mainly expert based (except for the work of Lesgold and Ramsden et al). That is, the models have been constructed from the expert physician's behaviour and then compared with less experienced physicians. Furthermore, no formal model of students' reasoning has been proposed.

iii) The clinical reasoning has been decomposed in terms of its contents (i.e. the knowledge used) and its form (i.e. the reasoning itself that support its contents) with a stronger emphasis on the content than on the form of medical diagnostic process.

iv) Differences between novice medical students and experienced physicians were found to be on the contents of the diagnostic process rather than on its form.

- From the review of the teaching of medical problem solving, it was found:

i) Medical students learn models of clinical reasoning (of the expert) in an implicit way.

ii) There are a number of problems in medical education which have implications for the students' learning. Teaching methods have tended to help students on the learning of factual knowledge rather than providing them with feedback on their reasoning processes.

These findings have implications for the features of a student model to be based on students' reasoning. In particular, it is suggested that student modelling for medical tutors should focus:

i) on the form of medical reasoning, that is, on reasoning strategies applied during medical diagnostic process

and

ii) on the need students have for self reflection and feedback on their own reasoning processes.

One should aim to provide a model which would consider students' reasoning strategies. This model could be integrated in a tutoring environment that would help the student examine what kind of reasoning processes she has been using. The form of the medical diagnostic process constitutes one of the features of the student model which is to be designed.

The review in this chapter has also showed that students' reasoning has been examined from a developmental perspective (with the work of Lesgold and Ramsden et al). Since the research work aims to design a model that takes into account students' reasoning, this approach seems worth pursuing. Hence, the development of medical reasoning constitutes a second feature of the student model. This approach implies modelling different phases of reasoning a student goes through - from novice to intermediate to more experienced physician and hence viewing the changes in student's reasoning over time as a developmental process rather than a static one and subset of the expert. Developmental models are not new, and the next chapter examines their application to medicine and in a tutoring context, before discussing the design of a developmental student model for medical tutors in chapter five.

Chapter Four

DEVELOPMENT OF EXPERTISE

The review of the previous chapter has shown that students' medical reasoning has been examined from a developmental perspective. This approach is pursued in the thesis because the objective of this thesis is to design a student model for a medical tutor that would model the development of medical diagnostic skill. The purpose of this chapter is to investigate what kinds of developmental models for tutoring have been built in the past, particularly developmental models for medical problem solving. The first section gives a general overview of the research about expertise. Some developmental models are then reviewed. The application and importance of development of expertise for tutoring is discussed in a further section. Lastly, some conclusions from this review are drawn.

4.1 Development of expertise

Research in the area of expertise has been studied in terms of three directions: expert behaviour, novice behaviour and the differences between experts and novices, in various domains such as physics (e.g. Larkin et al 1980) and programming (e.g. Jeffries 1982). The research on expertise makes it clear that experts and novices differ in fundamental ways. These differences extend to a variety of behavioural responses such as problem solving performances, perception, preferences and social attitudes (Leventhal and Instone 1982). However, as Leventhal and Instone point out the rich body of literature in expertise has tended to focus on the characteristics of experts and novices and relatively little work has emphasised the process of becoming an expert. It is that issue of how one becomes expert and the processes one goes through to becoming expert, that is relevant to the design of a student model to be based on the development of medical expertise.

It should be pointed out that in reviewing the area of expertise, it was found that researchers use various terms such as acquisition of expertise, acquisition of skill, or development of expertise. These terms are not always consistently defined across various studies. The following gives the meaning of these key words as used in the thesis. *Acquisition of skill* is taken to refer to the process by which a person acquires a skill (e.g. a concept), that is, comes in contact with the skill to be learned and internalises it. A skill may be one of many components that makes up expertise. By contrast, *the development of expertise* refers to the process by which a person from being a novice becomes an expert. The concept of development therefore has associated with it the notion of qualitative change over time. The research reported in the thesis focuses on the development of medical reasoning strategies, rather than on the acquisition of these strategies. However, when reviewing a research work, the terminology used by the cited researcher is kept unchanged.

4.2 Domain specific developmental models

Domain specific developmental models refer to models that deal with the development (or acquisition) of expertise of a particular domain. By contrast, generic developmental models (see section 4.2) are concerned with the development (or acquisition) of expertise of any cognitive skill. This section examines developmental models in the areas of physics, radiology and logic and probability. Three main models that are found in the literature, are presented using the following points: 1) scope of the research, 2) the domain of application and 3) the proposed developmental model.

Acquisition of physics expertise

White and Frederiksen (1986, 1987) have researched modelling the possible evolution of students' reasoning about electrical circuits as they come to understand more about circuit behaviours. They have proposed representing the evolution from novice to expert via causal model progressions. The transition from novice to expert is thus regarded as a process of model evolution. Novices need to evolve not just a single model but a set of models that embody alternative conceptualisations of the domains. One important contribution to the field of expertise is that this work emphasises the importance of qualitative models in the acquisition of expertise.

Acquisition of perceptual diagnostic skill in radiology

The process of acquisition of medical skill in radiology was described in section 3.2.4. Lesgold (1981, 1984) found that the differences that exist between novices, intermediates and expert radiologists lie in the rules of the interpretations of the films. One aspect of Lesgold's work has been to stress the importance of organised knowledge in acquiring expertise.

The genetic graph

The aim of this research (Goldstein 1982) has been to construct a model for representing the development of procedural knowledge from an evolutionary viewpoint. The domain of application is a maze exploration game called WUMPUS (Yob 1975) in which children can exercise basic skills in logic and probability. Goldstein proposed a genetic graph to represent the evolution of the learner's procedural knowledge. The genetic graph consists of nodes which represent procedural rules and links between nodes which represent various evolutionary relationships. These evolutionary links include processes such as generalisation, specialisation, analogy and refinement. Goldstein used an overlay approach to the genetic graph in which the student's knowledge is described in terms of the nodes of the graph and her progress in terms of the paths of the graph.

4.2.1 Discussion of domain specific developmental models

Issues such as the domain of application, representations of the developmental process and implementations of that process are now discussed.

Domains of application

The studies reported in chapter three on medical problem solving have been oriented towards differences between novice and expert physicians. The previous section shows that limited research has been carried out on the acquisition of medical diagnostic skill. The combination of these two findings support Leventhal and Instone's view of the current state of research on expertise, that is, to focus on novice/expert differences rather than on the process of becoming expert (see section 4.1). The work of Lesgold (1981, 1984) in the medical domain is itself limited to the acquisition of *perceptual* skill in medicine. But not all domains of medicine involve the use of visual information to drive diagnostic decisions. Rather, the work of Lesgold has implications for understanding the acquisition of expertise in other domains (e.g. equipment maintenance) where large amounts of visually available information are part of the process of diagnosis and treatment. In addition, in contrast to the medical domain, the other domains in which developmental models have been constructed include defined and structured, such as the domain physics and the toy blocks world.

Representing the acquisition of expertise

Several methods for representing the acquisition of expertise have been used such as causal model progressions and genetic graph. The causal model progressions correspond to explicit stages of the acquisition of expertise. White and Frederiksen (1986, 1987) show that expertise does not consist of a single model that represents a deep understanding of the domain. Rather, expertise is characterised by the coexistence of a set of complementary models that vary given the level of expertise. Moreover, the construction of the causal model progressions demonstrates that one can integrate models of various types into a flexible understanding of the domain. Each model contains information about the structure of the circuit, that is, the devices (e.g. switches) and their interconnections. The causal explanation contains information on changes of the device states that occur during an operation of the circuit and the reasons for those changes. Causal explanation of a model therefore corresponds to how to reason about the underlying principles of circuit operations.

Goldstein's evolutionary links are reasoning processes that are used to represent how new knowledge evolves from old one. Hence, the genetic graph makes explicit a number of strategies such as generalisation or refinement for reasoning over the rules. More recently, Bretch and Jones (1988) have expanded the definition of the genetic graph and tested it with other domains such as subtraction and ballet. Nodes which describe skills have been extended so that a skill can consist of components. Additional links such as a component link and a correction link have been incorporated for the modelling of the two other domains. These extensions of the genetic graph have shown that new relationships can be easily incorporated into the genetic graph to cover additional and more complex situations.

Lesgold's work (1981,1984), similarly to Goldstein's (1982), is focussed on procedural knowledge and demonstrated the role of proceduralisation in the acquisition of medical skill. The rules of interpretation of films from being context-free become more and more complex and compiled, containing contextual factors. Lesgold also illustrates the role of organised knowledge for diagnostic reading of X-rays films. Expert radiologists acquired organised bodies of knowledge that constitute radiological anatomy, and relationships between variations in anatomical structure and patterns seen in X-ray plates. Rather than viewing the development of procedural knowledge via reasoning processes like in the genetic graph, the development of procedural knowledge is achieved by restructuring knowledge that the learner has acquired.

Implementations of domain specific developmental models

Most of the developmental models have been implemented. Causal model progressions have been implemented into the system QUEST (White and Frederiksen 1987) an intelligent learning environment which will be discussed in section 4.4. The genetic graph was partially implemented and was tested only with simple game playing situation (Bretch and Jones 1988). The exception is the model of acquisition of diagnostic skill in radiology which has not been implemented (as already mentioned in chapter three).

In summary, one can conclude that research on the development of medical expertise has been limited but also it has stayed at a descriptive level. Moreover, research on the development of medical expertise has not focussed on the development of medical reasoning strategies but rather on the importance of organising knowledge. However, developmental models in other domains have explicitly made use of reasoning processes such as generalisation and specialisation.

4.3 Generic developmental models

This section reports on the ACT* theory, a five-steps stage model proposed by Dreyfus and Dreyfus and work on machine learning.

ACT theory*

The aim of this research (Anderson 1982) was to present a theory about the changes in the nature of a cognitive skill over a period of time and about the basic learning processes that are responsible for it. The theory includes two major stages in acquisition of a cognitive skill. The first stage is declarative in which facts about a domain are interpreted while the second stage is procedural in which domain knowledge is directly embodied into procedures for performing the skill. The evolution from the declarative stage to the procedural is modelled by the process of knowledge compilation where new productions are created. Once the skill has been proceduralised, further learning processes (e.g. generalisation, discrimination and strengthening) operate on the skill to make the productions more selective in their applications.

A five stages model of skill acquisition

One of the aims of this work (Dreyfus and Dreyfus 1986) has been to understand human skills and what goes into becoming a human expert. Dreyfus and Dreyfus have studied the skill acquisition process of airplane pilots, chess players, automobile drivers, and adult learners of a second language. They have observed a common pattern in all cases which lead to a five stages of skill acquisition. The model was then compared with the acquisition of nursing skill. One interesting aspect of this model of skill acquisition is that it encompasses skill acquisition of problem areas which Dreyfus and Dreyfus refer to as unstructured. Such areas contain a potentially unlimited number of possible relevant facts and features, while the ways those elements interrelate and determine other events is not always clear.

Dreyfus and Dreyfus have suggested that there are five steps that one goes through to become expert: (1) The novice learns to recognise various objective facts and features relevant to the skill and acquires rules for determining actions based upon those facts and features, without reference to the overall situation in which they occur. (2) Through practical experience in concrete situations, the advanced beginner starts to recognise the situations in which meaningful elements are present. (3) The competent performer can choose a plan to organise the situation and then examines the small sets of factors that are most important given the chosen plan. (4) The proficient performer is characterised by being involved, having an intuitive understanding of the task followed by detached decision-making. (5) An expert generally knows what to do on the experience-based understanding.

Machine learning (ML)

Research on machine learning (e.g. Kodratoff 1988, Michalski, Carbonell and Mitchell 1983) is mentioned in this review because one of its aims is to understand the principles underlying human learning abilities, in particular, the process of skill acquisition. ML techniques have been applied to various domains such as mathematics, physics and medical diagnosis. In medical diagnosis, ML has been used as knowledge acquisition tool. One example is in the domain of vascular diseases. NIVTIS (Schijven et al 1989) is a system for the interpretation of non-invasive test data obtained from patients that may suffer from peripheral vascular disease in the legs. The system learns the concept of pressure curves from examples.

4.3.1 Discussion of generic developmental models

Issues such as representations of the developmental process and implementations of that process are examined.

Representing the acquisition of expertise

The activity of proceduralisation is common to a number of models. Both generic models of expertise such as ACT* and domain specific models of expertise such as found in Lesgold's in radiology (1981, 1984), emphasise the importance and the use of proceduralisation. Dreyfus and Dreyfus (1986) do not explicitly mention declarative or procedural knowledge in the process of skill acquisition. However, one can draw two analogies between their model and the ACT* model: firstly, there is an analogy between the context-free elements that the novice learn to recognise and the declarative kind of knowledge that the novice first acquires. Secondly, there is an analogy between situational elements that are meaningful elements for the advanced beginner who can recognise it and the procedural knowledge that the learner compiles from the declarative knowledge. In the same way that there is a transition between declarative and procedural knowledge, there is a transition between context-free knowledge and situational knowledge. Another similarity between the ACT* model and the five-stages model concerns the role of practical experience in skill acquisition. In ACT*, experience triggers the knowledge compilation process which will result in procedural knowledge being created. In the five-stages model, experience triggers situational elements which (as seen) can be viewed as procedural knowledge.

While both models may have similarity in the way a skill is acquired, they differ in the tuning of the skill. In ACT*, once the skill has been acquired further learning processes such as generalisation operate on the skill, whereas in the five-stages model once situational knowledge has been acquired the learner uses plans to organise the situation which simplifies and improves her performance.

In addition to the importance of proceduralisation, the reasoning process of generalisation is also found to play an active role in skill acquisition. For instance, generalisation over production rules occurs both in the genetic graph and the ACT* theory. Likewise, the process of specialisation over productions is used in these two models, though specialisation is referred to as discrimination in ACT*.

Compared with the other models of skill acquisition, Dreyfus and Dreyfus emphasise other dimensions for consideration in the process of skill acquisition. This includes for instance, the kind of commitment that the performer has towards her task; ranging from a detached to an involved attitude. The nature of the decision that the performer made is also important, whether it is analytical or intuitive. These features have not been found, at least explicitly, in the other models.

Implementations of the generic developmental models

The ACT* theory has been incorporated in the ACT system. In the system, facts are encoded in a propositional network and procedures are encoded as production rules. ACT has a set of conflict resolution principles which specify how productions are selected to apply. The ACT* theory has also been implemented in a LISP tutor. The five stages of skill acquisition have not been implemented. Dreyfus and Dreyfus suggest that the first three stages of skill acquisition, novice, advanced beginner and competent could be simulated by a computer. However, they argue that implementation of the last two stages, proficient and expert, would not be feasible because modelling intuition into a computer program has not yet been achieved. Proficient performers and experts are *not* aware of looking for facts and inferring goals and actions; they are *not* aware of choosing any goals or actions. In contrast, improved performance in computer programs will result from more and better organisation of their context-free facts in terms of goals and hence from more and better rules inferred.

To summarise, reasoning processes such as generalisation and specialisation which are made explicit in skill acquisition were found in the domain-specific and generic developmental models. The role that these reasoning processes play in skill acquisition has implications for the design of a student model for a medical tutor that will model development of medical reasoning processes. For instance, reasoning strategies have been used to pursue the developmental process, as in the genetic graph where an evolutionary link helps the development of procedural knowledge, or as in ACT where the reasoning processes help in improving the acquired skill to tune the acquired skill. However, there is no report, in literature that was reviewed, of how such reasoning strategies evolve during the development of expertise. In particular, there is no report of the sort of interactions that occur between these reasoning processes.

4.4 Developmental models and tutoring

This section discusses the issue of development of expertise and its application for student modelling in intelligent tutoring systems. Specifically, this issue is examined with reference to the developmental models that were reviewed in the previous sections. The notion that user models in ITS should capture developmental processes is not new. For instance, Self (1979) pointed out developmental student models as one of the difficulties in building student models: the student model is intended to represent the student's state knowledge and as the student learns, the contents of the student model should also represent these learning changes. This problem of development of expertise for student modelling has been investigated from a number of perspectives - pedagogical, novice/expert and cognitive psychology. These are discussed below.

Pedagogical perspective

The causal model progression and the genetic graph have been built with instructional goals in mind. White and Frederiksen (1987) have proposed a learning environment, QUEST, based on a progression of causal models. The learning environment lets students solve problems, hear explanations produced by a speech synthesiser and perform experiments, all in the context of interacting with a dynamic simulation of circuit behaviour. Instruction is viewed as producing in the student a progression of models. In the system, the student model, the tutor and the domain simulation are incorporated within a single model that is active at any point in learning. This model is used to simulate the domain phenomena, generate explanations by articulating its behaviour and provide a desired model of the student's reasoning at that particular stage in learning.

QUEST illustrates the advantages of using models of development of expertise for tutoring purposes. The causal model progressions not only correspond to possible models of the evolution in students' reasoning about electrical circuit but are also used as the basis for a tutoring environment that helps students learn. While the use of causal models of progressions has been found to be successful for tutoring, one of its limitation lies in the domain of application. White and Frederisken (1987) claimed that the application of causal model progressions would be limited to:

> "...any domain whose phenomena can be represented by laws affecting the behaviour of objects..." (p.2).

Instances of such domains that are suggested include Newtonian mechanics, economic systems and biological systems such as the human heart. Even if some aspects of medicine are open to this approach, it seems difficult to envision causal models progressions for teaching medical diagnosis as this task involves dealing with incomplete and uncertain knowledge.

The genetic graph has been incorporated in WUSOR an expert-based coach which is a tutor for the game WUMPUS. The genetic graph guides the tutor in two ways. First, it suggests which skills to discuss with the student, that is, those at the edges of the student's position in the graph. The assumption made here is that learning is facilitated by being able to explain a new skill in terms of those already acquired. Secondly, once a topic has been selected, the genetic graph can explain that skill in diverse ways. This capability of the genetic graph derives from the fact that a rule can be explained in terms of its genetic links. So for instance, a rule that has two links such as analogy and refinement may be explained in terms of either of these processes.

One disadvantage of the genetic graph is that the graph is predetermined and tracing the student's progress is limited to the initial static graph. For a larger domain such as medical diagnosis for instance, this is an unrealistic approach. Bretch and Jones (1988) have suggested a dynamic structure of the genetic graph that would require maintaining nodes and links. To generate new sections of the graph and to discard old ones, maintaining nodes would consist of adding new nodes as new skills are acquired, updating current nodes for which new information is available and removing old nodes which are no longer relevant to the student model. Bretch and Jones have shown how the genetic graph can be extended for better student modelling and hence for enhancing tutoring. Moreover, the feasibility of adapting the genetic graph to other domains has demonstrated once more the usefulness of student models based on the developmental processes.

Novice/expert perspective

The work of Lesgold (1981) and of Dreyfus and Dreyfus (1986) on the development of expertise has some implications for tutoring. Lesgold proposed the general research direction of using their model of the acquisition of radiological skill as the basis for a formal model of the the diagnostic process that could be used for tutoring. For instance, students might be able to access simulated diagnoses on the computer, and the computer could then help the student to learn from these simulations by providing, for example, appropriate problems. However, no further details of how this could be achieved were suggested.

In contrast to Lesgold's position, Dreyfus and Dreyfus adopt a pessimistic view of the computer as a tutor in general and of modelling skill acquisition for tutoring in particular. This issue is related to what levels of expertise can be implemented (discussed in section 3.2.1). Dreyfus and Dreyfus have argued that only the first three levels of their model could be successfully modelled into a computer and further levels which involve intuition and common sense could not be modelled. Similarly, for tutoring purposes, only some of the levels could be modelled. At the beginner's level, the computer can be useful to teach facts, rules and procedures such as spelling and subtraction. However, expertise in teaching does not solely consist of knowing complicated rules about the discipline or of coaching. Rather the expert teacher relies on her intuition and common sense to tutor students. Dreyfus and Dreyfus do not reject entirely the potential of modelling the skill acquisition process for tutoring. Their position is that

> "...one should not attempt to tutor any higher level of skill, for that would require giving logic machines skills that are proved to be beyond their capacities." (1986, p.157).

Cognitive psychology perspective

The ACT* theory of skill acquisition has been embodied into a LISP tutor (Anderson et al 1989). The LISP tutor aims to get the student to mimic the steps of the ideal production model. The tutor immediately provides feedback if the student makes a mistake and gives the student the opportunity to make the correction. The tutor also provide a correct step into a solution if the student appears to repeat the same type of error, or if the student request an explanation.

In Anderson's view, the mechanism of knowledge compilation can change the granularity level of expertise by combining existing rules. Knowledge compilation plays an active role in the representation of expertise. However, *skill* acquisition is not equivalent to the acquisition of *expertise.* One can certainly become *skilled* in programming in LISP, but could not become an *expert* LISP programmer using the LISP tutor. Whereas skill acquisition can be tested by straightforward measures, it is not the case with expertise. The model of Dreyfus and Dreyfus has suggested that expertise is a more subtle notion which goes beyond learning complex rules. As Wenger (1987) pointed out skill acquisition is a necessary aspect of the acquisition of expertise and hence it is an important aspect of learning. In this context, the LISP tutor will be best suited for the instruction of novices rather than for developing more experienced programmers.

To summarise this discussion, modelling developmental processes for tutoring, in particular for student modelling, is a research issue which has been investigated for some time now. Three perspectives (pedagogical, novice/expert and cognitive psychology) from which this modelling can be achieved have been discussed. As regards the medical domain, work in that direction has been limited and as said before has remained largely theoretical. All the developmental models that were reviewed in this chapter have implications for instruction mentioned by their authors, whether in theory such as with Lesgold's work or in practice such as with the LISP tutor.

Modelling skill acquisition in general and for tutoring in particular is very much dependent on the model of reference. The more sophisticated the developmental model is, the more difficult it may be to implement it. For instance, the five stages model proposed by Dreyfus and Dreyfus has brought to light several features of expertise which cannot be implemented due to non-availability of relevant tools. In contrast, as Anderson et al (1989) claim, while the ACT* theory is complex, the process of acquiring a complex skill like LISP programming is simple. All the complexity is due to the structure of the domain, reflected in the structure of the productions, and not in the learning process.

The need to design student models that model the development of expertise has been stressed recently by Richardson (1988). Realistic student modeling in ITS requires the modelling of the student's cognitive development throughout the course of acquiring expert-level competence in a domain. It should track and monitor the changes that occur as a novice becomes expert.

4.5 Summary

This chapter has reviewed the development of expertise. Domain-specific as well and generic developmental models were discussed. Developmental models are not new and this review has illustrated the variety of research which exists. However, as regards the medical domain, research in that direction has been restricted to the theoretical level. Furthermore, this review has also showed that some developmental models contain explicit reasoning processes (such as generalisation and specialisation) that help in describing the developmental process. However, there was no report from the literature reviewed, of developmental models that focus on the development of reasoning strategies, particularly of medical reasoning processes. By development of reasoning strategies one means how strategies evolve over time, and the kind of interactions between these processes that occurs as part of the developmental process.

The review also showed that modelling the development of expertise for tutoring was important and further research was needed. Again, regarding the medical domain, the development of medical expertise for tutoring has been proposed but not yet implemented.

This chapter concludes the necessary literature review which forms the basis for the research conducted in this thesis. Given the interdisciplinary approach adopted in the research work, three research areas were reviewed - ITS in medicine, medical problem solving and teaching of medical diagnosis and development of expertise. The next chapter synthesises the important findings of these reviews, and describes the design of a developmental student model for a medical tutor based on these findings.

Chapter Five

DESIGN CONSIDERATIONS FOR A DEVELOPMENTAL USER MODEL

In the previous three chapters, research in the areas of ITS in medicine, medical problem solving and developmental models have been reviewed. The purpose of these reviews was to examine the state of the art in these different research areas in order to investigate the development of medical reasoning processes and how student modelling of such processes can be achieved in ITS. This chapter synthesises important findings from the literature review which led to the design of a developmental student model for medical diagnosis called DEMEREST[1]. The first section of this chapter describes the idea of a developmental student model for medical diagnosis. An overview of the system is then presented. The planning approach chosen to build such a system is discussed and related research in planning in medical problem solving is examined. The role that DEMEREST could play within a medical tutoring system is examined in a further section. Lastly, the domain of medicine in orthopaedics in which the system has been built is detailed.

5.1 Developmental user model

The review of intelligent medical tutors in chapter two helped establish which aspects of intelligent medical tutors that the research should address. This review showed that the development of ITS in medicine

---

[1] DEMEREST stands for DEvelopment of MEdical REasoning STrategies.

has been *expert systems oriented.* Given the increased development of expert systems in Medicine (Clancey & Shortliffe 1984), such systems have become the main resources in the development of intelligent tutoring systems in medicine (e.g. GUIDON, Clancey 1979 based on MYCIN). Moreover, this review indicated that the approach to student modelling for tutoring medical diagnosis has been expert based. For example, GUIDON contains a subset student model whereby the student's knowledge and understanding of the domain is entirely represented in terms of the expert physician's. Such a model is not sufficient to explain the student's reasoning. Not only may the medical student take a different approach to the expert physician, but also her reasoning is not static and evolves over a period of time. An alternative approach (Alpay 1988b) is to suggest that an intelligent medical tutor should have an understanding of the clinical reasoning of the student from the student's point of view and not from the expert physician's. In other words, it is proposed that student modelling for medical tutors should be based on the student's perspective of the problem in hand and not on the expert's.

The review of the medical problem-solving and medical education literature in chapter three provided a discussion of the features of a student model to be based on the student's medical reasoning processes. The review showed that the medical reasoning process can be decomposed in terms of its *contents,* that is, the medical knowledge, and its *form,* that is, the reasoning strategies that support that knowledge. The form of medical problem solving used by students and experts was found to be similar, while the contents (not surprisingly) were found to differ. The literature review also indicated that medical students have various needs in their learning phase. Firstly, they need to strengthen their medical knowledge. Medical students have to learn and assimilate a vast amount of medical knowledge. An illustration of how a computer tutor could help in that task is given in the system QMR (reviewed in chapter two) which offers medical students a means to explore medical knowledge in a flexible way. QMR can operate as an electronic text book: its knowledge base describes the clinical manifestations of some 600 diseases in the domain of internal medicine. Secondly and more importantly to the objectives of this research, students need to get feedback on their strategic knowledge (Balla et al 1989, Irby 1986, Buchanan 1990). Students usually have access to their performance rather than their reasoning processes which are not part of their preparation for clinical practice. Moreover, medical students are expected to learn an increasing amount of knowledge in ways that favour passive reception learning instead of stimulating the use of strategic methods to manipulate that knowledge. In the context of tutoring, it is important for the intelligent medical tutor to know how the student progresses and how her reasoning develops. As a result of this literature review, it was decided that student modelling for medical tutors should focus i) on the form of medical reasoning, particularly on the reasoning strategies applied during the diagnostic process and ii) on the development of these strategies.

The literature on developmental models in chapter four demonstrated that although the concept of developmental process (from novice to expert) is not new, its application to medicine has been limited (e.g. medical perceptual skill, Lesgold 1984) and mostly researched at a theoretical level. Moreover, the review showed that this concept of developmental process had not focussed on the development of medical reasoning strategies.

The new type of student model which is being proposed for the teaching of medical diagnosis is referred to as *developmental* (Alpay 1989b, 1990b). The main features of the developmental user model include the following:

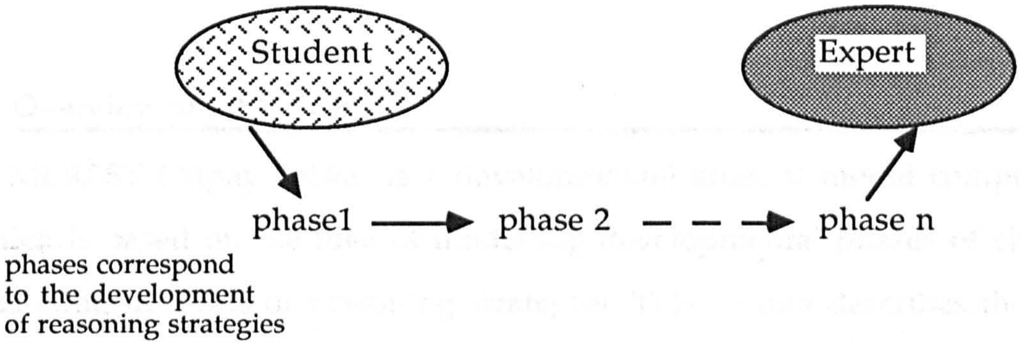


Figure 5.1: Notion of the developmental student model

- The student model is designed to maintain a representation of the current state of the student from the student point of view by incorporating the clinical reasoning strategies of the student (figure 5.1). Thus, the student's knowledge is not represented in terms of the expert physician. By modelling the reasoning processes of the student, the medical student's own view of the problem is being represented.

- The student model is designed on the assumption that the student's knowledge is a progressive and dynamic entity. That is, the student goes through a process of development of medical diagnostic expertise. The developmental student model that is being proposed differs from other student models (e.g. subset model, perturbation model) discussed in chapter two since in these models the student's knowledge is represented in terms of the expert's knowledge. It also differs from the bounded student model which is based on the learning process of the student. The bounded user model does not take a developmental approach to the student's knowledge, that is, it does not integrate different levels of expertise. There is no claim that a developmental model is the answer to the representation of the student's current state. However, it is believed that such a model can provide a complementary or alternative way to the already existing user models.

5.2 Overview of DEMEREST

DEMEREST (Alpay 1989a) is a developmental student model component which is based on the idea of modelling developmental phases of clinical reasoning in terms of reasoning strategies. This section describes the role of each of the components of the system (see figure 5.2).

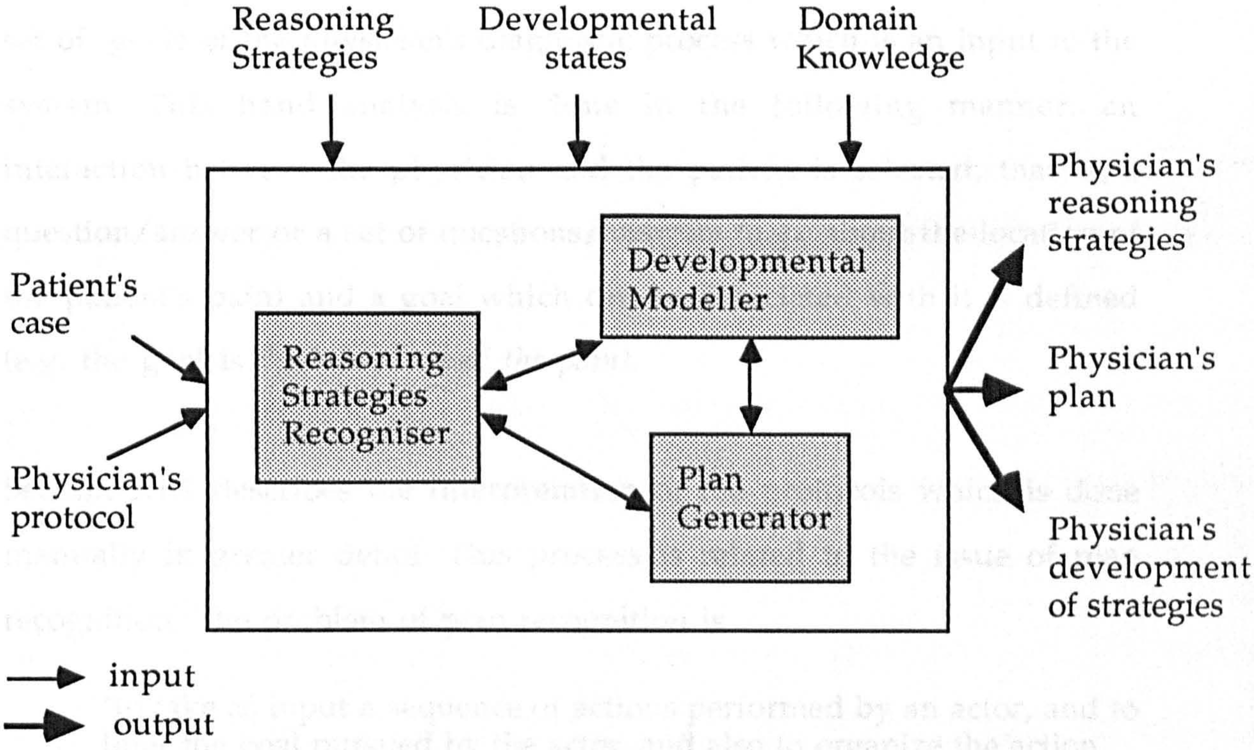


Figure 5.2: Overview of the system DEMEREST

The system analyses the physican's diagnostic actions and level of expertise and uses planning as a formalism to decompose the physician's medical problem solving task into a set of goals[2]. In particular, the tasks of the system are 1) to diagnose the reasoning strategies that the physician has used, 2) to identify development of these strategies that will help the system in determining the physician's level of expertise and 3) to produce a plan corresponding to the application of these strategies.

5.2.1 Input to DEMEREST

As an input, DEMEREST takes the physician's protocol that corresponds to a consultation between the doctor and a patient. The physician has been interviewed and his or her protocol recorded and analysed by hand analysis (see chapters seven and eight). As part of the hand analysis, the physician's protocol is transformed into a plan. The plan is made up of a set of goals of the physician's diagnostic process which is an input to the system. This hand analysis is done in the following manner: an interaction between the physician and the patient is selected, that is, a question/answer or a set of questions/answers (e.g. about the location of the patient's pain) and a goal which can be associated with it is defined (e.g. the goal is *check location of the pain).*

Section 9.1.1 describes the interpretation of the protocols which is done manually in greater detail. This process is related to the issue of plan recognition. The problem of plan recognition is

> "to take as input a sequence of actions performed by an actor, and to infer the goal pursued by the actor, and also to organize the action sequence in terms of a plan structure" (Schmidt et al 1978 p.52).

Plan recognition is by itself a complex research problem which was not pursued in this thesis but will be discussed in further work (section 11.4).

---

[2] The word physician refers to medical students as well as more experienced physicians.

The other input to the system is the patient case. Information of the patient case is included into the goals (as will be explained in section 5.3.3). DEMEREST has three components: 1) a reasoning strategy recogniser, 2) a developmental modeller and 3) a plan generator. For each component, the role, the information needed as input and the output produced are described.

5.2.2 Reasoning Strategies Recogniser

This component carries out the first task of the system which is to diagnose the reasoning strategies applied by the physician. Given a goal, the reasoning strategy recogniser identifies the strategy associated with it. Its inputs are a goal (and the data objects associated with that goal) from the physician's plan which was generated through the hand analysis, and the set of reasoning strategies (described in chapter six). The output is an *instantiation* of one of these reasoning strategies, that is, the context in which the strategy has been applied (e.g. the goal, the hypotheses generated etc). It should be pointed out that there may be more than one strategy associated with a goal. A strategy is instantiated by accessing from a given goal the necessary medical knowledge in the medical database that characterises that strategy.

An example of an instantiation of the hypothesis generation strategy applied by a 4th year medical student is shown below:

Strategy applied: Hypothesis Generation
Goal: check location of the pain
Observation: right sided back pain
Hypothesis: kidney problem

The student asked the patient about the location of the pain and being told that the pain is on the right side of the back, concluded that there may be a kidney problem.

5.2.3 Developmental Modeller

This component carries out the second task of the system. That is, its role is to identify interactions of the strategies applied by the physician, and determine the physician's level of expertise. The level of expertise is determined by the interactions between strategies that are identified. Given the physician's goals, the developmental modeller not only generates the physician's reasoning strategies (using the reasoning strategies recogniser), but also determines the physician's level of development. The inputs are the physician's set of goals, the set of reasoning strategies, and the expected development of reasoning strategies (derived from the empirical study described in chapter eight). The outputs are instantiations of reasoning strategies and interactions of these strategies applied by the physician along with a level of expertise determined for the physician. Since the developmental modeller contains a model of clinical reasoning at different levels of expertise, it can match the physician's reasoning with this model and determine the physician's level of expertise.

5.2.4 Plan Generator

This component fulfills the last task of the system which is to generate the physician's plan. The inputs of this component are the physician's goals and the reasoning strategies she has applied, and the output is the physician's plan. The plan contains the goals and their associated strategies in the order which corresponds to the physician's diagnostic process during a consultation with a patient.

5.3 The planning approach

In this section, related research on planning in the context of medical diagnosis is first examined and the use of planning in DEMEREST is then discussed. Finally, a number of planning features of the system are described in the remaining subsections.

5.3.1 Planning in medical diagnosis: Related research

The view that expert physicians use plans to make medical decisions has been supported by a number of studies. In one study (Kuipers, Moskowitz, and Kassirer 1988), the decision process observed in the protocols is described as

> "one of planning by successive refinement of an abstract plan, combined with opportunistic insertion of plan steps" (p.193).

Expert physicians make an initial decision at an abstract level and go on to specify it more precisely. Although this work provides interesting insights of the structure of a medical decision, it takes the concept of medical decision in a broad sense, that is, on the management of the patient (which includes diagnosis, treatment, referral etc).

One aspect of the patient management involving planning is the task of taking a present illness (Miller 1975). A present illness corresponds to a description of the patient's presenting problem. Taking a present illness is different from performing a complete diagnosis, in that it is limited to doing what can be achieved during an initial consultation, and excluding investigations. The patient usually presents a 'chief complaint' that becomes the initial focus of the consultation, and diagnosis is based on only very low cost sources of information (such as patient history, physical examination and routine laboratory tests). High cost or risk procedures that may be necessary for a complete diagnosis are not used.

Miller suggests that there are two distinct but closely related planning activities involved in taking a present illness: data acquisition and diagnosis. Data acquisition planning specifies what data to look for next, whereas diagnostic planning specifies what to do with each piece of data once it has been obtained. Miller's study concentrates only on data acquisition planning. The work in this thesis focuses on the diagnosis of the patient, and hence is restricted to patient diagnosis planning and not patient management planning. The idea of diagnosis planning is different from Miller's approach for two reasons: 1) the research is concerned with the planning approach taken by novices (e.g. students) as well as more experienced physicians, rather than just by experts and 2) the research focuses on the reasoning strategies that the physician has applied during her diagnostic process.

Another aspect of patient management which has been considered as a planning activity and reported in the literature is treatment (Langoltz et al 1987). Patients with oncological problems require treatments which are often complex. In this study, the therapy planning pattern identified is as follows: first, expert physicians have been observed to develop a set of possible plans that are reasonable to administer. Then they envision the possible consequences of administering each of these plans. Finally, they assess how well their predicted consequences of each plan meet the treatment goals. As in Miller's research, Langoltz et al consider the use of therapy planning only by expert oncologists and not by novices.

Planning is used differently in both studies. In the first study (Miller 1975), the idea of planning is general. A plan specifies what data to look for next; goals are what a doctor hopes to accomplish by obtaining the requested data. Data acquisition strategies determine the contents, form and sequence of the questions that are asked. The focus is on developing a strategy frame model which describes the strategies and on providing a mechanism for the selection of a strategy. The study remains at the theoretical level; no system has been implemented to incorporate the idea of data acquisition planning. In contrast, in the second study (Langoltz op.cit.) the idea of planning is central to the design of a therapy planning system. A new planning architecture for such a system is described where techniques of planning from Artificial Intelligence (AI) are combined with a decision theoretic approach. In the implemented system called ONYX, a plan specifies the therapy to administer therapeutic goals. These goals represent what the oncologist hopes to obtain by administering a certain therapy. In addition, both studies consider planning as a problem solving behaviour. However, Langoltz et al introduce the concept of plans as the result of observing expert oncologists performing, whereas Miller does not provide any data or argument that will support the idea of data acquisition and diagnosis as planning activities.

Although, planning in the context of medical diagnosis deals with most AI planning issues (e.g. interactions of goals, planning process reported in the next sections), it also raises a number of important matters that are not always considered in AI planning research. A first issue is related to uncertainty: the state of the patient and the effects of actions are both uncertain. In Langoltz (op.cit.), the problem of uncertainty in planning is dealt with by integrating decision theoretic technique to the planning system.

A second issue is related to planning with incomplete knowledge: planners such as NOAH (Sacerdoti 1974) or STRIPS (Fikes and Nilsson 1971) implicitly assumed that the planner has complete knowledge. Making a medical diagnosis usually means reasoning with incomplete information. Some researchers have investigated the use of planning with incomplete knowledge. For instance, Morgenstern (1987) has proposed a highly flexible model of action and planning that is well suited for partially specified plans.

A third issue is related to the domain where planning is applied. Traditionally, AI planning has been tested with toy worlds such as the block world (e.g. Sussman 1975). However, more recently planning applied to real world situations has attracted attention (e.g. Hayes-Roth and Hayes-Roth 1979 and Wilensky 1983). The first two issues have not been addressed in this thesis. The third issue has influenced the use of planning in this research work and further investigation of its application for a real world and complex situation such as medical diagnosis. Moreover, the studies mentioned earlier on (Kuipers, Moskowitz, and Kassirer 1988, Miller 1975 and Langoltz et al 1987) have illustrated the importance of planning in medical diagnosis.

5.3.2 Planning in medical diagnosis and DEMEREST

Planning involves decomposing a problem into subparts before achieving a problem solution. Indeed, medical diagnosis is a complex problem solving skill which can be broken down into manageable pieces to be worked on and decided separately, and then combined to reach a diagnosis for the patient case. For instance, the physician will usually take the history of the patient, and examine the patient before reaching a final diagnosis (although the physician may have already formed initial diagnoses during the history taking).

In DEMEREST, planning is used as a means of *representing* medical reasoning, and no new planning mechanisms are proposed. By adopting this approach, it is suggested that medical reasoning be viewed as a planning process and hence be decomposed into a set of goals, the goals being associated with the reasoning strategies. Each level of expertise corresponds to a view of the medical diagnostic process, and at each level of expertise the applied reasoning strategies and interactions of strategies associated with their goals form a plan for that level. The planning component of DEMEREST possesses several features that are common to most AI planners. These include the representation of plans, the planning process, and interactions of the goals. The following sections describe each of these features.

5.3.3 Representation of Plans

As mentioned in the previous section, a plan is used to represent the physician's diagnostic processes. Specifically, the physician's decisions are viewed as goals that she needs to achieve in order to reach a diagnosis for the patient case. A plan consists of a hierarchy of goals designed to achieve the desired goal i.e diagnosis of the patient. The goals represent the *operators* of the planner. A goal contains the following information:

Name of the goal
Precursors of the goal
Subgoals of the goal
Action of the goal
Effects of the goal

The name of the goal indicates the specific goal to be achieved. In the usual AI approach to planning, one refers to preconditions that must be true before a goal can be executed. Precursor rather than precondition is used here. A precursor is a specification of the state of the goal that has occurred before the goal can be applied. The distinction between precondition and precursor allows one to differentiate between a physician's preconditions that she had in her mind and the preconditions that can be inferred from her consultation with a patient (as reported in a protocol). For instance, the physician may have wanted to carry out the goals G1, G2 and G3 before achieving G4. The transcription of this plan in her protocol could be "carry out G1, then G2 then G3 and then G4" or it could be "carry out G2, then G1 then G3 and then G4". One cannot be sure which ordering of the goals the physician used. Moreover, it does not make medical sense to say, for instance, that the precondition *no radiation of the pain* needs to be true in the world before applying the goal *check location of the pain.* Rather, one can say that the precursor *no radiation* is a pointer to a previous goal *check radiation of the pain.*

A goal may be decomposed into a set of subgoals. An action is an action in the world which carries out a goal. An effect is a state of the world after an action has been performed. The effect of an action corresponds to some observations (signs, symptoms and test results). Attached to each name of a goal is the prefix *check* and to each action of a goal the prefix *ask* to distinguish between the slots of the goal. The prefix *check* is taken in a broad sense - checking if the pain radiates, checking if there is tenderness in the back, checking for an X-ray of the back. The prefix *ask* is also taken in broad sense - asking the patient a question, asking for a test to be done.

As an example of these points, the decision to gather some information about the onset of the pain is represented as the goal below:

Name of the goal: check onset of the pain
Precursors of the goal: no radiation
Subgoals of the goal: *none*
Action of the goal: ask about the onset of the pain
Effects of the goal: onset of the pain this morning

A typical plan contains one or more goals to achieve (see figure 5.3). The main goal is the diagnosis of a patient's problem. Figure 5.4 shows an example of a plan. DEMEREST may be viewed as a non hierarchical type of planner. Unlike hierarchical planners, it does not generate a hierarchy of representations of plans in which the highest is a simplification of abstraction of the plan and the lowest is a detailed plan, sufficient to solve the problem. Instead, in DEMEREST all the goals are at the same level of abstraction and the assumption made is that all goals are important to the diagnostic problem solving task.

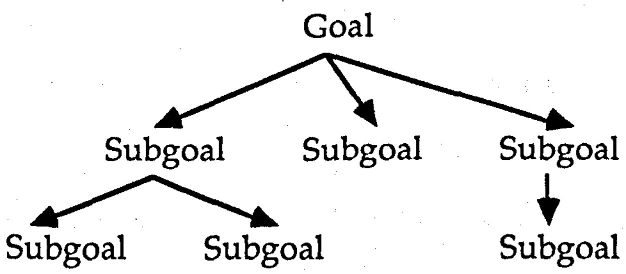


Figure 5.3: The structure of a typical plan

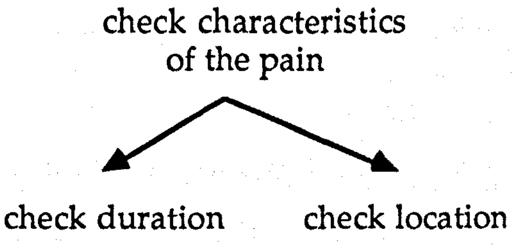


Figure 5.4: An example of a plan

As mentioned in section 5.3.1, Kuipers et al (1988) reported that expert physicians combined successive refinement of an abstract plan with opportunistic insertion in plan steps. In DEMEREST, the opportunistic aspect of the medical diagnostic process is not included. This does not imply that opportunistic planning is rejected but simply that it is not investigated in this research work.

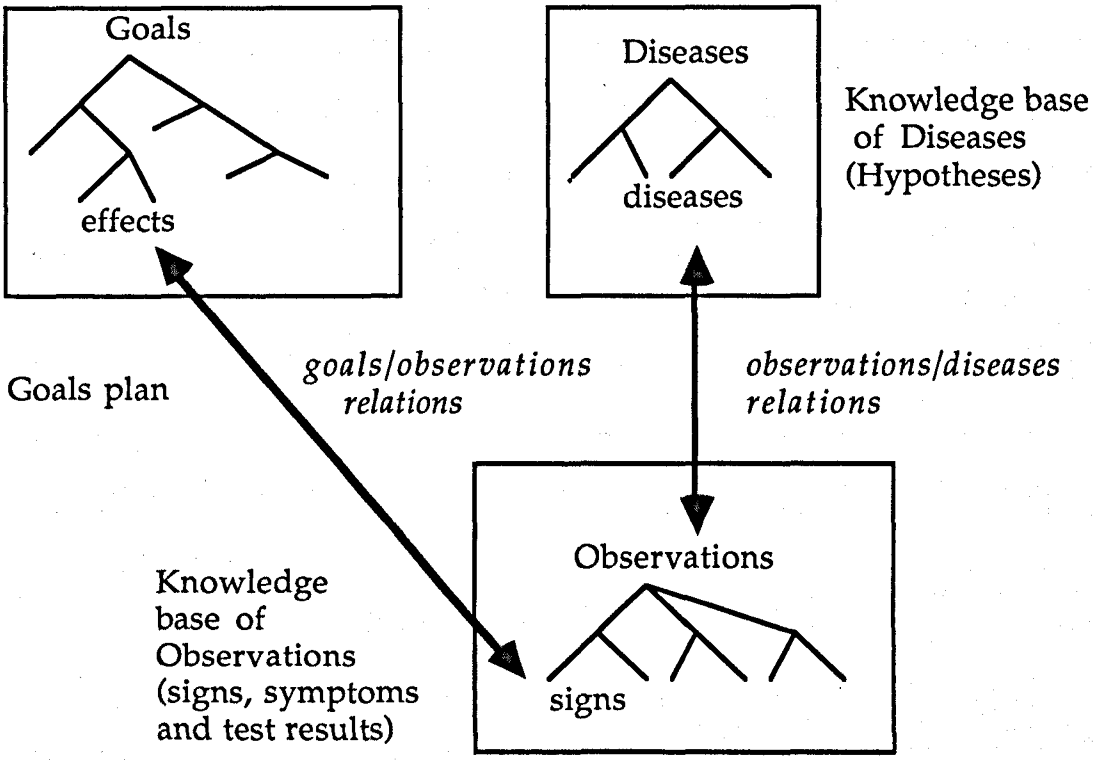


Figure 5.5: Goals plan and the knowledge bases

The set of information present in a plan does not convey all the necessary knowledge to diagnose medical problems in general and back problems in particular. There is a large body of knowledge that the plan representation cannot handle in a flexible way. Hence, in order to reduce this limitation, the goals of the plan are extended and complemented by two knowledge bases which form the medical knowledge of orthopaedics: the knowledge base of observations contains knowledge about signs, symptoms and test results of back pain. The knowledge base of diseases contains knowledge about diseases of back problems.

Goals in the plan are linked to the knowledge base of diseases and observations in the following way (see figure 5.5): an effect of a goal corresponds to questions put to a patient, the effect of which is a reply such as a sign, a symptom or a test result. The effect is connected to the knowledge base of observations which contains signs, symptoms and test results. These elements of the knowledge base of observations are in turn linked to their associated diseases and thus linked to the knowledge base of diseases. For example, the effect *right sided pain* of the goal *check location of the pain* is a symptom found in the knowledge base of observations which is caused by the hypothesis *kidney infection* in the knowledge base of diseases (see figure 5.6).

5.3.4 Planning Process

The planning process is concerned with how the plan is to be achieved. It involves the goals which are to be carried out and the ordering of these goals. In DEMEREST, the way a plan is processed is driven by the reasoning strategies that have been applied. As mentioned in section 5.2.2, it is the role of the reasoning strategy recogniser to associate goals of the plan and strategies by accessing appropriate information in the slots of a goal and in the knowledge bases that characterise the strategy.

A goal linked with a strategy corresponds to what the physician tries to achieve, and also to the "context" in which the strategy is applied. Specifically, a goal may be tied to one or more strategies in the following cases:

i) *Reacting to data:* The physician may react to some piece of information that the patient has volunteered or that the physician had gathered by asking the patient. For example, the patient may say that she has *right* *sided back pain* which brings to mind the possibility of a *renal infection.* The physician then generates the hypothesis *renal infection.* The goal is *check location of the pain,* and the strategy here is one of hypothesis generation. In this case, one may say that the strategy hypothesis generation is associated with the goal *check location of the pain.*

ii) *Posing a question:* the physician may want to acquire specific information by posing a question. For example, the physician may be thinking of a possible renal infection, and hence will probe the patient specifically to check whether she had right sided back pain. As above, the goal is *check location of the pain,* and the strategy is of hypothesis generation. This example illustrates a hypothesis generation driven data acquisition. In this case, one may say that the strategy hypothesis generation has led to posting the goal *check location of the pain.*

The usual approach in AI has been to treat planning as being independent of execution, that is, the activity of planning is fully completed before any execution takes place. In contrast, in DEMEREST, the building of a plan and its associated reasoning strategies are intertwined: applying reasoning strategies does not only form a plan, but also reflects the execution of the plan. The presence of different reasoning strategies applied either by the medical student or the experienced physician mirrors the fact that goals of the plan have been manipulated in different ways. For example, (taken from empirical data - chapter eight) figure 5.6 shows that both a 3rd year student and a house officer have the same goal of achieving *check_location_of_the_pain* in their plans. However, the student has applied a problem refinement strategy with that goal by considering the location of the pain as part of a routine protocol for the characteristics of the pain, while the house officer has applied a hypothesis generation strategy by generating the possible hypothesis *kidney infection*. In other words, in this particular case, the student was not hypothesis directed, whereas the house officer was.

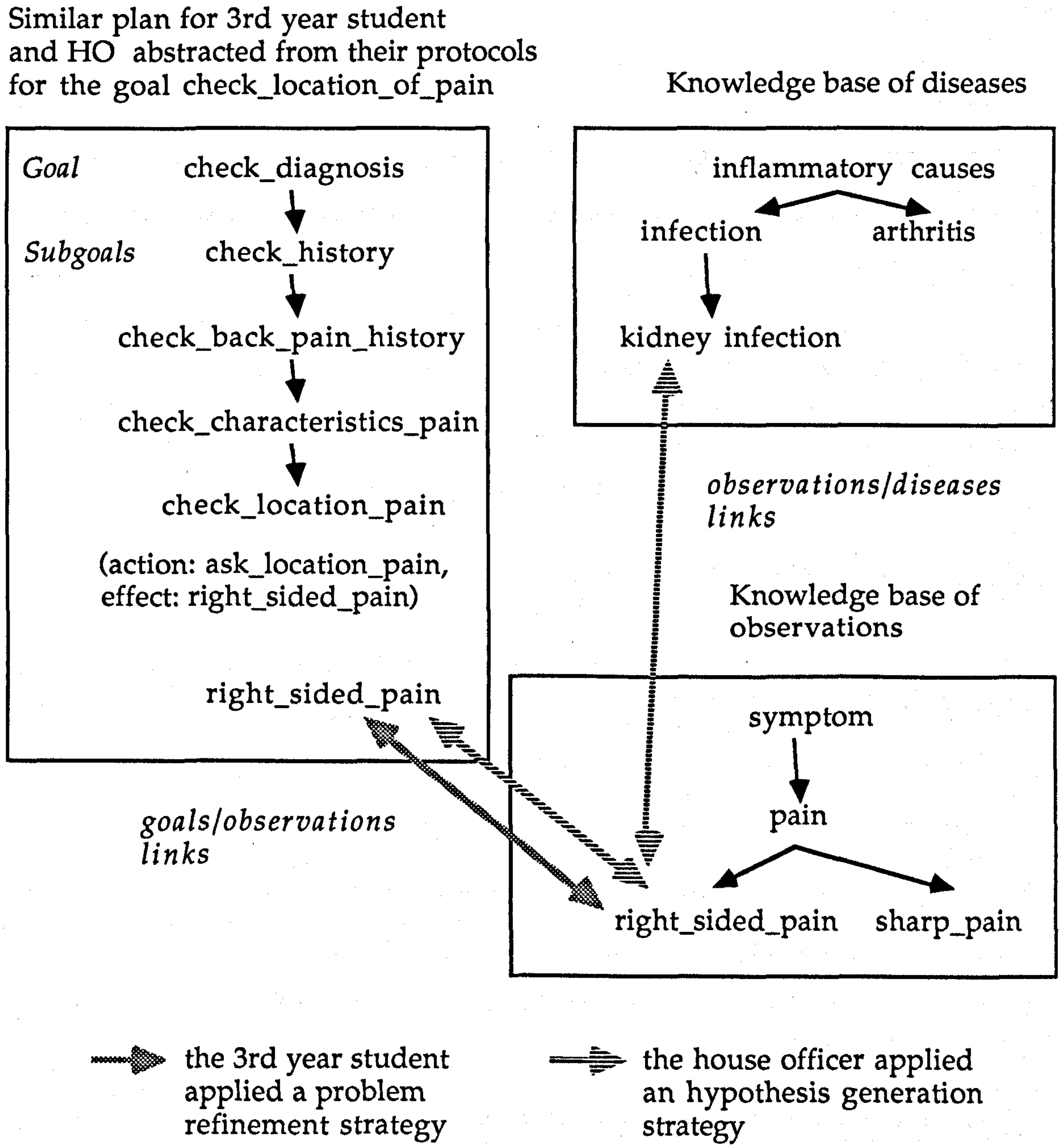


Figure 5.6: Example of plans and reasoning strategies

5.3.5 Interactions of Goals

Goals rarely occur in isolation. In most cases, a problem is solved using a conjunction of goals. The interactions of goals are directly associated with the reasoning strategies applied to produce the plan, and represent the order in which the plan steps are being carried out. In the system, goals relate to one another by 'effect/precursor' links, that is, the effect of one goal is a precursor of another goal. This kind of goal interaction corresponds to sequencing of goals: the achievement of one goal allows another goal to be carried out. For example, consider the actual case of a 4th year student whose plan contained the goals below.

Goal1(4th year student)

Name of goal1: check radiation of the pain
Precursor of goal1: *none*
Subgoal of goal1: radiation of the previous pain
Action of goal1: ask about radiation of the pain
Effect of goal1: no radiation of the pain

Goal5(4th year student)

Name of goal5: check location of the pain
Precursor of goal5: no radiation of the pain
Subgoal of goal5: *none*
Action of goal5: ask about the location of the pain
Effect of goal5: right sided back pain

The student had asked the patient about radiation of the pain, and was told that there was no radiation (goal1). Then the student asked the patient about the location of the pain, and was told that the pain is on the right side (goal5). Once goal1 had been activated and its associated reasoning strategy generated, the system is looking for a next goal to achieve to pursue the plan. To do so, the system takes the effect of goal1 and searches for a goal (in this case goal5) whose precursor slot is the same as the effect. Goal5 can only be achieved if its precursor *no radiation of the pain* exists, that is, if goal1 has already been achieved.

In summary, this section has discussed the planning approach adopted in the design of DEMEREST. In this approach, the diagnostic process is decomposed into a set of goals, each goal being associated with one or more strategies. Related research which have used planning in the context of medical diagnosis was also examined.

5.4 DEMEREST and intelligent medical tutors

In this section, the role of DEMEREST within a medical tutoring system is examined. While this thesis does not report on the construction of an intelligent medical tutor as it is outside the scope of the research work, some suggestions are made on how the system would be used by a medical tutor.

Components of intelligent tutoring systems

The following reports on the components usually found in ITS and thus provides some background before discussing the possible role DEMEREST could play within a medical tutoring system. There is some disagreement about how many components constitute an ITS. The traditional model of a tutoring system (Self 1974) is the trinity model which includes a domain knowledge component (also referred to as expert module) because it contains knowledge of an expert in a particular domain, a tutoring component and a student model. A five component paradigm was proposed by O'Shea et al (1983); it includes a teaching strategy, a teaching generator, a teaching administrator, a student model and a student history. More recently, another five component paradigm, which includes the expert, the student model, the psychologist, the instructional module and the interface, has also been suggested (Bretch and Jones 1988).

In chapter two, a number of tutoring systems in medicine were reviewed, each of them including some of these components: GUIDON contains an expert module, a tutor which also acts as the interface with the student by conducting a dialogue, and a student model. In contrast, ATTENDING, whose prime function is not to teach, includes an expert module, a limited teaching interface, but no student model. QMR contains an expert module which is the knowledge base of INTERNIST, a "diagnostic spreadsheet" which serves as a tutor to the student but like ATTENDING does not have a student model. For the purposes of this discussion, the structure of an ITS is said to include an expert (or domain) module, a teaching module, a student model and an interface.

The domain module contains the domain-specific knowledge (e.g. the domain of orthopaedics). The knowledge can be, for instance, in the form of production rules such as in WUMPUS (Goldstein 1982) or semantic nets such as in SCHOLAR (Carbonell 1970). The expert module fulfils two functions (Wenger 1987). First, it acts as the source for the knowledge to be represented which includes generating explanations and responses to the student, as well as tasks and questions. Secondly the expert module serves as a standard for evaluating the student's performance.

The teaching module determines what to teach, when to teach and how to teach the student. Every student is unique and the intelligent tutor should have the ability to vary its teaching methodology depending on the student. Hence, this module should include a number of teaching strategies such as coaching (e.g WEST, Burton and Brown 1982) or a socratic method (e.g. SCHOLAR, Carbonell 1982). The teaching module should have access to knowledge of what is being taught from the expert module and knowledge of who is being taught from the student model. In addition, the teaching module is used by the interface module to channel tutorial communication to the student.

The student model represents the student's understanding of the domain being taught and maintains a representation of the student's current knowledge state. Moreover, this module diagnoses what the student knows. In chapter two, a number of user models for ITS were discussed: the overlay model, the perturbation model and the bounded user model. As discussed in section 5.1, although these models are useful and adequate for certain aspects of tutoring, the argument in this thesis is that one should try to model and diagnose what the student knows from the student's point of view rather than from the expert's. DEMEREST attempts to embody and test the feasibility of this proposal by taking a developmental approach. Specifically, DEMEREST models combinations of interacting reasoning strategies which correspond to various levels of medical expertise.

The last component of a tutoring system, the user interface, is responsible for the interaction and dialogue between the student and the tutor by means of natural languages, pointers, windows, menus, icons etc.

DEMEREST as a component of a medical tutor

The usefulness and importance of modelling of expertise for tutoring purposes has been discussed in section 4.4. DEMEREST, like other development models, attempts to capture development of expertise, in this case, of medical reasoning strategies. The following section discusses how DEMEREST could be used by an intelligent medical tutor and possible properties that a medical tutor should have to support such a system (Alpay, in press).

The tasks of DEMEREST are to identify the student's reasoning processes, determine her level of expertise and produce the plan corresponding to the application of the strategies. Once these tasks have been carried out, the information resulting is passed on to the teaching module which will determine how to use it to teach the student further. In particular, the teaching module can propose alternative strategies and plans given the student's level of expertise; it can also use the information about the student's reasoning strategies and level of expertise as a basis for advising and testing the student on the application of medical reasoning strategies. The student not only has the chance to reflect and get feedback on the performance of her medical reasoning, but also has access to various levels of medical expertise. For instance, let us assume that the system has diagnosed a 5th year student at a level of expertise which does not correspond to the student's level since the student did not apply a particular pattern of reasoning strategies specific to her level. The teaching module may then select the pattern of level of a 5th year student and presents it to the student through examples. A particular pattern, expected of a 5th year student, can be generating hypotheses using a hypothesis generation strategy and then generating a more specific hypothesis using a specialisation strategy (see chapter eight which describes a model of changes of strategy at various levels of expertise).

Since the approach taken in DEMEREST is developmental, that is, the system possesses a model of changes of strategies at different levels of expertise, the domain knowledge that DEMEREST has access to, contains knowledge (in this case of back pain) for each level of expertise, and not only for the expert's level like in other tutors. There may be of course knowledge common to all levels since some pieces of information may be used at more than one level of expertise. For instance, asking the patient about the onset of the pain should be found at more than one level of expertise.

The nature of the task taught, i.e. medical diagnosis, and the kinds of users that would be tutored, i.e. medical students, suggest that the role of the medical tutor should be to aid and advise the medical student, and let the student have control of the interaction, rather than forcing the student into a Socratic tutoring session. In the context of tutoring, it is important for the intelligent medical tutor to know how the student progresses and how her reasoning develops. Since DEMEREST diagnoses what the student does in terms of the student's medical reasoning strategies and represents the reasoning strategies which have been applied as plans, a medical tutor that would use DEMEREST as one of its components should aim to tutor medical reasoning processes as well as emphasise a planning approach to medical diagnosis.

5.5 DEMEREST and its domain of application: Orthopaedics

The medical domain used with DEMEREST is discussed in this section. First, reasons for choosing this domain are given, followed by descriptions of how to diagnose back pain.

5.5.1 Motivation for the domain of orthopaedics

The domain of orthopaedics, specifically low back pain, was selected for a variety of reasons. It is a difficult and vague domain, which is important in the medical curriculum but which is not taught well. Back pain is a symptom with a large number of different causes and has various forms of management. Table 5.1 shows the major causes of back pain. In Great Britain, back pain affects a large number of the people. In fact, back pain is one of the major problems in our society and one of the commonest reasons for requiring time off work (Jayson 1981). Back pain is common and widespread, and occurs at all ages and in all levels of society. This means that at some point in their medical career physicians in general practice will see patients with back pain problems. Likewise physicians in hospital, mainly in the orthopaedics department, much less commonly in the department of rheumatology and also in the departments of neurology, gynaecology and general medicine will see back pain sufferers.

The problem of back pain in the population is quite different from the problem of back pain in hospital. As Jayson (1980) points out the role of the general practitioner is to screen the patients to be referred to the hospital, whereas the role of the hospital doctor is to get the patient earlier in order to improve the chances of helping the patient. This means that the environment in which the physician performs (general practice or hospital) may have implications on the diagnostic process (of back pain).

For instance, in the hospital a number of investigations are carried out that would not be possible in a general practice.

5.5.2 Diagnosis of back pains

In many cases of back pain, it is not possible to arrive at a precise diagnosis. Nevertheless, it is possible to make sure that the patient's problem is correctly categorised (e.g mechanical in origin, inflammatory). Most textbooks give a long list of causes of backache; a possible way to make a diagnosis is to match individual patients against such lists. However as Waddel and Hamblen (1983) point out this is an illogical and inefficient method of diagnosis. They suggest an alternative to clinical diagnosis which is to use the history and physical examination as a basis for a few fundamental decisions, allocating the patient's presentation to one of the three categories: mechanical, non mechanical (spinal pathology) or nerve root. The roles of the history and physical examination is also stressed by Jayson (1983).

The following paragraphs describe what should be done during a consultation with a patient complaining of back pain and how a diagnosis can then be reached. In particular, taking the history, the physical examination and investigations are discussed. Different literature sources on back pain problems were used. These included Jayson 1980, 1981 and 1983, Macnab 1977, Evans 1982, and Waddell and Hamblen 1983. This is not meant to be an exhaustive description about all the action a physician can perform to diagnose back pain, but rather to provide the reader with sufficient background and understanding of the domain used in this thesis.

CAUSES OF BACK PAIN

Mechanical (structural)	prolapsed intervertebral disc	disc protudes backwards and usually to one side or the other
	spondylosis	wear and tear damage of the spine
	spinal stenosis	narrowing of the vertebral canal so that the nerve roots are at greater risk of damage
	fractures	breaches in the structures of bones (in this case of the vertebrae of the back)
	non specific	back pain of mechanical origin (e.g muscle strains) but without precise identification of the problem
Inflammatory	ankylosing spondylitis and related inflammatory spondylo-arthropathies	a systemic disorder associated with inflammation in the joints of the spine, and occasionally elsewhere, together with back pain
	rheumatoid arthritis	a systemic disorder involving changes on the affected joints such as thickening of the lining of the joint followed by involvement of the joint's cartilage which becomes damaged
	infection	for example, bacterial infection, tuberculosis of the spine
Neoplastic	primary tumors	usually tumours of the lining of the spinal cord (meningioma) or of supporting cells in the nervous tissue (glioma)
	metastases	secondary tumors - the spine is the most common site of metastic spread in the skeleton
	myelomatosis reticuloses	2 types of malignant tumour of cells of the body's immune system which are produced in bone marrow
Metabolic	osteoporosis	weakness and collapse of the bones due to a deficent fibrous framework
	osteomalacia	weakness of the bones due to calcium deficiency
Referred pain		referred pain from abdominal or pelvic disorder and felt by the patient to be located in the spine

Table 5.1: Causes of back pain (adapted from Jayson 1983)

History

In taking the history, it is essential to obtain a description of the pain in great detail as back pain is a subjective sensation felt and described by the sufferer and there is no objective means by which the pain can be proven. Different people feel pain to different extents; some have a high pain threshold while others have a low pain threshold. A patient's grasp of anatomy is not always precise; hence patients should demonstrate the location of the pain, not just stating where it hurts. Once the physician has obtained a clear description of the pain, it is necessary to find out more about the patient's personality and her activities in order to correlate the pain to the disability about which the patient is complaining. Most patients do not come because of the pain but because of the disability it produces. During the history, the physician should distinguish between referred pain (from the pelvic or abdominal viscera) and root pain. Pain referred to the back is unrelated to spinal movements, to posture or to coughing, while pain arising in the spine is influenced by posture and movements, and the sudden stress of coughing.

The history also helps to differentiate between mechanical causes such as a torn ligament and a spinal pathology such as an infection. Most cases of backache are of mechanical origin. Mechanical pains are usually worse on activities such as bending, coughing or sudden movements, and better with rest or a back support. In contrast, non mechanical backache is unrelated to time or physical activity. Non mechanical pain is not relieved by rest and may be worse at night. The non mechanical causes of low back pain are those in which the bones themselves are affected by tumour, infection or metabolic disease. Along with the characteristics of the pain, the age of the patient is of some help in the diagnosis. Mechanical cause such as prolapsed intervertebral disc most frequently occurs in the 30-50 year age group. Ankylosing spondylitis (inflammation in the joints of the back) is more common in younger people (15-25 year group).

Physical Examination

Jayson (1983) stresses the importance of the physical examination for patients presenting with back pain for the first time, and patients with recurrent mechanical problems in whom the character of symptoms change which could indicate a new reason for the back pain. Macnab (1977) points out that examination of the back should be conducted in an orderly predetermined manner so that all possible physical findings may be evaluated. In other words, examination of the patient should not be directed solely at eliciting signs of a specific disease suggested by the history. Examination of the back should include 1) posture, 2) movements, 3) palpation and 4) neurological examination.

- Posture: The normal person stands upright with a lumbar lordosis (curve of the lumbar curve) and a slight forward curvature of the dorsal spine. A fixed kyphosis (inability to alter the curve of the back by standing up) is a sign of the inflammatory problem ankylosing spondylitis. A sharp angular kyphosis indicates localised disease such as vertebral collapse. The posture of the lower limbs and leg length should be checked; a real or apparent shortening of one of the lower limbs may produce a tilt in the pelvis and a scoliosis (abnormal sideways curve of the spine) which may be a cause of premature degenerative changes in the back.

- Movements of the spine include forward flexion, extension, lateral flexion and lateral rotation. During the examination of movements, the physician should observe any specific abnormalities such as limitations of the degree of movement. For example, in forward flexion, the normal lordosis flattens, then flexes until the spine is a smooth curve from sacrum (portion of the spinal column near the lower end) to occiput (lower part of the head when it emerges into the neck). In case of ankylosing spondylitis there is a localised loss of flexion since the spine has becomes more rigid. In the case of a prolapsed intervertebral disc, forward flexion will not be possible but extension and rotation will be relatively free.

- Palpation: The physician should look for tenderness which may indicate the source of an underlying pathology. For example, extreme tenderness in a specific area may indicate the site of a fracture, a tumour or an infected abscess, or osteoporosis.

- A neurological examination includes a straight leg raising test (SLR), and checking for muscle power and sensation. In the SLR test, the patient lies supine (on her back) with both lower limbs extended and the physician elevates each in turn by raising the heel. The normal person can achieve straight leg raising of about 80 degrees or more. Patients with prolapsed intervertebral disc and sciatica may only tolerate 10 degrees. Muscle power is examined in the lower limbs; the distribution of weakness may suggest which nerve root is involved. Sensory examination includes testing for light touch and pin prick sensitivity.

The examination that has been described above is confined only to the back. Jayson (1983) makes mention of a general examination that should be part of the physical. The general examination should include the abdomen as back pain may represent pain referred from abdominal or pelvic disorders.

Investigations

Investigations that the physician may order include blood tests, X-rays, and tissue typing. In cases of mechanical backache or root pain, blood tests and radiographs are of little or no value. Blood test should serve to reassure the physician that 'nothing else is going on'; the radiographs may or may not show degenerative changes due to mechanical problems. However, these tests are useful screens for spinal pathology. For instance, elevation of ESR (erythrocyte sedimentation rate) or PV (plasma viscosity) may indicate some inflammatory disorder or a tumour.

Abnormalities of the serum calcium or alkaline phosphatase may suggest bone disease such as osteomalacia. In ankylosing spondylitis, radiological changes are seen in the sacro-iliac joint and elsewhere in the spine. X-rays and bone scan may confirm and localise the site of spinal pathology, but are only 60% to 70% accurate in distinguishing between tumour and infection. In this case, histological and bacteriological examination of the affected tissue from a biopsy is best to help in making a diagnosis. Some diseases such as ankylosing spondylitis, or Reiter's disease are associated with an abnormal tissue type. For instance, one type of white blood cell known as HLA B27 occurs in about 95% of ankylosing spondylitics but only 5 % to 8% of the normal population. (Though this is rarely a useful test e.g. if 99% of the population does not have Reiter's disease, then the majority of people HLA B27 positive will *not* have that disease).

5.6 Summary

In this chapter, design considerations for a developmental student model called DEMEREST were put forward. The idea of a developmental student model for medical diagnosis was introduced. Specifically, the developmental student model aims to maintain a representation of the student's knowledge from the student point of view and not from the expert's. The features of the model are i) the focus on the form of medical reasoning, particularly, on the reasoning strategies applied during the diagnostic process and ii) the development of these strategies from novice to experienced physicians.

The design of DEMEREST was described. The tasks of the system are 1) to diagnose the reasoning strategies the physician has used, 2) to identify development of these strategies that will help the system in determining the physician's level of expertise and 3) to produce a plan corresponding to the applications of these strategies. The planning modelling approach to build the system was then discussed. Planning has been proposed as a means to decompose medical problem solving into a set of goals; the goals being associated with the strategies. By taking this approach, medical reasoning is viewed as a planning process. In further sections, the role of the system within an intelligent medical tutor was examined, and the medical domain which is used for the system was presented.

Before describing how DEMEREST was implemented, two features of the system need to be researched: 1) the reasoning strategies which the system will model, and 2) the development of these reasoning strategies. Chapter six investigates the reasoning strategies that medical students apply, and chapters seven and eight examine the development of the strategies from an empirical study and the modelling of the development of these strategies.

Chapter Six

REASONING STRATEGIES

In the previous chapter, some design considerations for DEMEREST have been discussed and its architecture described. One issue remains to be discussed, that is, the reasoning strategies that the system contains. In this chapter, these strategies are examined. The concept of reasoning strategies in the context of medical diagnosis is first discussed and then reasoning strategies used by students which were identified in the medical problem solving literature are reported. The strategies need to be described in detail if one wants to have the system recognise and diagnose the reasoning strategies used by a physician (whether novice or experienced) during a consultation. The descriptions of the strategies stem from this investigation in the medical problem solving literature. However, the literature does not always provide specific information about these strategies. Hence, the descriptions of the strategies have been refined and complemented through discussions with a medical doctor. During a consultation, more than one strategy is applied and thus a number of possible interactions between strategies is also discussed. The strategies considered in this research have similarities with those used in other medical AI systems which are examined at the end of the chapter.

6.1 Concept of reasoning strategy

Medical problem solving refers to processes by which physicians make medical diagnoses. As reviewed in chapter three, medical problem solving has been characterised in several ways and various models have been proposed. From this review it was also found that the medical reasoning process is decomposable into its form and its content. The specific knowledge relating to a medical problem clearly plays a role. In the present research, the focus has been on the form of medical reasoning, that is, on the reasoning strategies used to reach a diagnosis.

In the context of medical diagnosis, a reasoning strategy is used to refine the details of the patient case and to generate one or more hypotheses which correspond to a diagnosis. A medical reasoning strategy is related to how one makes inferences between findings (e.g. signs, test results) and diseases. In using a medical reasoning strategy, the physician makes a decision about what move to make in the current state. This decision describes a choice between two or more actions and the move is based on the physician's knowledge.

In the research reported in this thesis, a distinction is made between reasoning processes such as forward or backward reasoning and reasoning strategies such as generalisation or hypothesis generation. Forward and backward reasoning are concerned with the direction in which to conduct the search through the space, in this case of the domain of back problems, either top down or bottom up. Patel and Groen (1986) have studied the reasoning processes of expert cardiologists in terms of forward and backward reasoning. Their results showed that experts with accurate diagnoses used bottom-up forward reasoning whereas experts with inaccurate diagnoses used at least some top-down backward chaining. In contrast to forward and backward reasoning, reasoning strategies result in a search space (e.g. possible hypotheses for a back pain problem) and reflect the degree of specificity of the solution i.e. choice of a hypothesis for medical diagnosis. As Clancey pointed out (1986), a strategy "reasons" about operators and problem solving methods. Physicians do not use strategies randomly. There is some logic behind each choice (i.e. the strategy applied) which describes a line of reasoning in diagnosing the patient case. Applying this analysis to DEMEREST, concepts such as questions to the patient, clinical and laboratory tests correspond to the operators, and problem solving involves applying some strategy for manipulating these concepts.

6.2 Medical students' reasoning strategies

In order to build a model of development of strategies, our starting point was to examine in the literature the strategies used by medical students. It is worth mentioning here that experienced physicians also apply these strategies (and probably other ones as well): the expert strategies built in NEOMYCIN (Clancey 1985) are similar to the ones used in this research and thus provide evidence that novice and expert physicians use these strategies (this point is discussed further in section 6.5). However, the medical problem solving literature does not provide much insight into the kinds of reasoning strategies medical students possess. Most of the research work on medical reasoning has studied what expert physicians do (as reviewed in chapter three); and it has been shown that students use the same general form of reasoning as experts but with less powerful and organised domain knowledge (Gale & Marsden 1983, Feltovitch et al 1984a). The research reported in the thesis shows that students and more experienced physicians combine reasoning strategies in different ways and therefore do not necessarily use the same form of reasoning.

A number of observations were found in the medical problem solving literature from which a set of reasoning strategies that medical students apply was derived. Barrows and Tamblyn (1980) have examined changes in student reasoning associated with learning. They reported that medical students in their first year are not inhibited by specific medical knowledge and as a result tend to generate hypotheses which are relatively non specific. Hence their hypotheses are too *general.* As they increase their medical knowledge, they tend to use the opposite strategy, that is, their hypotheses are too *specific.* The hypotheses formulated by students are inclined to be too confined. This can be a problem as hypotheses which are too specific become less adaptable to change as new data appears and often will prevent the student from realising that there are other data that would suggest alternative hypotheses. As the knowledge of the medical students increases, they are able to obtain data that can support or weaken hypotheses. That is, they can *confirm* or *eliminate* their hypotheses.

Medical students are trained to take a complete history and perform a complete physical examination so as not to miss something important. To do so, students use routine protocols. Traditionally, the routine format of the medical record presented below has served as a guide for students to gather information about the patient (Gale and Marsden 1984).

1) patient's presenting complaint
2) history of the present complaint
3) symptomatic survey (by system)
    a) cardiovascular
    b) respiratory
    c) locomotor
    d) genito-urinary
    e) gastro-intestinal
    f) neurological
    g) general

4) past medical history
5) family history
6) social history
7) drug survey

Barrows and Tamblyn (1980) also found that students use comprehensive assessments. When presented with a patient case, students (more often than experienced physicians) make use of routine protocols to *refine* the patient's problem. For example, the student may use a routine protocol for the pain and asks the patient about all its characteristics i.e. severity, duration, episodic/continue, location, radiation, associated symptoms and relieving/aggravating factors. In addition to protocol based refinement, students (less often than experienced physicians) may use experience based refinement which involves refining an observation to another one e.g. stiffness to morning stiffness. For example, the student may have learnt that the distinction between morning and evening stiffness is important because it supports different diagnoses. Both kinds of refinements are not hypothesis driven, that is, no hypothesis is generated.

Elstein et al (1978) found that medical students (as well as more experienced physicians) applied hypothesis generation strategy whereby clues, i.e. observations, are used to generate hypotheses. Elstein's model of the diagnostic process is one of hypothesis generation and testing. However, the definition of hypothesis generation as used in this research does not include the 'testing' of the hypothesis generated. The testing phase of a hypothesis corresponds to a confirmation or elimination of the hypothesis.

In the first two years of medical schools, students learn a lot about *anatomy, pathology and physiology*. Thus, as a result, they use reasoning strategies which incorporate these kinds of knowledge. Lesgold (1984) has showed that novice radiologists (first and second years) try to explain X-ray film features by using their anatomical knowledge, whereas intermediates (third and fourth years) use anatomical and pathophysiological knowledge. Likewise, Ramsden et al (1988) have observed that fourth year medical students employ pathophysiological reasoning. In theory, one might use physiological and pathological strategies separately, but in practice it is most common to find these two strategies combine into one, namely, pathophysiological.

In some instances, students may attempt to make a diagnosis by relating the present patient case to a previous case (or a set of cases) that they had encountered. This kind of reasoning is referred to as *case based*. Patel and Frederiksen (1984) have examined the recall of written patient cases by medical students (as well as more experienced doctors) for the formation of case understanding. A number of typical and atypical cases was presented to the students and results showed that students make far fewer inferences from these cases than expert physicians.

In their studies, Ramsden et al (1988) observed that fourth year medical students use a *pattern matching* strategy. In using this strategy, the student selects a diagnosis which is associated with one or more selected clinical features. The student ignores the rest of the clinical features, or simply believes that they fit the selected diagnosis. The selected diagnosis matches a disease already known to the student i.e. the pattern and thus the diagnosis is accepted. Similarly, Barrows and Tamblyn also found that students have the tendency to force the patient's problem to fit a pattern the student has learned about a disease. In the same line of reasoning, they also noticed what they called 'eureka thinking' in which the student has recognised in the patient case a symptom or a sign for a particular disease and tries to prove that the patient has it without questioning other alternatives.

This list of medical students' reasoning strategies found in the literature is not exhaustive and is not meant to be since the interest has been to identify an initial set of strategies used by students. Some of the reasoning strategies that have been reported in the literature need to be considered with caution. This is because these studies make use of a number of methodologies, each of which constrain possible research outcomes. Section 3.2.5 has discussed the variety of methods to study medical reasoning and the methodology dependency of the results. Hence, interpretation and conclusions reached in some of these studies might need further clarification. For instance, in the study of Ramsden et al, different strategies used by students are suggested. However, no connections between these strategies are proposed. Moreover, as discussed in section 3.2.4 the interpretations and conclusions reached in this study need further explanations.

Medical students do not always apply a strategy properly. Incorrect use of strategy by students was reported in some cases. For example, Elstein et al mentioned that students tend to generate hypotheses that are either too general or too specific and thus they *overgeneralise* or *overspecialise*. The space of reasoning strategies includes strategies used by medical students as well as more experienced physicians. It is hypothesised here that experts and students use their own strategies as well as sharing common strategies. There is no claim that novices' strategies are uniquely applied by medical students. In some cases, intermediates and more experienced physicians may use the same reasoning strategies as novices.

Reasoning strategies can be classified as *domain dependent* and *domain independent* strategies. The former category includes strategies directly related to the medical domain, while the latter category contains strategies that can be used to solve other kinds of problem solving (e.g. diagnosis of a faulty circuit) as well as a medical problem. From the list of strategies gathered from the literature, domain dependent strategies are anatomical, pathological and physiological strategies, while the rest of the strategies are domain independent. This classification may be useful in determining whether in the development of reasoning strategies physicians rely more on domain dependent or domain independent strategies.

6.3 Definitions of reasoning strategies

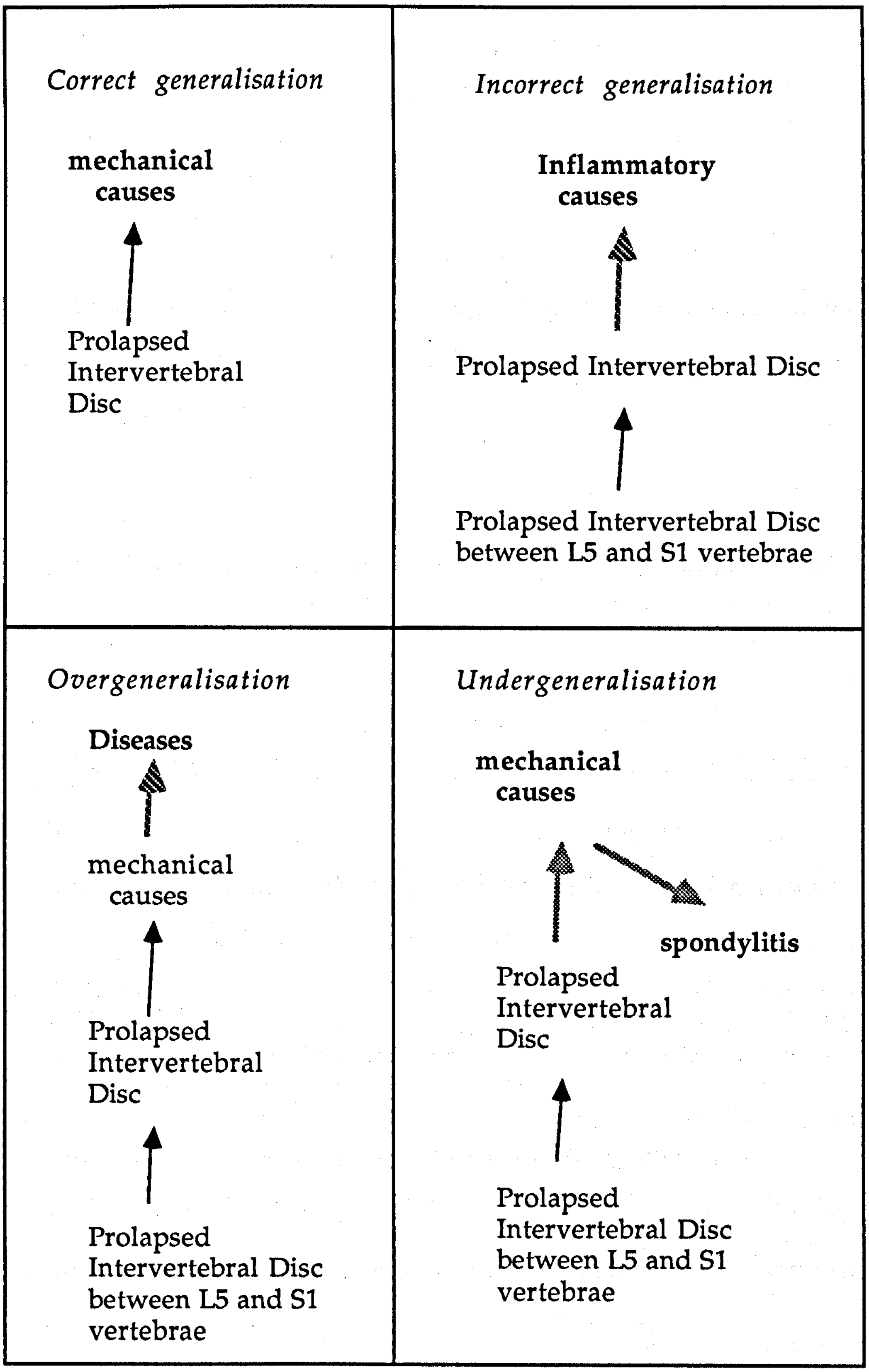
The strategies (reported in the previous section) which are believed to be identifiable from verbal protocols were selected to build the system. Furthermore, one was looking for a set of strategies that could account for the whole physician's protocol (i.e. consultation with a patient). That is, strategies that could carry out transformations such as clinical data to other clinical data, data to hypothesis and hypothesis to another hypothesis. Case based strategy was rejected because relating a present patient case to a previous one may be rephrased in terms of other strategies (e.g. hypothesis generation and confirmation) and hence may be viewed as a kind of meta strategy. Pathological and physiological strategies were ignored because anatomical strategy was already selected. In addition, it is difficult to separate anatomical and physiological knowledge and the anatomically based strategy selected here also attempts to combine physiological knowledge.

In this section, each of these strategies is elaborated. In order to recognise and diagnose the reasoning strategies used by the physician (novice or experienced) during a consultation, one needs to describe in detail the properties and features of these strategies. The descriptions of these strategies were refined and complemented through discussions with a medical doctor. It should be pointed out that the reasoning strategies described in this section need to be represented in a formal way if one wants to identify them in protocols (see chapter seven) and to implement them in the system (see chapter nine). The formal description of these strategies is postponed until chapter seven since it is most relevant to the coding of protocols. The examples given as illustrations in the following sections are all hypothetical.

6.3.1 Generalisation

The generalisation strategy is used to generalise a hypothesis (In the case of medical diagnosis, the hypothesis is a disease). Hence, one generates from a subclass of diseases. Generalisation may occur in the following cases (see figure 6.1): 1) correct generalisation, 2) incorrect generalisation, 3) overgeneralisation and 4) undergeneralisation. The last three kinds of generalisation are incorrect uses of the strategy.

The arrows in the figure 6.1 indicate the direction through which the strategy is being applied i.e. from one specific hypothesis to a general hypothesis. That link is a hierarchical link joining two diseases in the knowledge base of diseases. In all the cases the reasoning is bottom up.



→ correct step
▨→ incorrect step
░→ missing step

Figure 6.1: Generalisation Strategy

- **Correct generalisation:** the student knows the diagnosis and abstracts it into a general form. A correct generalisation will be *prolapsed intervertebral disc (PID)* is *a mechanical problem of the back.*

- **Incorrect generalisation:** the student abstracts her diagnosis into an incorrect general form. For example, integrating *PID* to *an inflammatory cause* is an incorrect generalisation; likewise by saying that *'all mechanical problems are related to PID'.*

- **Overgeneralisation:** the student abstracts her diagnosis to higher level than necessary. An example of this is to integrate *mechanical causes* to *diseases.* There are far too many diseases that one can look up.

- **Undergeneralisation:** the student is not able to abstract her diagnosis to a more general form. For example, the student has made her diagnosis as being *PID L5/S1,* and she cannot get to *spondylitis* (which may also be a reasonable cause) because she cannot generalise to *mechanical causes.*

6.3.2 Specialisation

A specialisation strategy is used to specify a subclass of hypotheses. This strategy is the opposite of the generalisation strategy. Thus, one generates a subclass disease from a general class of diseases. Specialisation may occur in the following cases (see figure 6.2): 1) correct specialisation, 2) incorrect specialisation, 3) overspecialisation and 4) underspecialisation. As with generalisation, the last three kinds of specialisation are incorrect uses of the strategy.

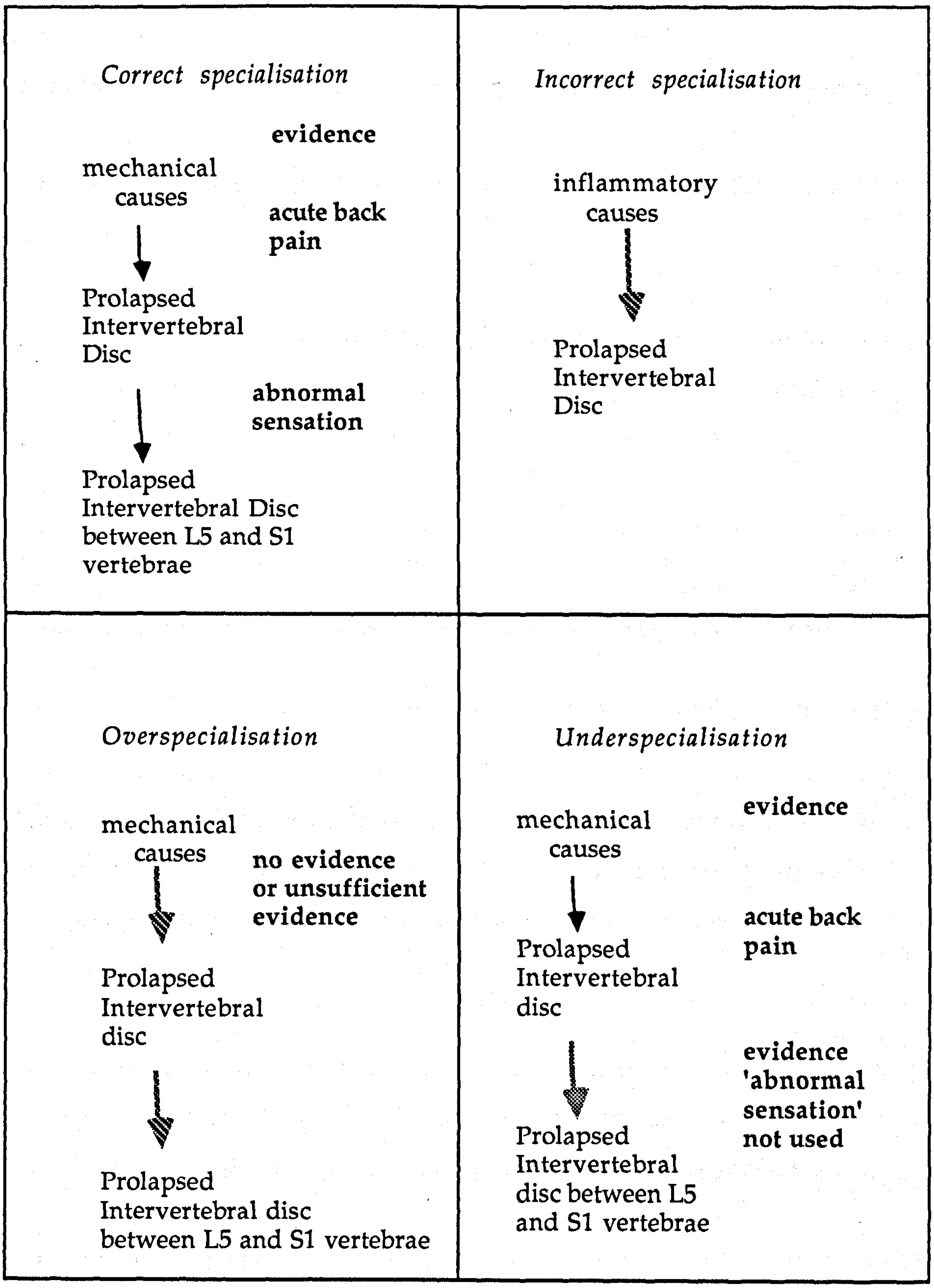


Figure 6.2: Specialisation Strategy

Once more the arrows in the figure 6.4 indicate the direction through which the strategy is being applied i.e. from one general hypothesis to a more specific one and that link between the two hypotheses is a hierarchical link within the knowledge base of diseases. In all the cases, the reasoning is top down.

- **Correct specialisation:** the student goes through the necessary inferences to reach a specific diagnosis, using correct evidence. For instance, from *mechanical* to *PID* and from *PID* to *PID L5/S1* with the evidence of *acute back pain* and *abnormal sensation.*

- **Incorrect specialisation:** the student has accessed a wrong subclass. For example, going from *inflammatory cause* to *PID* is an incorrect specialisation.

- **Overspecialisation:** the student jumps too quickly to a conclusion; her diagnosis is too specific given that the student has considered no evidence (or insufficient evidence or incorrect evidence) to reach her hypothesis. For example, the student with not enough evidence specialise from *mechanical causes* to *PID* and from *PID* to *PID L5/S1.*

- **Underspecialisation:** the student has some evidence to make the correct diagnosis but is not able to use it. For instance, the student has evidence of *acute back pain* and *abnormal sensation;* she is able to specialise from *mechanical problems* to *PID* but not able to use the second evidence to go from *PID* to *PID L5/S1.*

6.3.3 Confirmation

The confirmation strategy is used to attempt to confirm a hypothesis, that is, to validate that hypothesis as the diagnosis. Confirmation may occur after data gathering, physical examination, or results of investigation have been obtained. Confirmation of a hypothesis is possible at a general level e.g. *mechanical causes* or at a more specific one e.g. *PID L5/S1*. Confirmation may occur in the following cases (see figure 6.3):

- **Confirmation of a single hypothesis:** there is only one hypothesis under consideration (in the differential) and therefore the confirmation is of this hypothesis only. For example, the student may have considered *PID* only as the diagnosis for the patient case.

Confirming a hypothesis is related to the type of evidence available to the physician. There are different types of evidence such as positive, sufficient, necessary (Miller 1975):

- **Confirmation by positive evidence:** a hypothesis may be confirmed by positive evidence, that is, by findings which are characteristics of the disease. In the absence of sufficient evidence, one relies on positive evidence. For example, positive evidence for *PID* includes *pain in lumbar region* and *pain radiates down the lower limb.*

- **Confirmation by sufficient evidence:** a hypothesis may be confirmed if sufficient evidence is provided. For example, the finding from a myelogram test (X-ray of the disc) of *an obstruction to the flow of dye opposite a disc space* is sufficient evidence for confirming *PID L5/S1*.

Confirmation of a single hypothesis Differential Diagnosis: Prolapsed intervertebral disc	*Confirmation by positive evidence* Differential diagnosis: Prolapsed intervertebral disc Positive evidence: pain in lumbar region pain radiates down lower limb
Confirmation by sufficient evidence Differential diagnosis: Prolapsed intervertebral disc between L5 and S1 vertebrae Sufficient evidence: result of myelogram test	*Confirmation by necessary evidence* Differential diagnosis: gyne problems Necessary evidence: subject is female

Figure 6.3: Confirmation Strategy

- **Confirmation by necessary evidence:** a hypothesis may be confirmed by ensuring that necessary evidence is present. That is, the prerequisite for confirming a hypothesis is that necessary evidence is true. For instance, the prerequisite for *gynaecological problems* is that the patient be *female.*

A confirmation strategy is applied incorrectly when the evidence to support the confirmation is erroneous. For example, trying to confirm *PID* with the evidence that *the pain does not radiate* is incorrect.

6.3.4 Elimination

The elimination strategy is used to attempt to rule out a hypothesis, that is removing it from the set of hypotheses currently under consideration (see figure 6.4). It is the opposite of the confirmation strategy. Elimination may occur either at an early stage of the diagnostic process (e.g. elimination of *inflammatory diseases* as a cause of back pain given the patient's past history), or at a later stage when the physician changes her differential diagnosis to another class of problem (e.g. from *mechanical* to *inflammatory problems).* Usually, a hypothesis is eliminated after the physician has gathered additional information, performed a physical examination, or test results have been checked. Elimination may occur in the case that the hypothesis to eliminate is substantially less likely than the others. Similarly to confirmation, the ruling out of a hypothesis is related to the type of evidence available to the physician (Miller 1975, Patil et al 1982):

- **Elimination by insufficient evidence:** there is not enough evidence to support the hypothesis. For example, if the patient has *pain in the lumbar region* but the *pain does not radiate* then *PID* is less likely.

- **Elimination by negative evidence:** negative evidence can disconfirm a hypothesis, and can be obtained by asking questions about the absence of findings. An example of negative evidence for a *L5/S1 disc prolapsed* is that *bony tenderness* is not present.

- **Elimination by absence of necessary evidence:** lack of necessary evidence can eliminate a hypothesis. For example, if the subject is not *female* then *gynaecological problems* are ruled out.

Elimination by insufficient evidence	*Elimination by negative evidence*	*Elimination by absence of necessary evidence*
Differential diagnosis: Prolapsed intervertebral disc	Differential diagnosis: Disc prolapse lesion between L5 and S1 vertebrae	Differential diagnosis: gyne problem
Insufficient evidence: pain in lumbar region mild pain	Negative evidence: bony tenderness is not present	Necessary evidence: female

Figure 6.4: Elimination Strategy

Similarly to confirmation, an elimination strategy is applied incorrectly when the evidence to support the elimination is erroneous. For instance, trying to rule out *PID* with the evidence that *the pain radiates* is incorrect.

6.3.5 Problem refinement

The problem refinement strategy is used to refine the patient case, that is, to get more details about it (see figure 6.5). Two types of refinement may be considered:

- **Experience based refinement** involves refining one observation (a symptom, sign, or a test result) to get another observation.

- **Protocol based refinement** involves refining one observation by using a routine protocol.

Refining an observation means to gather more information (by asking the patient questions, by examining the patient or by lab test) about the nature of that observation. Moreover, the refinement of observations is done in the absence of considering any hypothesis. Figure 6.5 shows an example of experienced based refinement. The symptom *lower back pain* is being refined as *acute lower back pain.* In this hypothetical situation, the physician would have asked about the duration of the pain and the patient stated that it started a few hours ago (if the pain occured suddenly it is an acute event, while if the patient had the pain for a long time it is chronic).

Figure 6.5 is an example of a protocol based refinement where the symptom *right sided back pain* was gathered as a result of the physician asking about the location of the pain. Asking about the location of the pain is part of a routine protocol for defining *the characteristics of the pain.*

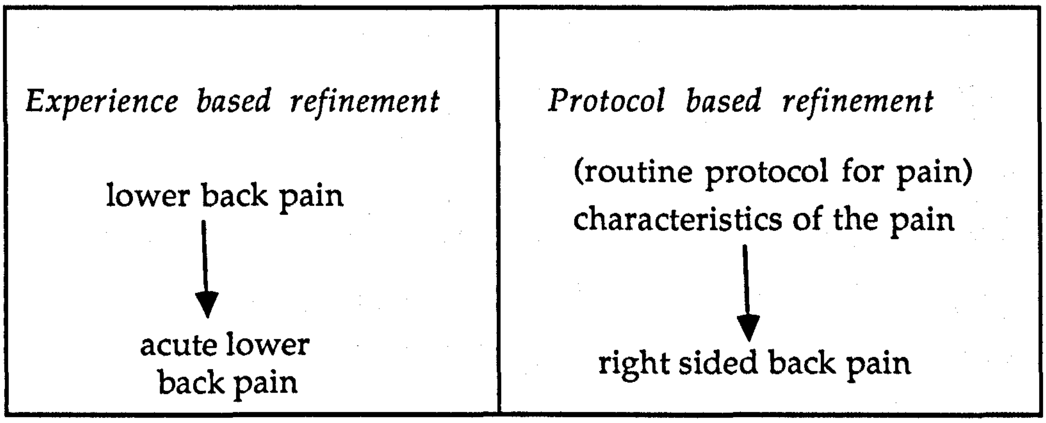


Figure 6.5: Problem refinement

The link between two observations, in the case of experienced based refinement, is not hierarchical as in the case of generalisation and specialisation. Rather it is a surface link (as opposed to a deep representational link) which links i) two entities of the same kind i.e. observation in the case of experienced based refinement and ii) an observation to a protocol routine in the case of protocol based refinement.

A problem refinement strategy is applied incorrectly when one of the observations is improper in the context in which it is used, or when the refinement from one observation to another one is vacuous. For example, trying to refine stiffness to intermittent stiffness is incorrect. One will refine stiffness to morning stiffness or to evening stiffness.

6.3.6 Hypothesis generation

This strategy is used to generate a hypothesis from one or more observations. For example (see figure 6.6), the hypothesis *PID* is being generated from the the symptom *pain in the lower back.* The link between the hypothesis and any signs and symptoms is a causal link since the hypothesis is inferred from a sign or a symptom associated with it.

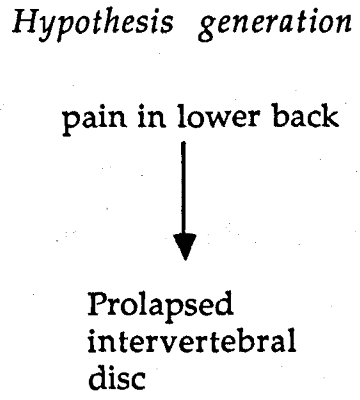


Figure 6.6: Hypothesis generation

A hypothesis generation is applied incorrectly when either the observation or the hypothesis generated is erroneous in the context in which it is used. For example, generating *sciatica* from *stiffness* is incorrect. One would generate *sciatica* from *radiation in the leg.*

6.3.7 Anatomically based strategy

Anatomical reasoning is structural reasoning. In the medical context, structural reasoning describes the structure and anatomy of each organ of the body. In this research, the focus is on the anatomy of the back e.g. structure of the spinal column, components of vertebrae, etc. By anatomical strategy, what is meant is reasoning about the anatomy of the back to explain the patient's problem. An improper use of anatomical based strategy would be to link elements of the anatomy of the back incorrectly or use elements of the anatomy of the back which are wrong.

Let us consider the following (simplified) explanation of the anatomy of (Jayson 1982): The back consists of the spinal column, the hinder parts of the rib, the wide spreading pelvic (iliac or haunch) bones, and the sacrum. The spinal column is built up of a number of bones placed one upon another, called vertebrae. The bones are covered with (thick and powerful) muscles. Between the bodies of the vertebrae lie a series of thick discs called intervertebral discs. The spinal cord which is the lower portion of the central nervous system is situated within the spinal column. In its course from the base of the skull to the lumbar region, the cord gives off a number of nerves. These nerves are issued between the vertebrae of the spine.

Now let us assume that the patient had *pain in the back* and *pain in the leg*. Given the above description of the back, an example of anatomical reasoning would be to use knowledge of the nerve supply to explain the patient's pain: the sciatic nerve supplies the area of the back of the legs. Thus, pain suffered by the patient in the back of the legs may be caused by a lesion of the sciatic nerve. Depending on which side the disc is pressing on, pain can be felt on the left leg or on the right leg. The link between the patient's symptom and the anatomical explanation is a surface link (as with problem refinement), that is, it does not convey a deep representation of the anatomy of the back.

6.4 Reasoning strategies interactions

While the literature provides evidence of the kinds of strategies medical students apply, there are no reports as to how these strategies interact. It is hypothesised in the thesis that in most cases, students (and probably more experienced physicians as well) do not use reasoning strategies in an independent manner. That is, students apply one strategy which interacts with other strategies. Moreover, one might ask about the frequency of reasoning strategy interactions i.e. how often an interaction occurs during a consultation. The identifications in the empirical data of the hypothesised interactions will be discussed in section 8.1.2. These interactions are described below:

- **Generalisation followed by specialisation.** Once the student has classified the problem category of the patient case (generalisation strategy), she may need to narrow down to one or more hypotheses that will explain the patient case (specialisation). For example, the patient complaint may be classified by the student as a *mechanical problem*. From this, the student generates a set of hypotheses (e.g. *PID)* related to *mechanical problem* which can explain the patient case.

- **Generalisation followed by elimination.** The student may eliminate a class of hypotheses (e.g. *mechanical, inflammatory)* which corresponds to a general hypothesis obtained through a generalisation strategy.

- **Problem refinement followed by specialisation.** Problem refinement involves obtaining more detailed information and hence it implies a higher level of specificity. The student may have selected to refine an observation (e.g. signs, symptoms, or test results). She can then use this new information as evidence to subclassify a hypothesis. For example, the student may refine the symptom *back pain* to *acute back pain,* which provides evidence to specialise from *mechanical problems* to *disc prolapsed.*

- **Specialisation followed by elimination.** In order to eliminate a hypothesis, the student should have enough evidence for the hypothesis to be ruled out and the hypothesis to be ruled specific. For example, if the student knows that there is no *bony tenderness,* she then can rule out the specific hypothesis *PID L5/S1.*

- **Specialisation followed by confirmation.** Similarly to the previous interaction, in order to confirm a hypothesis, the student should have enough evidence, and the hypothesis to be confirmed specific. For instance, if the student is told that the result of a myelogram test (X-ray of the disc) shows *an obstruction to the flow of dye opposite a disc space,* then the student can confirm the specific hypothesis *PID L5/S1.*

6.5 Related research

This section i) discusses how the strategies considered in DEMEREST have been labelled differently in some other research work and ii) examines the similarities between the strategies in DEMEREST and other medical AI systems such as NEOMYCIN.

The strategies discussed in the previous section correspond to strategies used by medical students. Evidence from the medical problem solving literature indicates that experienced physicians apply some of these strategies such as specialisation, confirmation, elimination, and problem refinement. However, these strategies are not always referred to by the same names. The strategy specialisation has been labelled as a refinement strategy (Patil et al 1982) and is defined as

> "The refinement strategy ... used to split a hypothesis about a general class of diseases into more specific hypotheses" (p.217).

The use of confirmation and elimination strategies by experienced doctors has been reported in (Miller 1975). However, Miller's classification of these two strategies is more elaborate than that presented in this thesis. For example, he not only differentiates between different evidence associated with confirmation but also between direct and indirect confirmation. Direct confirmation uses direct methods such as use of expert witnesses and prima facie evidence. Expert witnesses are those doctors who have observed the patient at some time in the past and have evaluated their medical status. Also, the doctor has reason to trust the conclusions of previous doctors who saw the patient. Prima facie direct confirmation is restricted to physical examination and laboratory findings. Indirect confirmation is based on the use of finding-disease association rules; each of these rules is of the form "if a collection of findings is found then the clinical condition is considered confirmed". The refinement strategy, referred to as the explore strategy in (Miller 1975), has been described in a more specific way i.e. as

> "checking for additional problems in the presence of a hypothesis structure that is already sufficient to explain the known findings" (p.217).

The difference with the definition used in the thesis and Miller's is that in the latter it is assumed that already some hypotheses have been generated. By using this strategy the physician is trying to make a complete diagnosis, that is, not to miss any other diseases that the patient may have.

Other research work on medical A.I. has used strategies similar to the ones considered in the research reported in this thesis. The system NEOMYCIN developed by Clancey (1985) models a physician's diagnostic reasoning and contains physicians' strategies which were acquired from protocol analysis. However, the strategies in NEOMYCIN are *expert* strategies. NEOMYCIN and DEMEREST incorporate strategic knowledge and domain knowledge. Domain knowledge in both systems consists of factual knowledge and the causal relations between factual knowledge, expressed as domain rules in NEOMYCIN and as a set of facts in DEMEREST. Regarding the strategic knowledge, NEOMYCIN has a strategic meta-level and hence captures a higher level of abstraction than DEMEREST. That is, strategies in NEOMYCIN are represented as a network of tasks. Each task has meta-rules associated with it which are used to order and select the applications of the domain rules. Each task which corresponds to a strategy is more elaborate than in DEMEREST. As will be examined below, it was found that the descriptions of the strategies used in this research are 'components' of some of these tasks. All the strategies but two were found in the strategic knowledge of NEOMYCIN. This gives evidence that these strategies are applied by students as well as more experienced doctors. The two outsider strategies are elimination and anatomical reasoning. Regarding the elimination of hypothesis in NEOMYCIN, Clancey states that

> "disconfirming a hypothesis involves discovering that required and highly probable findings - causal precursors or effects - are missing. NEOMYCIN's domain lacks this kind of certainty. Therefore, the program does not use a ruleout strategy." (1985, p.74).

The strategic knowledge in NEOMYCIN corresponds to a meta-level of reasoning, and thus does not include anatomical knowledge.

Generalisation

The concept of looking for more general hypotheses of the hypothesis considered (what Clancey referred to as 'ancestors' of the hypothesis) is found in one of the meta-rules of the task GROUP-AND-DIFFERENTIATE. This tasks attempts to establish the disorder categories that should be explored.

Specialisation

The concept of looking for more specific hypotheses from a more general one (referred to as 'child' of the hypothesis by Clancey) is found in one of the meta-rules of the task EXPLORE-AND-REFINE. This task chooses a focus hypothesis from the differential.

Confirmation

The confirmation of a hypothesis is the role of the task TEST-HYPOTHESIS. Confirmation is based upon findings that trigger the hypothesis (as in DEMEREST) or causal precursors to the disease.

Problem refinement

Seeking more information about a finding is the role of the task CLARIFY-FINDING. Two kinds of questions can be asked: specification questions (e.g. specifying from medication which drugs the patient had taken) and process questions (e.g. from the finding headache, the program will ask when did it start).

Hypothesis generation

Generating a hypothesis from a finding is the role of the task PROCESS-FINDING. Findings are applied to conclude about activated hypotheses.

The strategies used in NEOMYCIN were re-utilised in another system ProHC (Prolog Heuristic Classification) developed by Park, Tan and Wilkins (1989). The idea behind ProHC was to design a new representation for NEOMYCIN. ProHC is a refinement of HERACLES with the NEOMYCIN knowledge base, based on the experience gained in developing ODYSSEUS (Wilkins 1988) the apprenticeship learning program for HERACLES. ProHC can support multiple tasks such as problem solving and learning. The strategic knowledge to diagnose a patient in ProHC is represented by knowledge source and task. Each knowledge source models one of the expert's diagnostic reasoning methods, and a task is called by a knowledge source to execute an action part of the knowledge source. Knowledge sources and tasks contain similar strategies used in NEOMYCIN and in DEMEREST. As discussed in chapter three, other medical A.I. systems (e.g. PIP) which have been developed prior to NEOMYCIN do not make explicit a physician's reasoning strategies. Rather, these programs usually use other methods of inference combined with mathematical scoring schemes to reach a conclusion about diagnosis.

6.6 Summary

From medical problem solving sources, a set of reasoning strategies applied by students during the diagnostic process was identified and refined to be used in the research. The characteristics of these strategies were described in detail and related research which have used similar reasoning strategies discussed. The next chapter reports an empirical study which was carried out to examine the use of these reasoning strategies at various levels of expertise.

The seven strategies are summarised below. Throughout the rest of this thesis, abbreviations (put in parentheses) as well as their full names will be used to refer to the strategies:

- *Generalisation (GEN):* generating a general hypothesis from a specific one (e.g. mechanical cause of back pain from prolapsed intervertebral disc).

- *Specialisation (SPEC):* generating a specific hypothesis from a general one (e.g. disc prolapsed from mechanical cause).

- *Confirmation (CONF):* validating a hypothesis based on evidence and including it in the differential diagnosis (e.g. tending to confirm disc prolapsed if there is tenderness).

- *Elimination (ELIM):* ruling out a hypothesis based on evidence and removing it from the differential diagnosis (e.g. eliminating disc prolapsed if there is no bony tenderness).

- *Problem refinement (PREF):* refining the problem presented by the patient by gathering more details (e.g. refine pain to acute pain, or, refine patient case by asking about social history).

- *Hypothesis generation (HGN):* generating one hypothesis from a symptom, signs or test results (e.g. disc prolapsed from pain in lower back).

- *Anatomical (ANAT):* generating an information (which may or may not contain a hypothesis) using anatomical knowledge (e.g. pain in the leg is related to the activity of the sciatic nerve).

Chapter Seven

EMPIRICAL STUDY OF DEVELOPMENT OF REASONING STRATEGIES: Methodology

As examined in the previous chapter, a set of reasoning strategies was identified in the medical problem solving literature. Although the literature describes on these strategies, no evidence of the development of these strategies over time i.e. from novice to intermediate to expert physicians was found. This chapter reports on an empirical study which was carried out to investigate the development of medical reasoning processes (Alpay 1990a). Specifically, the intentions of this study were: 1) to identify the reasoning strategies collected from the medical problem solving literature which are applied during the medical diagnostic process and 2) to examine the development of these strategies at various levels of expertise. The first section of this chapter provides the scale of medical expertise that has been considered for the study. Then, the subjects taking part in the experiment are presented. Methodological considerations are described in a further section.

7.1 Scale of medical expertise

The scale of medical expertise considered in this research is based upon levels of hierarchy in the medical profession. The following levels were used.

- *Medical students*

Pre-clinical: The first two years of medical school are pre-clinical. Students study subject matters such as anatomy, physiology, biochemistry. They do not have hospital training or contact with real patients.

(The 3rd year is optional for the award of a Bachelor of Science degree.)

Clinical: The next three years are clinical, at the end of which students can qualify as doctors if they pass an examination. During that period in hospital, students learn by participating in the diagnosis and treatment of patients. Students get their training primarily by working with patients rather than through lectures and laboratory work.

- *House Officer (HO)*

After finishing their undergraduate studies, students must spend one year prior to registering as a doctor. The year, spent in hospital, is divided into six months of surgery and six months of internal medicine.

Training to be a specialist:

- *Senior House Officer (SHO)*

One can spend two years in hospital as an SHO. It is during that period that one starts to specialise in an area of medicine e.g. orthopaedics.

- *Registrar*

Following a senior house officer post, a doctor typically spends three years as a registrar. This is the period during which most experience is gained in the speciality of interest.

- *Senior Registrar*

After holding a registrar's position, a doctor typically spends between 2 to 5 years as a senior registrar.

- *Consultant*

A doctor training as a specialist finally becomes a consultant, usually after holding the post of senior registrar.

Training to be a GP:

- *Trainee in general practice (GP-T)*

After two years as an SHO, a doctor may choose to train as a *general practitioner* (GP) rather than specialise. If so, she must have done two years as an SHO in a range of specialities, usually four e.g. obstetrics and gynaecology, paediatrics, casualty and psychiatry and additionally spend a year in general practice.

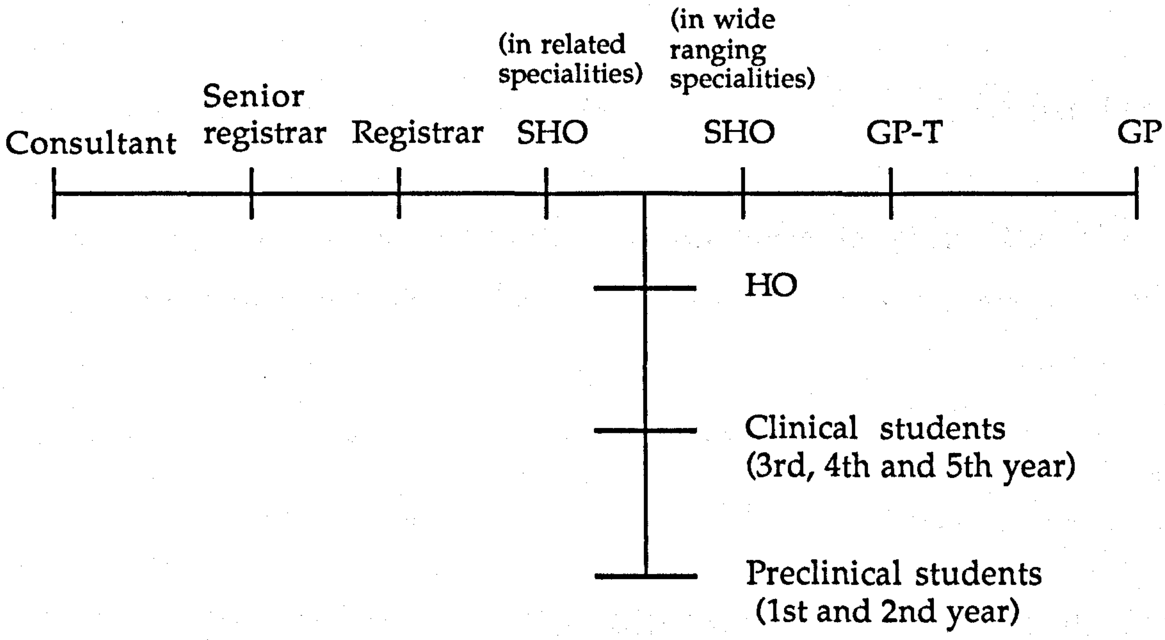


Figure 7.1: Scale of medical expertise

This scale of medical expertise is not linear. For instance, a senior house officer is not at a higher level of expertise than a general practitioner. It is true that an SHO will know more in her domain of speciality than a GP.

However, both have developed medical diagnosis skills. Figure 7.1 illustrates how this scale of expertise is viewed: preclinical students (1st and 2nd year) are at the bottom of the scale, followed by clinical students (3rd to 5th year), in clinical *training*. Next, come the house officers starting their clinical *experience*. Senior house officers, registrars, senior registrars and consultants are in the specialised group, whereas general practitioners and GP trainees are in general practice group. Both groups are on the same line of expertise, since it is considered that each group is 'expert' in its own medical field, that is, specialisation or general practice.

7.2 Subjects

The scope of this research made it impossible to observe physicians during their entire skill acquisition process. The approach chosen instead has been to observe physicians at different levels of performance i.e. novice, intermediate and expert. Ten subjects (5 males and 5 females) took part in the experiment[1], each having a different level of expertise in medical diagnosis in general and in back pain problems in particular. The distribution of subjects from novice to expert clinicians is as follows:

> Two **3rd year students** (1st year clinical) with no knowledge or clinical training of back pain problems.
>
> One **4th year student** (2nd year clinical) who related back pain problems to neurology and kidney diseases, and with no clinical training in orthopaedics.

---

[1] Nine subjects participated in the experiment, and an additional subject was interviewed at a later stage to be included in the testing of the system (see chapter 10).

Two **5th year students** (3rd year clinical) with some knowledge but no clinical training of back pain problems. They had been attached to units that specialise in internal medicine and gynaecology.

One **house officer** with some knowledge of back pain and one year of clinical experience (but not of back problems). As a student she had spent time in units that specialise in orthopaedics, internal medicine and gynaecology.

One **senior house officer** in orthopaedics with three years of general clinical experience as well as six months specialist experience in orthopaedics.

One **general practitioner trainee** with 10 years of previous clinical experience. None in orthopaedics.

One **general practitioner** with 13 years of experience in general practice. No specialist experience in orthopaedics.

One **consultant** in orthopaedics with 16 years of experience, a large part of which was specialist experience.

The medical students were in clinical practice in Milton Keynes General Hospital. There is no medical school in Milton Keynes and consequently students came from medical schools outside the Milton Keynes area: one 3rd year was from Bath Medical School, the other 3rd year from Barts Medical College (London); the 4th year student was from West Germany and the 5th year students were from Leeds Medical School. The general practitioner and the trainee were from Hanslope surgery (in the Milton Keynes area) and the rest of the physicians were practising in Milton Keynes General Hospital.

Pre-clinical students were not considered for the experiment for two reasons: firstly they know very little about medicine; secondly and more importantly the experiment involves the subject being in a consultation with a patient and pre-clinical students do not have any practice with real patient consultations. Other groups of physicians such as registrars and senior registrars were ignored for logistical reasons. That is, it was difficult to find a registrar or a senior registrar available in the hospital over the period that the interviews were carried out.

7.3 Methodology

The methodology of the study is described in this section. Physicians at different levels of expertise were put into a consultation with a simulated patient[2], followed by a post interview session between the subject and the experimenter. The verbal protocols were transcribed and coded. Prior to the main experiment, a pilot study was carried out to help in the setting of the experiment. This section is organised as follows: first, the patient case used in the study is presented, followed by a description of the experiment. Coding of the verbal protocols are then detailed and assessed.

---

[2] A simulated patient is a person who simulates a patient.

7.3.1 The patient case

Details of the patient case were extracted from a real consultation[3]. The patient, Mrs. A.F. is a 42 year old lady presenting with right sided back pain. A summary of the patient case can be found in appendix A1. This summary reports on three consultations, each of them including a diagnosis. The three consultations are spread over a period of time and it is at the end of the third consultation that the patient was diagnosed accurately and treated for the problem she had. Each consultation contains details about the presenting problem, the patient's history, the results of physical examination, the results of investigations (if any), a diagnosis and the patient's management (if any). General information about the patient and her previous past medical history are given in the first consultation.

The three diagnoses that were made over a period of two years were as follows: 1) non specific mechanical low back pain, 2) 90% psychological overlay, 10% non specific low back pain and 3) prolapsed intervertebral disc. The patient case may have looked simple given the chief complaint but in fact it contained confusing characteristics such as the patient's underlying depression.

7.3.2 Interviews

The experiment was divided into two parts, a think aloud session and a post interview session. Instructions provided to the subjects are found in appendix A2.

---

[3] Summary of the patient case was passed on to me by my external supervisor Dr. Mike O'Neil.

- In the **think aloud** session, subjects were asked to think aloud through their reasoning processes and verbalise what they were doing. The setting of the think aloud session was a patient consultation. Subjects were put into a consultation with a simulated patient and were asked to diagnose the patient. Subjects were not permitted to perform a physical examination, but physical findings were provided by an observer on the subject's request (e.g. if the subject asked the result of a SLR test she would be given just this result). Subjects were allowed about 15 minutes to complete the task. At the beginning of this session, subjects were given the instruction sheets and then the observer introduced the patient to the subject as "Mrs. A.F. is a 42 year old lady who presents with back pain" and from there the consultation between the patient and the doctor started. The session was to cover just one consultation. Thus, only the first consultation reported in the summary of the patient case (see appendix A1) was considered for the experiment. The subsequent consultations were included in the summary for the reader's own information.

- In the **post interview** session, subjects were asked to explain and clarify their decisions made during the think aloud session. The think aloud session was replayed to help subjects recall what they had said.

The simulated patient had not been given a fixed script. Rather, she learned about the patient case and was given information about the patient and a set of prepared answers. In case of unexpected questions to which the simulated patient did not have any ready made answer, she had been instructed to respond by the negative or to say that she did not know.

In the remainder of this section, the choice of the methodology for this experiment is discussed. First, lessons learned from a pilot study are examined, followed by a discussion on the methodology of the study.

Lessons of methodology from a pilot study

A pilot study was first conducted to help in the setting of the experiment, and highlight any problem of methodology. Moreover, other purposes of this pilot study were to define categories for the coding of protocols.

Subjects participating in the pilot study

Two subjects from Hanslope surgery (Milton Keynes area) took part in this preliminary study: a trainee in general practice with 10 years of previous clinical practice[4], and a GP with about 10 years of experience in general practice.

Patient cases for the pilot study

Two real patient cases were selected. The first patient case was of a 35 year old woman presenting with back pain. Her back pain problem was an awkward presentation of a gynaecological problem called endometriosis. The second patient case was of a 70 year old man also presenting with back pain. He had osteoporosis (which is the weakening of the bones).

Methodology of the pilot study

Subjects were put into a consultation with a simulated patient who acted for these two patient cases. The subjects were asked to diagnose the patient. As they asked the patient questions or made decisions (e.g. to carry out investigations), they were probed to explain why a particular course of

---

[4] This subject is the same GP-T who was interviewed in the main study.

action has been taken. Subjects were not given the simulated patient to examine, but physical findings were provided by the person acting as the simulated patient, on the subject's request. After subjects had carried out consultations for the two patient cases and made their diagnoses, a discussion about the patient cases followed. The consultations and ensuing discussions were recorded. No time constraint was put for the consultation or the discussion. The two protocols on tape were transcribed, providing the scripts of the subject and of the patient.

Problems in the design of the pilot study

There are several problems with the design of the pilot study which include the following:

1) The simulated patient was also the interviewer and is a medical doctor himself. This caused a confusion of roles. Each subject knew the interviewer well and hence the interview between the subject and simulated and on occasions the consultations became, more like an interview between two medical doctors than a consultation between a patient and a doctor.

2) The form of introspection adopted was not satisfactory, as subjects were often interrupted by the interviewer.

3) For each patient case the interview included more than one consultation with the patient. In the real life situation of general practice, when the patient comes in to see a doctor, she is often sent to undertake some tests and is given treatment until she comes back to see the doctor. This means that in some cases, more than one consultation is required for the doctor to have all the information needed to make a final diagnosis. However, during the first consultation the physician has usually made an

initial diagnosis for the patient case. In an experimental setting, mixing more than one consultation can be confusing in identifying the physician's reasoning processes for a single consultation.

4) The interview included the two patient cases and the discussion part took over one hour. Although no time constraint was given at the beginning of the interview, it was felt that the time taken was not close to a real diagnostic situation and a time constraint for the consultation should be introduced. Typically, the average time of a consultation is of 5 to 10 minutes for general practice and 15 minutes to half an hour for an hospital consultation. The time allocated for the simulated consultation should be reduced to these time ranges.

The problems of the pilot study were avoided in the main study: i) the interviewer did not also play the role of the simulated patient, ii) subjects were not interrrupted after each question, iii) only one consultation was considered, iv) only one patient case was used and v) time constraints were defined.

Discussion of methodology of the main study

Some problems of methodology were discussed in section 3.2.5. The reader may recall that two objections made against using verbal protocols as data were that: i) the reporting process might alter task performance and ii) reports may yield an incomplete record of cognitive processes. Ericsson and Simon (1984) have demonstrated however that verbal protocols can be reliable data. Their main claim is that verbalisation of information is shown to affect cognitive processes only if the instructions require verbalisation of information that would not otherwise be attended to.

Some of their arguments are applied here to support the methodology chosen for this experiment.

In the pilot study reported ealier in this section, concurrent probing was used during the consultation between the subject and the patient; subjects were probed concurrently with their performance of the task for specific information. As Ericsson and Simon point out, the negative effect with this form of verbalisation is that even if the verbal data reflect the cognitive processes going on during verbalisation, they would give an inaccurate picture of the normal course of those processes such as the processes for collecting information about the patient's pain. For this reason, this form of verbalisation was not continued for the main experiment. Instead, two kinds of verbalisation were chosen for the experiment: think aloud verbalisation (another form of probing) and retrospective verbalisation in what was referred to as a post interview session.

In the think aloud verbalisation, the cognitive processes described as successive states of the attended information are verbalised directly. In other words, by asking the subject to verbalise, a direct trace of information stored in short term memory (STM) is obtained. Retrospective verbalisation is best achieved when the retrospective report is given by the subject immediately after the task has been completed. Much of the information is still in STM and can be directly reported or used as retrevial cues. This is why in the experiment reported here, the post interview session followed the think aloud session immediately. Nevertheless one should keep in mind that in retrospection, the retrieval of information from long term memory (LTM) may cause problems. The retrieval process is fallible since other memory structures may be accessed instead of those created by the just-finished cognitive process. It should be pointed out that the method selected in this thesis for the study is similar to the one used in Elstein et al's study (discussed in chapter three), that is, a think aloud session followed by a session recalling the interview.

The choice of using written protocols was excluded because of a number of limitations they present. As mentioned in section 3.2.5, questionnaires using multiple choice questions (MCQ) are useful if the experimenter knows the right answers to the questions presented to the subject. Since this is not the case in the context of our experiment, questionnaires were ignored. In contrast to questionnaires, reading a technical expository text such as a medical text is a more demanding task in terms of the cognitive processes involved, as the work of Patel and Frederiksen discussed in section 3.2.5 has shown.

In the context of medical diagnosis, a physician does not usually diagnose patients by analysing a medical text containing the patient's case. In most cases, the physician deals with the patient directly and collects patient data verbally during a consultation. It is during a consultation with a patient that the physician's abilities to conduct an interview, examine the patient, and make appropriate observations are exercised. In contrast, the format of written patient cases is unreal and abstract. As Barrows and Tamblyn (1982) point out, in using written patient cases there is no challenge to the skills of interview and examination. All the important observations are written down in the abstract, linear format of words. Another important reason for not using a text in the study is that since the aim of the study was to identify the type and sequence of strategies used by each doctor, prompting the patient on a written text removes the chance of collecting this information. It is worth mentioning that written case histories have some advantages. The main advantage is educational. It is a familiar learning format. For example, the student can study the patient problem at any time convenient for her, at any speed, for any number of times until she has understood the problem. In the framework of our experiment, using a written patient case was not appropriate, not only given the disadvantages discussed above, but also since subjects were not just medical students.

In our experiment, a simulated patient was used. Researchers in medical problem solving have made use of real as well as simulated patients in their experiments. For example, Elstein et al (1978) chose simulated patients, whereas Gale (1983) preferred to use real patients. There are some reasons for <u>not</u> choosing a real patient. Barrows and Tamblyn (1982) mentioned a number of these reasons from an educational point view, but which can also be considered for an experimental setting. Firstly the patient may not be available at a particular time. In fact in our experiment, the same patient was to be interviewed by various physicians at different times and different locations (e.g. hospital and general practice). Physicians themselves have a busy schedule and are not available for an experiment at any time. Secondly, the patient may feel as though she is being used as guinea pig and would not tolerate repeated examination. Thirdly, the available patient may present complexities or unrelated problems which may be a source of confusion in the analysis of the diagnostic process.

There are of course some drawbacks in the use of simulated patients. First of all, it is an artificial and unfamiliar situation for physicians. It is probably more so for already experienced doctors than for medical students who are more exposed to these simulations as part of their medical education. In addition, it is difficult to simulate physical signs such as swollen joints or tenderness. Furthermore by using simulated patients, visual cues such as the patient looking pale may not (at all or accurrately) be simulated and reported.

As in the pilot study, the protocols (the think aloud and post interview sessions) were transcribed with the script of the subject, the patient and the observer. Pauses were not noted. Protocols of a novice subject i.e. the first 3rd year student and a more experienced physician in orthopaedics i.e. the consultant, can be found in appendix A6.

7.3.3 Coding of the verbal protocols

The results of the pilot study not only highlighted some methodological problems regarding the setting of the experiment but also helped in defining the coding categories (also referred to in this research as low level categories). A set of glossaries containing the encoding vocabulary for the low level categories was created. These glossaries were enriched by reading about back pain problems and during the analysis of the protocols of the pilot study. These glossaries can be found in appendix A4. The word lists are not complete: firstly, because there is reference to a couple of patient cases which do not contain all possible presentations of back pain and secondly, because the reading upon which the glossaries are based was not exhaustive of the whole domain. The pilot study also helped in the coding of the strategies: an analysis of the protocols used in the pilot study showed that the formalisations of the reasoning strategies were appropriate as it was possible to identify some reasoning strategies that the subjects had applied.

A consultation between a doctor and a patient contains rich data. Some of this includes observable facts i.e. what the patient or the physician says or does. Other information derives from the inferences which are made and which are not directly observable. The interest here is to link data which are directly observable in the interaction between the doctor and the patient in order to derive non observable inferences that were made. Our coding of the verbal protocols is designed to capture what happens during the consultation as well as what has occured as a result of this observable problem solving activity, that is, what reasoning strategies were applied. In the first case (referred to as a low level coding), the actual propositions are coded. In particular, this includes medical categories such as signs and symptoms. In the second case (referred to as high level coding), the reasoning strategies that are derived from the low level coding are coded. The coding procedure is described in greater detail below.

Segment encoding

Three basic units of encoding have been defined for the interactions between the patient and the doctor and for the doctor's verbalisations during the think aloud and the post interview sessions. Schematic presentations of these basic units of encoding and examples are shown after introducing the basic units of encoding and the low level coding categories.

1) The first unit of encoding is an interaction between a doctor and a patient (a question/answer or a set of questions/answers) for a particular context (e.g. location of the pain) which is treated as the basic unit of exchange of information and which is found in the consultation/think aloud session. The questions asked by the physician as well as the answers from the patient do not necessarily correspond to a single utterance; they

may also be multiple utterances. It should be pointed out that the analysis of the protocols is centred on the physician (and not the patient), since the research interest is on the physician's problem solving processes and also since the physician is usually the one who directs the consultation and hence the flow of information.

2) The second unit of encoding is also found in the consultation/think aloud session. It corresponds to a verbalisation about what the physician is doing.

3) The last encoding segment is found in the post interview session. It consists of the replay of the interaction between the doctor and the patient, and the physician's explanation for his or her questions to the patients along with any other comments she or he may want to make. In the second and third units of encoding, the physician's verbalisation can also be single or multiple utterances.

Low level coding

Within each of these three basic units of encoding, the propositions associated with the physician and the patient are coded. Low level categories include the following medical categories:

1. **Hypotheses** which correspond to diseases of back pain and any problem that the physician thinks is the cause of the patient's pain (e.g inflammatory problem).

2. **Symptoms** (e.g. pain) are subjective sensations reported by the patient, or any other information that the patient gives to the physician.

3. **Characteristics of the pain** (e.g. onset, severity, duration) are specific features that one might ask to characterise the pain in the back.

4. **Signs** (e.g. age, smoker, patient looks pale) are objective and observable by the physician.

5. **Examination tests** are physical tests and checks on the patient during physical examination. The results of examining the patient (e.g. tenderness in back) are included in category #4 (signs).

6. **Investigation tests** (e.g ESR, PR) are investigations that the physician has requested in order to get more information about the patient's condition.

7. **Investigation tests results** are the results of investigations.

8. **Treatments** correspond to the management of the patient's back pain.

9. **Results of treatments** are the patient's responses to the treatment(s) that the physician has prescribed.

10. **Differential Diagnosis** contains the current hypotheses (i.e. the working set of diagnoses) generated by the physician for the patient case.

A few remarks need to be made about these categories:

i) Characteristics of the pain are coded separately from symptoms. This is done in order to highlight the importance of the features of the pain informing a diagnosis.

ii) The differentiation between categories #1 and #10 is relevant and reflects what happens in the protocol. The subject may not only generate hypotheses which correspond to the working diagnosis for the patient case (category #10), but may also generate other hypotheses which are not considered in the differential (this point will be discussed further in section 8.1.1).

Evidence that physicians use a planning approach was discussed in section 5.3.1. In the encoding of the protocols, the position adopted is that subjects are goal oriented and that they take a planning approach during the consultation. Moreover, in constructing a plan for the subject's protocol, one needs to know with which phase of the consultation a given goal is associated. Consequently, two other categories are added to the low level coding - a planning category and a consultation category. These are:

1. **Goals** which are decisions taken during the consultation (e.g. check about location of the pain, check history of the patient, check palpation of the back).

2. **Phases of the consultation** which include history, physical examination, investigations[5].

---

[5] The treatment phase is not included because it is not part of the diagnosis of the patient but rather is part of the patient's management.

These two categories are not data directly observable from the protocol. Each goal created corresponds to the context of a doctor/patient interaction. For example, if the physician asks the patient about the location of the pain she has, the goal becomes *check location pain.* Similarly, for the phase of the consultation, the topic of the doctor/patient interaction provides information about to which phase of the consultation the interaction belongs. For example, if the physician asks the patient about her past medical history, it is part of the history phase. A list of goals that was created from the protocol analysis can be found in appendix A3, and the process of filling in the slots of the goals for inputs to the system is discussed in section 9.1.2.

As discussed in section 5.3.4, a goal may be associated with one or more strategies either as a result of the physician i) reacting to some data or ii) posing a specific question. In a protocol, it is not always easy to distinguish whether the physician is reacting to some data or whether she is posing a question. The post interview session provides some means of making this distinction by having the subject explain why she or he asked a particular question. The subject may reply, for instance, that she or he was following a protocol routine, or that she or he had a particular hypothesis in mind. However, the distinction between "reacting to data" and "posing a question" is not always clear cut and as will be explained in section 8.1.1 this may have some effect on the results obtained.

Schematic representations of the low level coding

Schematic representations of three basic units of encoding using low level coding are shown in figures 7.2 to 7.4 along with portions of actual protocols. Figure 7.2 shows a doctor/patient interaction. In the first example the 4th year student asked about the onset of the pain and in the other one the second 5th year student asked about the severity of the pain.

Schematic representation :

doctor <category> <utterance-1>.......<utterance-n>
patient<utterance-1> <category>.......<utterance-n>

Example 1:		
4th year st:	When did the pain start ?	category #3 (onset of the pain)
Patient:	It started this morning	category #3 (onset of the pain)
Example 2:		
5th year st.2:		
	Ok, what is the pain like ?	
	Was it a very sharp or dull pain	category #3 (severity of the pain)
Patient:	It was quite a sharp pain	category #3 (severity of the pain)

Figure 7.2: Schematic representation of a doctor/patient interaction in the think aloud session and portions of actual protocols

Figure 7.3 shows a verbalisation of the doctor's action during the think aloud session. In the example, the second 3rd year student explained the reasons for asking about whether the pain was constant by generating a number of hypotheses for it.

Schematic representation:

doctor: <utterance-1> <category> ....<utterance-n>

Example: The student has just asked the patient whether the pain is constant

3rd year st.2:		
	I want to know whether it is a persistant problem or an acute or chronic condition, really	category #1x3

Figure 7.3: Schematic representation of a verbalisation of the physician's action in the think aloud session and a portion of an actual protocol

Figure 7.4 shows a verbalisation of the physician's action, but one taking place during the post interview session. In the example, the trainee in general practice explains the reasons for asking about waterworks and he generated some hypotheses.

Schematic representation:

doctor: <utterance-1> <category> ......<utterance-n>

Example: Replay of a question in the think aloud session

GP-T: How is the waterwork now ?

GP-T: Again, back pain as I mentioned earlier could
be loin pain, urinary infections, and one could categories #2 & #1
find out more about whether there is any disturbance.

Figure 7.4: Schematic representation of a verbalisation of the physician's action in the post interview session and portion of an actual protocol

High level coding

Given the categories of the low level of coding, it is possible to define the reasoning strategies in terms of these categories - medical, planning and phases of the consultation. A notation of the form R(x,y,..) where R's are relations and the x,y, etc arguments can be used to formalise the strategies (Ericsson and Simon 1984). In the formalisation of the strategies, the arguments are the low level categories, and the predicates correspond to strategies. In the following, the formalisations of each of the strategies are described. Examples of encoding of excerpts of actual protocols using the schematic presentations can be found in appendix A9.

- *Generalisation:* the formal notation is

GEN(hypothesis-specific, hypothesis-general, evidence, goal, phase)

Hypothesis-general is being generated from hypothesis-specific using some evidence.

- *Specialisation:* the formal notation is

SPEC(hypothesis-general, hypothesis-specific, evidence, goal, phase)
Hypothesis-specific is being generated from hypothesis-general using some evidence.

As mentioned earlier, a subject may generate a hypothesis which corresponds to the differential or she may generate a hypothesis which is not part of the differential. However, it is not always easy to distinguish whether the hypothesis is considered as the diagnosis. Hence, in the formalisation of the strategies, the differential is applied to confirmation and elimination strategies only.

- *Confirmation:* the formal notation is

CONF(hypothesis, evidence, differential, goal, phase)
The hypothesis is being confirmed with some evidence to support the confirmation and the hypothesis is added to the differential. Knowing when a physician confirms a hypothesis is not always made explicit in the protocol i.e the subject saying "I try to confirm hypothesis x....". However, when the physician indicates that the hypothesis corresponds to a possible diagnosis, it is understood that the hypothesis is being confirmed.

- *Elimination:* the formal notation is

ELIM(hypothesis, evidence, differential, goal, phase)
The hypothesis is ruled out with some evidence to support the elimination and the hypothesis is removed from the differential. Similarly to confirmation, knowing when a physician rules out a

hypothesis is not always made explicit in the protocol and the update of the differential is also an indication of eliminating a hypothesis as a possible diagnosis for the patient case.

- *Problem Refinement:* the formal notation is

PREF(observation, observation-routine protocol,goal, phase)
The observation is refined to another observation or the observation is refined by using a routine protocol.

- *Hypothesis generation:* the formal notation is

HGN(observation, hypothesis, goal, phase)
The hypothesis is being generated from the observation which is also the evidence for the hypothesis.

- *Anatomically based:* the formal notation is

ANAT(observation, anatomical-information, goal, phase)
The anatomical information is produced from the observation.

Some further points need to be stated about the above formal notations:
i) In chapter six, correct as well as incorrect use of these strategies were described. The formal notations given above are appropriate in both cases since the arguments used are the same. In other words, no distinction is made between correct and incorrect evidence.

ii) The evidence referred in the notation can be from category #2 to #9, that is, a sign, a symptom, a characteristic of the pain, an examination test, an investigation test, a result of investigation, a treatment or a result of treatment.

iii) In the formal notations, there is no distinction for the kind of evidence that the physician has used e.g. sufficient evidence or positive evidence (see section 6.3). This is a limitation in the analysis of the data that will be discussed in section 8.1.3.

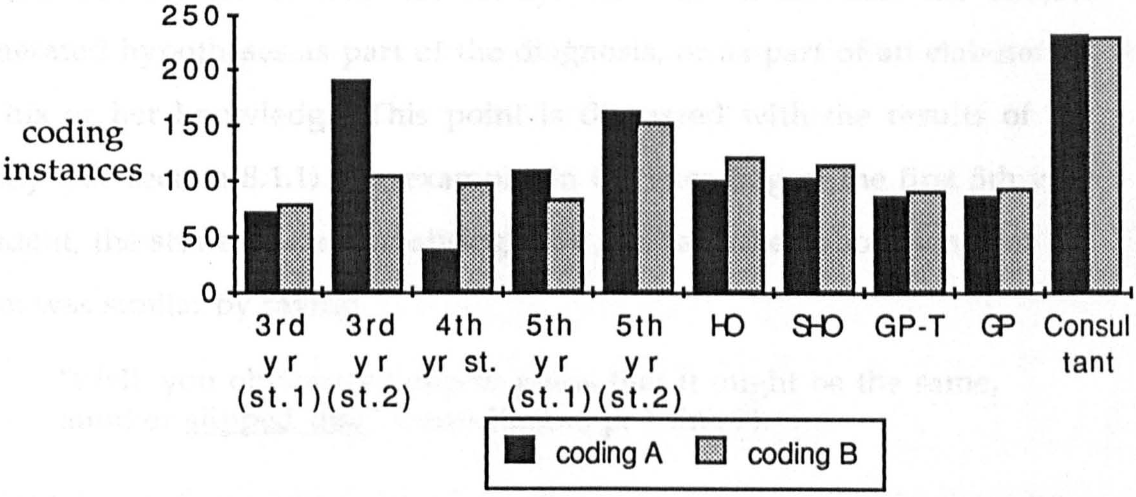
iv) It should be noted that the differential is updated via two strategies only (confirmation and elimination) as it was found in the analysis of the pilot study that maintaining the differential from protocol to implementation design is not straightforward. Another possibility would have been to update the differential every time a hypothesis is generated, but not necessarily confirmed or ruled out. However, not every hypothesis generated corresponds to the differential, as the results of the pilot and main study showed.

In the experiment, subjects may have applied strategies, either during the think aloud session, or the post interview session. The procedure to identify physicians' reasoning strategies was as follows. Each protocol (think aloud and post interview sessions) was first analysed for low level coding. Then, the analysis consisted of linking these low level categories to form reasoning strategies. Each interaction between doctor and patient for a particular context was analysed for high level coding. Any verbalisations (in the think aloud session) of the subject for that context, were also coded for reasoning strategies. This time, in the post interview session, verbalisations for the same context were also analysed. If the doctor repeated himself or herself in the post interview session, and as a result the same strategy was applied in both sessions then only the strategy in the think aloud part was considered. In some instances, the links of low level categories to form strategies were not found within a single unit of encoding, but rather by linking low level categories of a doctor/patient interaction with the corresponding verbalisation either in the think aloud or post interview session.

7.3.4 Consistency of the protocol categories

Low level categories were assessed for consistency by an independent encoder who was not medically trained. She was asked to identify the medical categories, and the phases of the consultation, but did not assess the goal category[6] and the reasoning strategies as it was thought to go beyond her activity as an independent encoder (instructions to the independent assessor are found in appendix A5). The independent assessor was given the glossaries mentioned in the previous section which contain the coding categories.

Figure 7.5: Coding per protocol



The chart of figure 7.5 shows the number of instances coded per protocol (the think aloud and post interview sessions) by myself (coding A) and the independent judge (coding B). Two major differences appear in the

---

[6] The goals were assessed for consistency at a later stage i.e. in the testing of the system (see chapter 10).

protocols of the second 3rd year student and of the 4th year student. This result is due to the fact that in the case of the 3rd year student I encoded more categories than the independent assessor (86 vs 38), and in the case of the 4th year student it was the other way around (18 vs 30). For the rest of the protocols, about the same quantity of categories were encoded.

The differences between my coding of the protocols and the independent assessor's coding were based on the following:

- Encoding variations

The assessor and myself sometimes did not encode the same categories, and sources of variation include:

1) Coding of the categories #1 and #10, that is, hypotheses and hypotheses in the differential. It was not always clear as to whether the subject generated hypotheses as part of the diagnosis, or as part of an elaboration of his or her knowledge. This point is discussed with the results of the study (see section 8.1.1). For example, in the encoding of the first 5th year student, the student was verbalising why she has asked about whether the pain was similar by saying

> "Well, you obviously jump to guess that it might be the same, another slipped disc" (consultation, p. 1 line 9).

Slipped disc was encoded by myself as a hypothesis for the differential (#10), whereas the other assessor coded it as category #1, hypothesis.

2) New instances not found in the glossaries which led to confusion of coding. This confusion probably stems from the fact that both encoders were not medically trained. Difficulty arose in distinguising between the

categories #1, #3 or #4 (i.e. hypotheses, signs and symptoms) in the coding of the house officer's protocol. The HO said

> "Basically the things which are running through my head are: Is the pain actually in the back or is it in the abdomen`" (consultation, p.2).

The underlined section was encoded by the independent assessor as a hypothesis (#10), whereas it was encoded by myself twice as characteristic of the pain (#3) for the location of the pain.

- Finding the categories

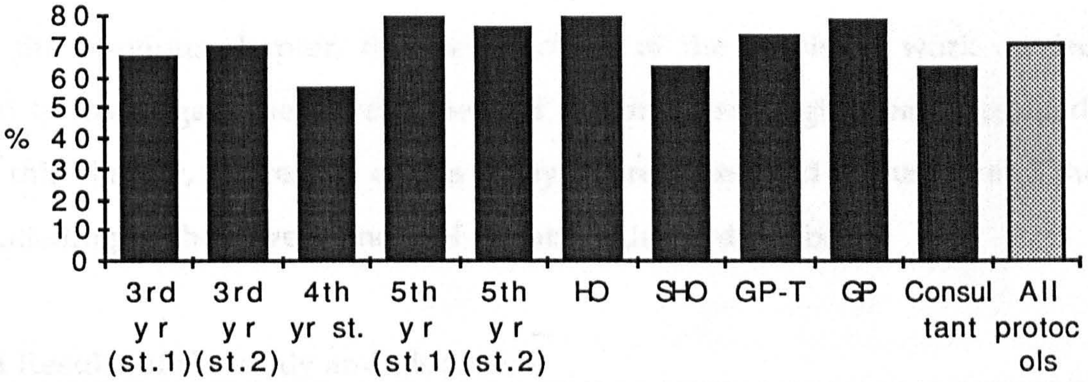
1) On occasions, the assessor and myself sometimes missed an encoding.

2) Coding more of the categories #1,#10, #4 and #5, that is, hypotheses, hypotheses in the differential, signs, and examination tests. I usually encoded the above categories more often than the assessor. A possible explanation for this is that the assessor had less exposure to medical terms than myself.

Overall there was a 71% agreement of the encoding of the low level categories with the independent assessor (see figure 7.6). This result shows that these low level categories were not purely subjective and also corroborated (to an extent) by an independent assessor. For each protocol, the percentage of agreement with the other encoder was calculated by adding the total of instances that I coded with the total of instances that the independent assessor coded, and then dividing this total of coding of instances by the total number of instances coded differently resulting in the percentage of differences. The percentage of agreement was then calculated by subtracting the percentage of differences from 100%.

This degree of agreement between the independent assessor and myself was judged to be sufficiently high to allow the coding to be used in the construction of the model of development of the reasoning strategies and its implementation.

Figure 7.6: Percentage of agreement of encoding



7.6 Summary

This chapter has reported on the methodology of an empirical study which was carried out in order to examine the development of the reasoning strategies used in medical diagnosis. Physicians at different levels of expertise were put into a consultation with a simulated patient. This was followed by a post interview session between the subject and the experimenter. The verbal protocol were transcribed and coded, and the low level coding categories were assessed by an independent assessor.

In the next chapter, the results of the study are reported and discussed, and the modelling of interactions of these reasoning strategies at different levels of expertise is described.

Chapter Eight

DEVELOPMENT OF REASONING STRATEGIES

Results of the Study and Modelling

In the previous chapter, the methodology of the empirical work carried out to investigate the development of reasoning strategies was presented. In this chapter, the results of this study are reported and discussed and the modelling of the development of the strategies is described.

8.1 Results of the study and discussion

The first part of this section reports on a quantitative analysis of the applied reasoning strategies, while in the second part, interactions between reasoning strategies identified at different levels of expertise are discussed. Finally, some limitations of the empirical study are examined.

8.1.1 Applications of reasoning strategies

A first step in the analysis of the protocols was to examine the reasoning strategies applied by the subjects quantitatively. At this point, it should be pointed out that subjects did not verbalise a great deal in the think aloud session and that the post interview session was very useful in providing additional insight necessary to identify which reasoning strategies were used.

Distribution of strategies accross subjects

Figure 8.1 shows the distribution of strategies. Three groups of strategy distribution were found by taking the total number of times a strategy was applied: 1) generalisation, confirmation and anatomically based, 2) specialisation, elimination and problem refinement and 3) hypothesis generation.

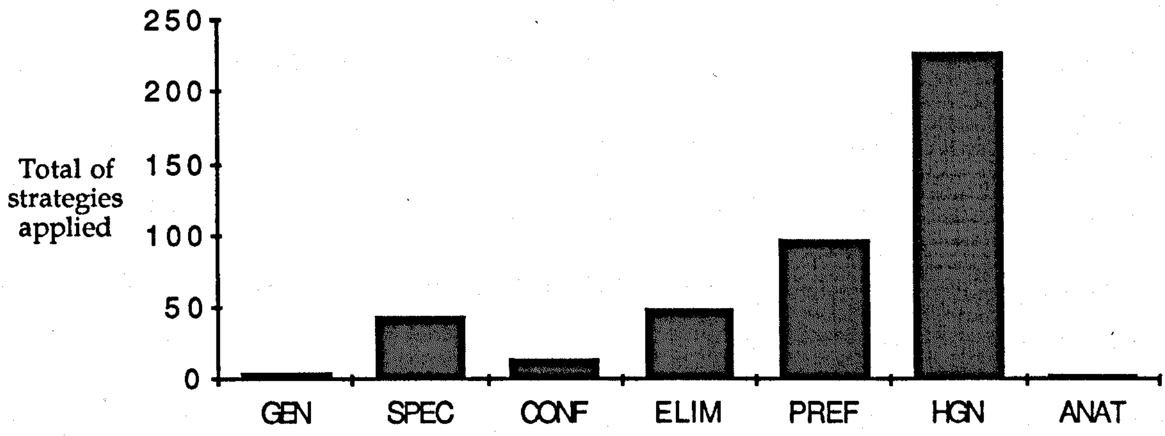
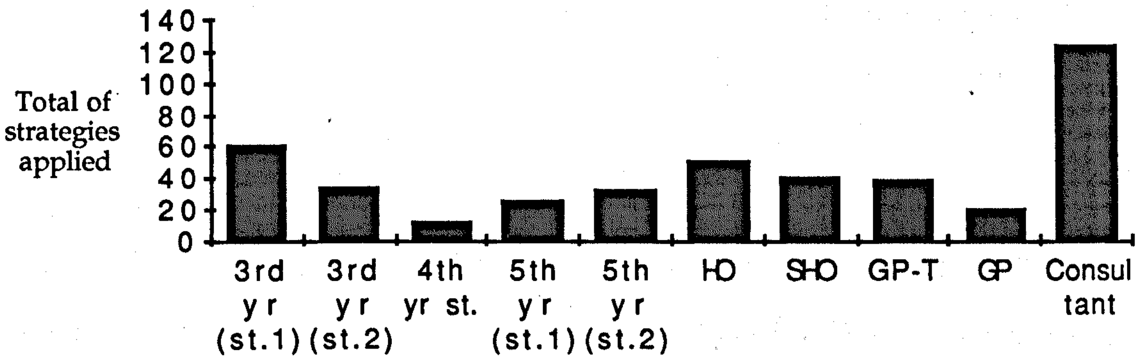


Figure 8.1: Number of strategies observed

Distribution of reasoning strategies per subject

Figure 8.2 shows the distribution of reasoning strategies per subject. A first significant result is that the consultant applied the greatest number of reasoning strategies. It would be premature to conclude that as one becomes more experienced one applied more reasoning strategies. A possible explanation for this result is that the consultant mentioned his plan of action which included not only the first consultation with the patient but also subsequent ones (although the consultant was told to consider only one consultation with the patient). Another possible explanation is that, the consultant has usually more time for the patient and more resource available to him (for investigations and treatment) than, say, the GP. The low number of strategies applied by the GP can be explained by the fact that the GP usually has limited time to see a patient (about 10 minutes).

Figure 8.2: Distribution of strategies per subject



A third possibility to explain the difference beween the consultant and the GP is related to the fact that subjects' protocols contain very rich data. Two distinct kinds of data were found in the protocols:

i) *diagnostic data* which seem to be used to diagnose the problem
ii) *performance data* which seem to correspond to how much knowledge a subject had, but not necessarily in relation to the present case.

There was not always a clear distinction between these two kinds of information. A lot of performance data is found in the expert's protocol, whereas more diagnostic data is present in the GP and GP-T's protocols. The post interview protocol of the consultant contains more information which is not directly related to the current patient case. For example, the consultant went on to explain the importance of the patient's age in diagnosing back pain by saying (see Appendix A6 p.330):

> "In a 20 year old there are disc problems but when you hit your 40s and 50s the commonest is degenerative changes, so when I say disc problems I say prolapsed intervertebral disc problems. 40s and 50s you are talking about degenerative changes that involve sagging discs not actually prolapsed ones, ones you would call backache rather than sciatica and degenerative - ah joint problems. And in the older group 60s and 70s then clearly tumour, secondaries come top of the list."

The consultant not only mentioned the hypotheses one considers when the patient is in her forties but also if the patient is in her 20's, 50's 60's and 70's. In the above paragraph, a number of hypothesis generation strategies was applied from the performance data. In contrast to the consultant, the GP did not mention explicitly the age when diagnosing the patient's back pain. The presence of these two kinds of data (diagnostic and performance data) in the protocols implies that in this experiment physicians applied reasoning strategies for the purpose of diagnosis, as well as for the purpose of elaborating on their knowledge about the domain.

Timing of the consultation

The difference in strategies applied by the consultant and the GP may also be related to the time each physician took over the consultation. The time constraint of the consultation was 15 minutes. The GP as well as other experienced physicians (i.e. house officer, senior house officer and GP trainee) perform below the given time (11<time<15). The time taken by the expert was the highest (time=19) and can be explained by the fact that the expert not only put himself into a simulated consultation, but also felt that his role was to express the totality of his knowledge. Medical students perform on time or slightly above that time (15<time<18).

Frequency of applying the strategies

The following paragraphs discuss how each strategy was applied by the subjects. The number in parentheses refer to the number of times a particular strategy was used.

Generalisation strategy

Generalisation was applied only a few times - at the intermediate level i.e. house officer (1), at expert level (3), not at all at the student's level.

Specialisation strategy

Specialisation was used almost at all levels of expertise, with slightly more concentration at the 3rd year student level (13) and at the expert level (11).

Confirmation strategy

Confirmation was applied more from the SHO level and above (12) and a few times at the student levels (3). There are remarkably few examples of confirmation being applied. One possible explanation for this result is related to the way confirmation was defined and extracted from the protocol. As mentioned in section 7.3.3, it is not always easy to know when a physician is trying to confirm a hypothesis. The position adopted in the study is that confirmation is said to be applied when the subject makes it explicit or indicates that he or she is trying to confirm a hypothesis. In the protocols collected, relatively few subjects verbalised their confirmation process which may explain the low number of confirmations.

Another plausible explanation for the low number of confirmations applied is related to the way a strategy is linked to a goal (see section 5.3.4) - by posing a question or reacting to data. In posing a question, it appears that the physician, in the process of generating a hypothesis, aims to try to confirm (or in fact rule out) the hypothesis. In contrast, when reacting to some data, the physician may generate a hypothesis, and then try to confirm or rule it out. For example, the physician wants to confirm *renal infection* and hence will ask if the patient has *right sided back pain* (case of posing a question). Alternatively, the physician is told that the patient has *right back pain* which brings to mind *renal infection* The physician may then try to confirm the hypothesis (case of reacting to data). Methodologically, the distinction between reacting to data and posing a question was blurred. Hence, it seems that, during the analysis of the protocols, more strategies were identified as the physician reacting to some data rather than posing a question. This may explain why more hypothesis generation strategies than confirmation strategies were found. In fact, some of the hypothesis generation strategies may have been confirmation strategies.

Elimination strategy

The same problem of methodology discussed above applies to the elimination strategy. However, more elimination strategies were found in the protocols (50). A possible explanation is that subjects were able to verbalise their elimination of hypotheses more often than they did with confirmation. Elimination was mostly applied at the expert level (26), and much less at other levels (between 1 to 6 - a total of 24). Although the consultant explicitly explained that his way of reasoning was to diagnose by exclusion (post interview session - "It's a diagnosis by exclusion", see appendix A6 p.333), one might question whether diagnosis by exclusion is purely the expert's style or whether it is related to his level of expertise. In addition, the clinical setting of the expert coupled with the nature of the medical domain of orthopaedics may favour the use of this strategy: a back pain problem may have a large number of different causes and it is not always possible to come to a definite diagnosis.

Problem refinement strategy

The application of the problem refinement strategy was found at all levels of expertise. It was applied most at the 3rd year student level (between 13 to 22 - a total of 35) and intermediate level (24), but less at the 4th and 5th year student and experienced levels (between 2 to 12 - a total of 38). Students applied protocol based refinement as well as experience based refinement (see chapter six). The #2 3rd year student and the #2 5th year student used both forms with a significant emphasis on the protocol based refinement. In contrast, the other students used mainly experience based refinement, although they all pointed out that if they had had more time and were in a real situation, they would have asked the patient more routine questions, that is, more questions based on routine protocols.

It should be pointed out that the routine protocols that students used were not only back pain related. Students are taught to ask a full protocol of questions about areas not related to the problem in hand to ensure they do not miss any important cues (see chapter six). Indeed, some students reported that they used routine protocols not to 'miss anything out'. The question raised here is whether this is developmentally important. Is it a good practice for students to apply by rote routine protocols ? Students learn to apply a pre-defined protocol and gradually select part of the protocol which is useful for a specific consultation. The routine protocol for characteristics of the pain is a good illustration of this point. All subjects asked questions about the features of the pain, but students tended to ask more in a rote manner than did the SHO or the GP.

Interestingly, the HO also used routine protocols during the consultation and hence applied problem refinement a great number of times. However, she seemed to have used routine protocols in a different way than the #2 3rd year student and the #2 5th year student: the students used routine protocols since that was the best way they knew to handle a consultation, whereas the HO used routine protocols because she had to handle a new patient and hence asked the patient some routine questions. Subjects were not told whether they already knew the patient. The HO assumed that she did not know the patient; the GP-T, the SHO, and the consultant alike ignored this issue; and the GP assumed that he had never met the patient before but still did not ask the patient routine questions. This may be an indication that experienced clinicians are more focussed in their approach.

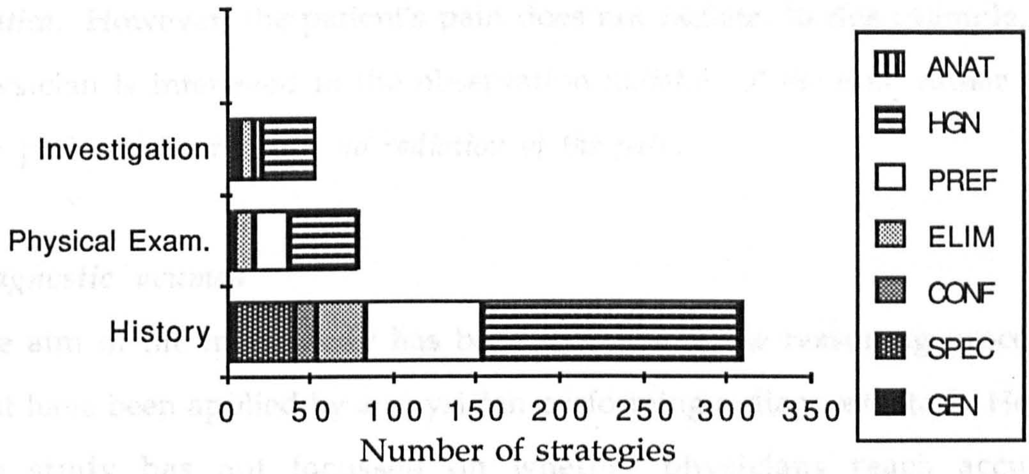
Hypothesis generation strategy

Hypothesis generation was the strategy applied most across subjects and phases of the consultation. It was applied most at the expert level (69) and much less at the other levels (between 7 to 33 - a total of 159). This result is not surprising and supports the evidence in the literature that novice and expert physicians alike use a hypothesis generation strategy. However, the context in which the strategy is applied varies among subjects. For example, all the students applied HGN during the history phase of the consultation but for different reasons: 3rd and 4th year students applied hypothesis generation for symptoms related to pain, social history and past medical history. More advanced students (5th year) broadened this application of hypothesis generation to other symptoms such as urinary infection symptoms and symptoms related to kidney problems.

Anatomically based strategy

Anatomically based strategy was applied at the 3rd and 4th year students' level (1, 3) but not at the other level. The anatomically based strategy was applied only a very few times (4). One expected that novice physicians like the medical student would rely on anatomical information to carry out a diagnosis. In fact, what seems to have happened is that students rarely verbalised that kind of knowledge in their protocols.

Figure 8.3: Distribution of strategies across phases



Strategies and phases of the consultation

Figure 8.3 shows the frequency of applying the strategies according to the phases of the consultation. Most of the reasoning strategies are applied during the history phase and then during the physical examination. This result was expected since history is the phase of the consultation where most clinical reasoning takes place. Moreover, this supports Jayson (1983)'s finding:

> "a careful history will usually indicate the nature of the problem, and this is aided by a full physical examination." (p.160).

Patient observations versus physician observations

It was found in the protocols that, when applying a strategy, physicians used observations e.g. symptoms which were either:

i) Patient observations, that is, observations specific to the given patient case. For instance, the physician may use the patient observation *right sided back pain* to generate the hypothesis *kidney problem*.

ii) Non patient observations (referred to as physician observation), that is, observations non specific to the patient case. For instance, the physician may generate from the observation *radiation of the pain* the hypothesis *sciatica.* However, the patient's pain does not radiate. In this example, the physician is interested in the observation *radiation of the pain* rather than the patient observation *no radiation of the pain.*

Diagnostic acumen

The aim of the main study has been to examine the reasoning processes that have been applied by a physician performing a diagnostic task. Hence, the study has not focussed on whether physicians reach accurate diagnoses. It is, of course, essential that students learn to make proper diagnoses. However, in their learning process, making the wrong diagnosis may sometimes be beneficial as they can learn from the diagnostic mistake they made.

SUBJECTS	DIAGNOSES	SUBJECTS	DIAGNOSES
3rd yr st.1	disc problem	HO	musculo-skeletal
3rd yr st.2	recurrent slipped disc, trapped nerve	SHO	muscle strain
4th yr st.	recurrent disc problem	GP-T	musculo-skeletal
5th yr st.1	slipped disc, kidney infection, stone, infection	GP	muscle strain
5th yr st.2	disc problem, kidney infection	Consultant	mechanical problem, inflammatory cause, infective disorder, tumour

Table 8.1: Diagnostic Acumen

The diagnostic acumen for each subject is shown in table 8.1. The reader may recall that the first diagnosis of the patient was "non specific low back pain". All the subjects also classified the patient's problem as possibly of mechanical origin. In addition, the 5th year students and the consultant mentioned other causes of back pain than mechanical.

After the consultation with the patient, each subject was asked what was his or her diagnosis. As mentioned before, it was not always easy to find out whether or not subjects confirmed or disconfirmed a hypothesis and hence whether or not he or she considered the hypothesis as part of the differential. The approach adopted here was to record and model only the confirmations and eliminations that were indicated in the protocol. As a result, it should be noted that the subject's working diagnosis found in table 8.1 may not always be present in the differential generated by the system since the subject may not have verbalised that this hypothesis is part of the differential. However, the program will generate the subject's diagnosis as any other hypothesis (this point is explained further in the next chapter).

8.1.2 Interactions of reasoning strategies

An interesting finding of this study is that there is no evidence of monotonic development of the strategies. Monotonic development, in this context, refers to an evolution of the strategies applied at each level of expertise. Instead, it was found that strategies may be combined with one another in meaningful ways, referred to as interactions of strategies. As will be explained, these interactions of strategies reflect a development from one level of expertise to another and are not random combinations of strategies.

Models of interactions of strategies at different levels of expertise were built by splitting the data collected into two equivalent halves. The first half was used to construct the models and includes the protocols of one 3rd year student, the 4th year student, one 5th year student, the HO and the GP. The second half was used to test the models and includes the protocols of the two other 3rd and 5th year students, the SHO, the GP-T and the consultant.

One kind of interaction between strategies which was found in the protocols, is referred to as direct. In a *direct interaction,* one strategy is first applied and the information used to apply that strategy (e.g. observation or hypothesis) is then used to apply a subsequent strategy. An example of a direct interaction between HGN and SPEC is found in the protocol of the 5th year student #1: the student generated the hypothesis *infection* given the evidence that the patient had *past waterwork infections* and then specialised into *urinary infection.* In the case of a direct interaction, the strategies concerned will be applied from the same goal. In the above example, the goal was *check past waterwork infections.*

The procedure used to identify interactions of strategies was as follows: Each protocol was laid out with the sequence of interactions in the order in which the strategies have been applied. The interactions of strategies which have been hypothesised in section 6.4 were the starting point in identifying some strategy interactions. In the following section, interactions of the reasoning strategies are reported for the levels of expertise of a 3rd, a 4th and a 5th year student, of a house officer and of a general practitioner. Examples of these interactions are given to illustrate the points discussed.

Interactions in the protocol of the 3rd year student #1

The 3rd year student applied four kinds of strategies: SPEC, HGN, PREF and ANAT. Direct interactions of these strategies found in the protocol are:

- Direct interaction between HGN and SPEC.

- Direct interaction between ANAT and HGN.
- Direct interaction between PREF and HGN.

- **Direct interaction between HGN and SPEC**

This interaction implies generating from an observation a hypothesis h1, and then generating from h1 a more specific hypothesis h2. For example, the student generated from the finding *renal pain,* the hypothesis *infection* and then specialised into *descending infection.* Specialisation is learned early on by medical students. As discussed in section 6.2, students do generate hypotheses which are specific even though they may not have the appropriate cues to do so. By using this interaction, the student can narrow down a particular area of the patient's problem and pinpoint a specific diagnosis, for example, whether it is a mechanical or an inflammatory problem in the case of back pain. It also means that the student does not generate the specific hypothesis directly and goes through an additional step by applying HGN. Moreover, this interaction provides information about the level of specificity the student has reached.

One should note that the level of specificity that the student has reached may depend on the ability of the student to express herself clearly. An experienced doctor may express himself or herself more easily than a novice subject, who may blurt out *infection* though meaning or thinking of *pyelonephritis.*

- **Direct interaction between ANAT and HGN**

This interaction implies taking an observation, ob1, producing from it anatomical information and then using the same ob1 to generate from it a hypothesis h. For example, the student generated from *renal pain* anatomical information about *renal pain being an obstruction of the* *ureters* and then from *renal pain* generated the hypothesis *infection.* The 3rd year student does not have knowledge or training in orthopaedics. Therefore, the student connected back pain problems with what she knew, that is, the anatomy of the back.

- **Direct interaction between PREF and HGN**

This interaction is concerned with refining one observation ob1 into another observation ob2 or a routine protocol rp and using ob1, ob2 or pr to generate a hypothesis h. For example, the student used a protocol based refinement strategy to refine the observation *cardiorespiratory symptoms* and then used that observation to generate the hypothesis *eroding aortic aneurysm.* Refining the patient case and generating hypotheses are two aspects of the diagnostic process that students learn during their medical education and this interaction shows how these two aspects can be combined to further one's reasoning skill.

Interactions in the protocol of the 4th year student

This student applied fewer different kinds of strategies than the 3rd year student and no new interactions of strategies were detected. A similar interaction found in the 3rd year student's protocol was present in this protocol. This interaction is:

- Direct interaction between ANAT and HGN.

As an example of this, the student generated from the observation *can move legs* the anatomically based explanation that *if the disc goes straight into the spinal hole and on one of the nerves down below you won't be able to move properly.* She then generated from the observation *movements of the legs* the hypothesis *retroverted prolasped.* In section

7.2, it was mentioned that this student's background was of neurology. Her reference to the nerves is an example of this.

Interactions in the protocol of the 5th year student #1

In this student's protocol, not only were more kinds of strategies applied than in the protocols of the 3rd or 4th year students, but also more interactions were found:

- Direct interaction between HGN and ELIM.
- Direct interaction between HGN and ELIM and SPEC.
- Direct interaction between HGN SPEC CONF and ELIM.

• Direct interaction between HGN and ELIM

This interaction implies generating from an observation, ob, a hypothesis, h (the observation is evidence for generating that hypothesis). Then, the opposite of the observation, not ob, is used to rule out the hypothesis h. In other words, in order to eliminate a hypothesis h, one needs to know the necessary evidence to generate that hypothesis in the first place. For example, the student generated from the observation *flexion not ok* the hypothesis *hip problem*. Since flexion was fine, the student ruled out that hypothesis.

• Direct interaction between HGN and ELIM and SPEC

This interaction is similar to the previous interaction. In addition, a specific hypothesis h2 is generated from the hypothesis h1. It should be noted that the specific hypothesis is not eliminated. Using the same example of the previous paragraph, after ruling out the hypothesis *hip problem* the student generated a specific hypothesis *arthritis of the hip*. By eliminating a hypothesis, the student already narrows downs her space of possible diagnoses. In generating a more specific hypothesis, the student is still in the same space of diagnoses which are candidate to elimination.

- **Direct interaction between HGN and SPEC and CONF and ELIM**

This interaction implies generating from an observation a hypothesis h1 (HGN) and then generating a specific hypothesis h2 from h1 (SPEC), and finally trying to confirm or disconfirm h2 giving the evidence ob (CONF ELIM). For example, the student generated from *past waterwork infection* the hypothesis *infection* and then specialised from *infection* to *urinary infection* and then tried to confirm or rule out that hypothesis giving the evidence of *past waterwork infection*. By using this interaction, the student needs to specialise before she can confirm or disconfirm a hypothesis. Interestingly, as will be discussed in a moment, at a more experienced level (i.e. GP), this step was not required.

This interaction can be viewed as decomposed into two other interactions, HGN SPEC CONF and HGN SPEC ELIM since the student was trying to either confirm or rule out the hypothesis.

Interactions in the protocol of the house officer

At the level of the house officer, further developmental features were identified: i) interaction with the same strategy, that is, the repetition of the same strategy and ii) new interactions of strategies. These are:

- Direct interaction between PREF's.
- Direct interaction between HGN's.
- Direct interaction between HGN and SPEC and ELIM and GEN.

- **Direct interaction between PREF's**

This interaction involves generating from an observation ob1 another observation, ob2, and then using ob2 to generate ob3 and so on. For example, the house officer generated from the observation *aggravating factor,* the observation *movements limited* and then used *movements limited* to generate *bony pain.* This kind of recursive process using previous findings shows that the physician has made associations between cues for a given patient case and further refines the patient's problem.

- **Direct interaction between HGN's**

This interaction involves generating a hypothesis h1 from some observation ob1, and then using ob1 with some other observation ob2 to generate h2, and so on. In other words, observations that have been previously gathered are grouped together to generate hypotheses. For example, the HO generated the hypothesis *kidney infection* from the observations *feeling hot sweaty and sick* and *pain in the back* The observation *pain in the back* was collected ealier in the consultation and before *feeling hot sweaty and sick.* By combining more than one finding, collected at different times during the consultation, in order to generate new hypotheses, the physician forms a global picture of the patient problem and makes connections between its different aspects. It should also be pointed out that the interaction between HGN's only seems possible because the subject knows more than one cause of the symptom. Hence, a student might not use HGN HGN, not because of a problem solving skill lack but because she lacks domain knowledge.

- **Direct interaction between HGN and SPEC and ELIM and GEN**

This interaction implies generating a hypothesis, h1 from some observation (HGN) then generating a more specific hypothesis, h2 from a more general one h1 (SPEC), ruling that hypothesis out (ELIM) and then generating from h2 a general hypothesis h3 (GEN). For instance, the house officer generated from *accident of the back* the hypothesis *inflammatory condition,* then generated from *inflammatory condition* the specific hypothesis *pyelonephritis,* ruled out this hypothesis and then generalised from *pyelonenephritis* to *serious problems.* This interaction shows that although the physician has generated a specific hypothesis, she then applied a generalisation, possibly to avoid getting too specific as students tend to do (as mentioned in Barrow and Tamblyn 1980), or to follow a new line of possible diagnoses. In other words, the subject generalised because a more specific option was eliminated. It should also be noted that the interaction HGN SPEC ELIM is the reverse to the one used by the 5th year student HGN ELIM SPEC. Whereas the house officer eliminated specific hypotheses, the student eliminated general hypotheses and then specialised.

Interactions in the protocol of the general practitioner

From this protocol, two additional interactions of strategies were found:

- Direct interaction between HGN and CONF.
- Direct interaction between HGN and GEN.

• Direct interaction between HGN and CONF

This interaction implies generating from an observation, ob, a hypothesis h and then trying to confirm that hypothesis. For example, the GP generated the hypothesis *similar disc problem* from the observation *similar pain* and then confirms that hypothesis. Unlike the student level, this interaction shows that at a more experienced level, the physician confirmed a hypothesis without having to specialise first from a more general one. As will be discussed in chapter ten, this interaction was often applied at other experienced levels e.g. GP-T and consultant.

- **Direct interaction between HGN and GEN**

This interaction implies generating from an observation ob a hypothesis h1 and then generalising to another hypothesis h0. For example, the GP generated from a list of observations *(no aggravating relieving factors, similar pain, sharp pain, right sided back pain, onset this morning)* the hypothesis *acute back strain* and then generalises to *muscularskeletal.*

Interactions shared among levels of expertise

Some of these interactions were found at more than one level of expertise. For instance,

i) The interaction HGN and SPEC was found in more than one protocol (e.g. protocols of the 5th year student and the SHO).

ii) The interaction between HGN and ELIM was applied by the 5th year student as well as the house officer and the general practitioner. For example, the GP generated from the (physician) observation, *radiation in the previous back pain,* the hypothesis *sciatic radiation* and then ruled out that hypothesis since the patient had no radiation in her previous event of back pain.

iii) The interaction between HGN's was applied not only at the house officer level but also at the general practitioner level. For example, the GP used the observations that in the previous event of back pain the treatment was a *(week in bed, pain killers were given, no tests were required)* and *(the patient did not have physiotherapy).* He then added to these observations of previous treatment, observations characterising the pain such as *(sharp pain, right sided back pain, no radiation)* and *(no aggravating or relieving factors)* to generate the hypothesis *a bit of ligament strain.*

Predictions of interactions

A number of interactions of strategies were hypothesised in section 6.4. The following points examine whether these interactions were found in the protocols.

i) Two of the interactions were found in the protocols. The interactions SPEC ELIM and SPEC CONF were found at the 5th year student level.

ii) The interaction PREF SPEC was not found. However, this interaction could be decomposed into two interactions PREF HGN and HGN SPEC which were identified at the level of a 3rd year student.

iii) The interaction GEN ELIM was not found. However, the interaction ELIM GEN also combined with SPEC was found at the house officer level.

iv) The hypothesised interaction GEN SPEC was not found.

Indirect interaction of strategies

In the process of examining direct interactions of strategies, another kind of interaction was found, referred to here as *indirect interaction.* In an indirect interaction, the link between two strategies is at the immediate higher level of abstraction. For example, the 5th year student #1 applied a protocol based refinement strategy to refine *aggravating factors of the pain* and then a hypothesis generation strategy to generate the hypothesis *sciatica* from *radiation of the pain.* These two strategies are linked at the abstract level of the characteristics of the pain because they include information about the characteristics of the pain i.e. its aggravating factors and its radiation. In the case of an indirect interaction, the strategies will be applied from two different goals but each will have in common a goal at an immediate higher level of abstraction. Indirect interactions have not been fully examined in the current research work and possible research directions are discussed in further work (section 11.4).

8.1.3 Limitations of the study

Some of the limitations of the empirical work will now be discussed. They include the following:

1) The study was carried out with a small number of subjects. Each subject not only corresponds to a class of individuals but also possesses his or her individual differences. With so few subjects, it is not always possible to differentiate the processes which are part of the class the subject represents from the processes which belong only to the individual.

2) The consultation was with a simulated patient. A question that can be raised is to what extent a simulated situation affects the subject's performance and reveals the acumen of her reasoning, specifically in relation to the physical examination phase. In this study, the expert's behaviour is a clear indication of the effect of a simulated consultation on the physician's diagnostic process. The expert generated much more explanation than was necessary for the given patient case. Furthermore, the general practitioner made it clear that he felt uneasy about this artificial situation. In particular, when it comes to the absence of the non visual cues underlying problems that the patient may be reluctant to talk about in the first place, the GP said

> "I suspect, I mean it's awfully difficult in a structured situation, but I suspect that if a young woman came in with an hour and a half/two hours back pain and had walked in without any limp and there was very little in the way of objective signs, you are going to say why has she come to see me, the back pain is the excuse."

Then he added

> " She [the patient] may have other problems she wanted to talk about." (post interview session).

3) No new interactions were found at the 4th year student level. This may be due to individual differences as mentioned earlier: that particular subject did not generate new interactions of strategies but another fourth year student may have. Another possibility is that at that level no new interactions are learned. A 4th year student may gain more knowledge but not necessarily improve her reasoning strategies to support that knowledge. Additional protocols of 4th year students would have to be collected and analysed to check these two explanations.

4) The analysis of the protocols was restricted to the set of pre-defined reasoning strategies. It did not contain, for example, the misuse of the reasoning strategies by the subjects and did not clarify the kinds of evidence associated with the strategies (sufficient, positive etc) that were mentioned in chapter six. This is a limitation of the encoding of the data; the kinds of evidence used could be included by altering the formal notation of the strategies to be encoded.

8.2 Models of interactions of reasoning strategies

Following the interactions of strategies found at different levels of expertise in the protocols, the modelling of the development of strategies consists of building not *one* model but a number of models that show how medical diagnosis varies at different levels of expertise. As explained in chapter five, these models of interactions of the reasoning strategies form the developmental modeller part of DEMEREST. While this section describes the models that have been constructed, section 9.3.4 will discuss the system with reference to other developmental models reviewed in chapter four.

Each level of expertise incorporates a model of reasoning strategies interactions constructed from the interactions discussed in section 8.1.2. Five levels of expertise are modelled:

Level 1: third year medical student
Level 2: fourth year medical student
Level 3: fifth year medical student
Level 4: house officer
Level 5: general practitioner

Levels 1 to 3 correspond to medical students at different levels of clinical training and as such are novice physicians; level 4 corresponds to an intermediate physician (in this case an HO) and level 5 to an experienced physician (in this case a GP). The models of interactions of strategies for each level of expertise are shown in figures 8.4 and 8.5 respectively.

Each model contains strategies applied at that level of expertise and the interactions between these strategies. At the level of the 3rd and 4th year medical students the models are fairly simple. It is interesting that the particular 4th year student applied fewer kinds of strategies than the 3rd year student. From the next level onwards, models become progressively complex, even if the transition from one level of expertise to another are not complex interactions of strategies. In contrast to changes from level 4 to level 5 (HO to GP), more changes of strategies occur from level 2 to level 3 (4th to 5th year students) and from level 3 to level 4 (5th year student to HO).

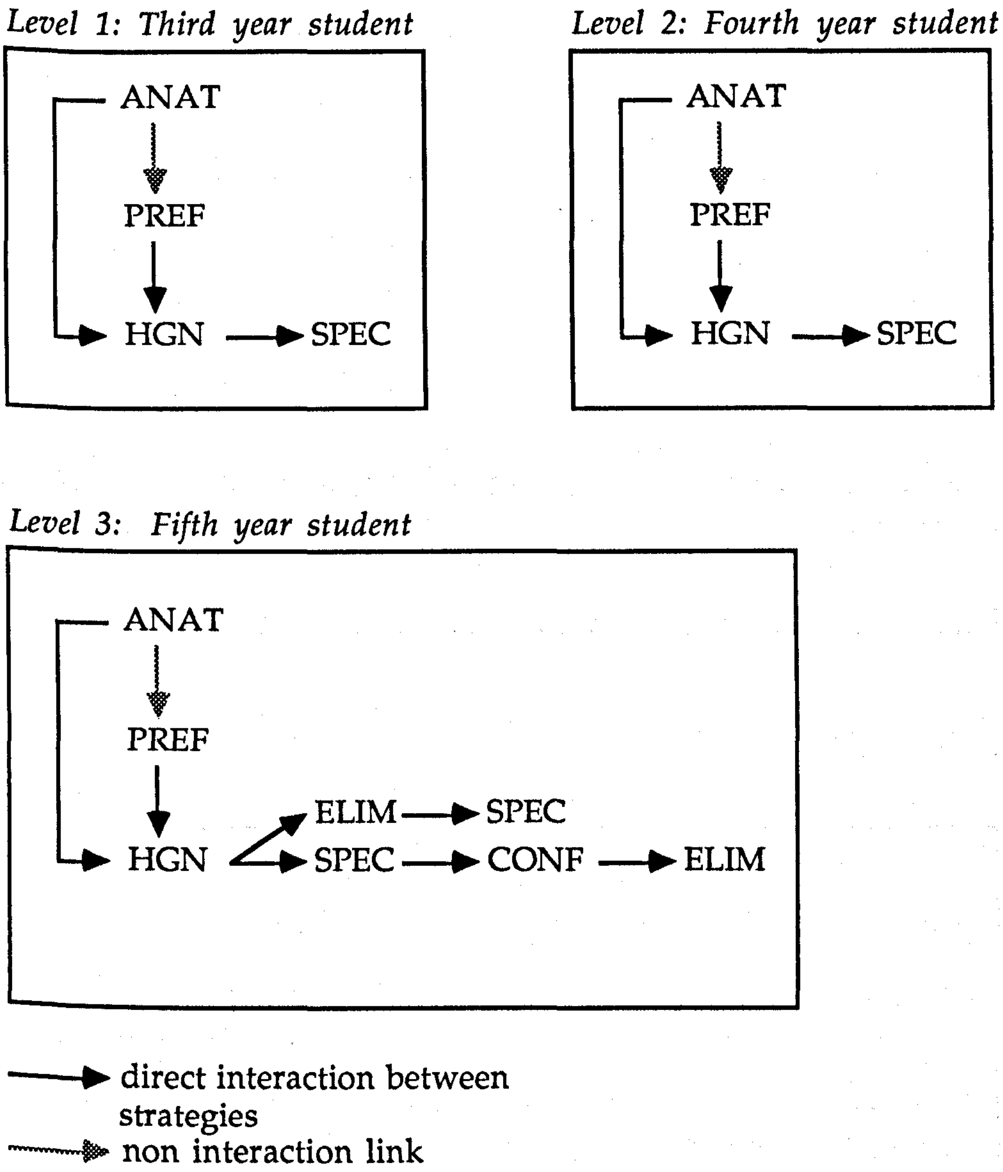
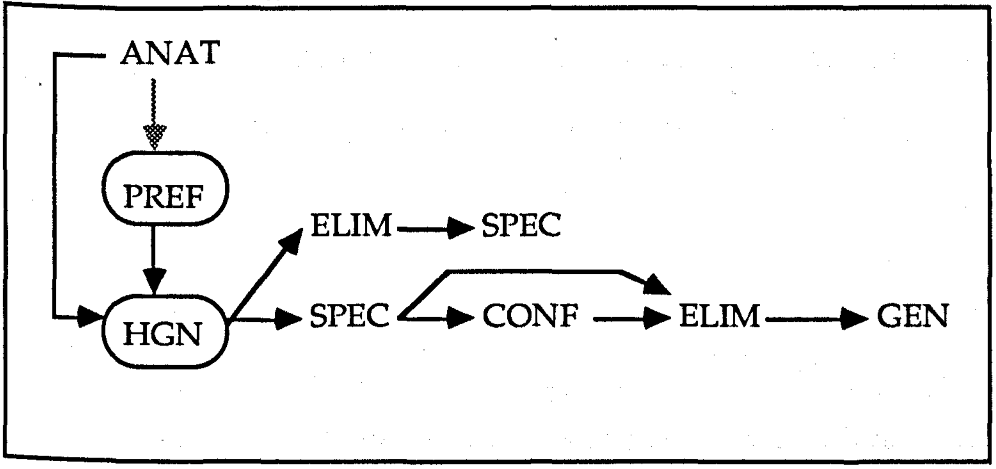


Figure 8.4: Models of changes of reasoning strategies at various levels of expertise (Levels 1 to 3)

The process of building these models is incremental, that is, each model is built as extension of the model at the level below. By doing so, the changes that have occurred from one level of expertise to another are made explicit. Changes can be either additions (e.g. level 3 - HGN ELIM), or modifications (e.g. level 4 - HGN SPEC ELIM) of strategies and interactions of strategies. Moreover, with this incremental approach, interactions of reasoning strategies which take place at more than one level of expertise are incorporated at the necessary level of expertise.

Level 4: House Officer



Level 5: General Practitioner

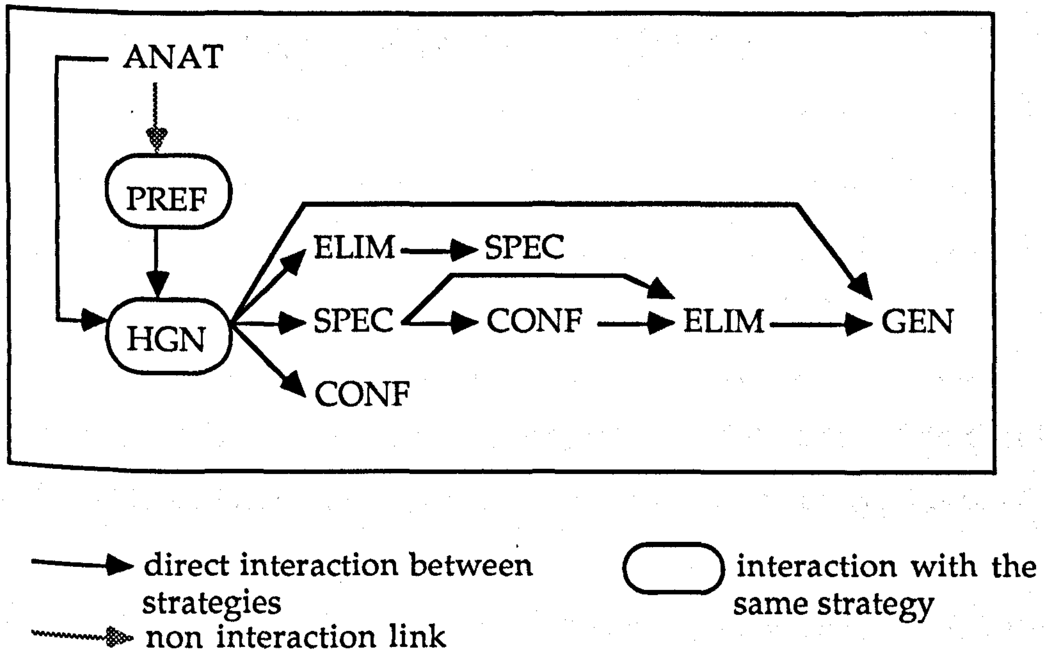


Figure 8.5: Models of changes of reasoning strategies at various levels of expertise (Levels 4 and 5)

Model of level 1:

There were three direct interactions used to build up this model. The model shows that PREF is between ANAT and HGN. This representation was chosen because in most of the protocols, the beginning of the diagnostic process started with PREF and HGN. ANAT is the first strategy in the model since the level is the most novice one and ANAT is a strategy that students used more than experienced physicians (even if not significantly). A consequence of building the model in that way was an extra link created between ANAT and PREF.

Model of level 2:

This model contains fewer kinds of interactions of strategies than the previous model as the student did not apply as many kinds of interactions as the 3rd year student. Since each model is constructed from the previous one this means that the level of the 3rd year is the same as the level of the 4th year. In other words, the interactions used by the 4th year student are already included in the model of the 3rd year student.

Models of levels 3, 4 and 5:

These models contain more interactions of strategies that were reported in section 8.1.2. In particular, one may notice, at level 4, the interaction of two similar strategies and the way in which two interactions of strategies from different levels can integrate. For example, one interaction found at level 3 is SPEC CONF ELIM and another interaction found at level 4 is SPEC ELIM GEN. The former interaction includes the latter and is extended by one strategy (GEN). When the system tests for the occurrence of the interaction SPEC ELIM GEN, it will skip the strategy CONF.

8.3 Summary

This chapter has reported on the results of an empirical study which was carried out to examine the development of the reasoning strategies, and on the modelling of interactions of strategies. The results of the study give evidence that this set of strategies form a coherent system of reasoning processes for medical diagnosis: all the defined strategies have been applied and the interactions connect and hold the strategies together to form the diagnostic process. The reasoning strategies that were identified in the literature and reported in chapter six are supported by the data. No evidence of monotonic development of these reasoning strategies was found. However, some changes of strategy at different levels of expertise were identified, which correspond to a change in the style of clinical reasoning. Using half of the data, a model of changes of strategy over time was put forward. Given the restricted number of subjects, this study has reported on some aspects of the development of the reasoning strategies, and has provided a conceptual basis for the developmental canvas, rather than a full descriptive model of the development of these reasoning strategies. The model of changes of strategy over time forms one of the components of DEMEREST, namely the developmental modeller (introduced in chapter five). In the next chapter, the implementation of this component as well as the rest of the system is described.

Chapter Nine

IMPLEMENTATION

The previous two chapters reported an empirical study that was conducted in order to determine the development of reasoning strategies. The model of changes of strategies over time derived from this study forms an integral part of DEMEREST and subsequently implemented in the developmental modeller component of the system. The rationale and overview of the system were discussed in chapter five. In this chapter, the implementation of DEMEREST is described and lessons learned from developing such a system are discussed.

9.1 Implementation of DEMEREST

The system has been implemented in LPA Prolog on a Macintosh SE. Listings of the program can be found in appendix B and outputs in appendix C. The design of the system DEMEREST was presented in section 5.2 (see figure 5.2). The reader will recall that DEMEREST was designed with three components: 1) a reasoning strategy recogniser, 2) a developmental modeller and 3) a plan generator.

The reasoning strategy recogniser aims to recognise the reasoning strategies applied by a physician, the developmental modeller to identify interactions of reasoning strategies and to determine the physician's level of expertise and the plan generator to generate the physician's plan corresponding to the applications of the strategies. In addition, DEMEREST has access to a database of medical facts. As will be explained, in the system, the reasoning strategies recogniser and the developmental modeller have been implemented together, that is, the definitions of the strategies and their interactions were grouped. In the following section, the implementation of the knowledge sources and other parts of the system are described.

9.1.1 Medical Knowledge

The medical knowledge contains different kinds of knowledge which are derived from medical textbooks and from the protocols obtained in the empirical study reported in chapters seven and eight:

- Medical knowledge derived from medical textbooks:
    1) knowledge about the diseases causing back pain
    2) knowledge about observations related to back pain i.e. signs, symptoms and tests results

- Medical knowledge derived from the protocols:
    3) relations between diseases of back pains
    4) relations between observations
    5) relations between action slots of the goals and observations
    6) relations between observations and diseases
    7) relations between action slots and diseases
    8) relations between observations and anatomical information
    9) knowledge to recognise that confirmation and elimination strategies have been applied
    10) details about the action slots of the goals

Knowledge about the diseases and knowledge about observations correspond, respectively, to the knowledge base of diseases and knowledge base of observations introduced in chapter five. Relations between diseases and between observations describe how the diseases and observations are linked in the knowledge bases. Relations between observations and diseases link the knowledge base of observations to the knowledge base of diseases.

The database contains medical knowledge at each level of expertise and not just at the expert's level. Common knowledge among the various levels of expertise was assumed since some pieces of information are used at more than one level. The following list details each kind of knowledge and its use with the reasoning strategies as well as providing some examples.

1) Knowledge base of diseases

The diseases in this knowledge base correspond to the hypotheses generated at any level of expertise. Each disease is of the form[1]:

```
        kinds(hypothesis, Hypothesis_name).
e.g.    kinds(hypothesis, slipped_disc).
```

This knowledge is accessed by the generalisation and specialisation strategies to check that the hypothesis exists in the database of diseases before generating the hypothesis. In the case of generalisation, the hypothesis being checked is a hypothesis child and in the case of specialisation the hypothesis checked is a hypothesis parent. In fact, one could use this knowledge with other strategies whenever a check is required.

---

[1] The argument with the capital letter means that the argument is a variable that will be instantiated.

2) *Knowledge base of observations*

This knowledge base contains signs, symptoms and test results of back pains. Each observation is one of the forms:

```
kinds(sign, Sign_name).                     e.g.   kinds(sign, age _42).
kinds(symptom, Symptom_name).               e.g.   kinds(symptom, back _pain).
kinds(test_result, Test_result_name).       e.g.   kinds(test_result, result_urine_test).
```

The knowledge base of observations is accessed by the problem refinement and anatomical strategies which check whether the observation considered exists.

3) *Relations between diseases*

The relations between diseases are of the form:

```
        kinds(Hypothesis_parent, Hypothesis_child).
e.g.    kinds(infection, urinary_infection).
```

Each relation is hierarchical and is accessed by the generalisation and specialisation strategies. In the case of generalisation, the system looks for the hypothesis parent of a given hypothesis child; and in the case of specialisation, the system looks for the hypothesis child of a given hypothesis parent.

4) *Relations between observations*

The link between two observations is of the form:

```
        kinds_obs(Observation1, Observation2).
e.g.    kinds_obs(movements_limited, bony_pain).
```

Links between observations are used by the problem refinement strategy. Each kinds_obs links the goals to the knowledge base of observations: Observation1 corresponds to the effect of a given goal and Observation2 is an element of the knowledge base of observations. In the above example, the house officer refined *movements limited* to *bony pain*. This kind of problem refinement is experience based problem refinement (see section 6.3.5).

5) Relations between action slots of the goals and observations

As explained above observations can be linked together. In some cases, instead of linking two observations, it is useful to link the slot Action of the goal to an observation. The relation becomes:

```
                kinds_obs(Action, Observation2).
e.g.            kinds_obs(ask_radiation_pain, refine_source_of _the _pain).
and not         kinds_obs(no_radiation_pain, refine_source_of_the_pain).
```

for the goal[2]

Name of the goal: check radiation of the pain
Precursor of the goal: none
Subgoals: none
Action of the goal: ask radiation of the pain
Effect of the goal: no radiation of pain

In many cases, associating the effect of the goal with another observation does not carry out the true meaning of what the physician did. The action of the goal conveys a problem refinement strategy rather than the effect. In other words, in this case the physician does not use a patient observation but rather uses her own observation to refine that observation or generate a hypothesis. For example, the house officer asked the patient if the pain

---

[2] The prolog notation for this goal would be:
`goal(check_radiation_pain,[],[], ask_radiation_pain, [no_radiation_pain]).`

radiates. The problem refinement here is to refine *radiation of the pain* and not refine *no radiation of the pain.* This alternative to use the action slot and not the observation slot with PREF is done with all the protocols.

The action slot helps in differentiating between a patient observation *(e.g sharp pain)* and a non patient observation *(e.g. pain radiates)* that the physician uses to carry out the diagnostic process. Moreover, it helps to differentiate between a patient/doctor interaction and a doctor/doctor interaction. By using the action slot instead of the observation slot, the evidence becomes in some cases the action slot.

6) *Relations between observations and diseases*

The relation between an observation and a hypothesis is of the form:

```
              causes(Observation, Hypothesis).
       e.g.   causes(radiation_pain, sciatica).
```

This relation links the knowledge base of diseases with the knowledge base of observations. This type of link is used by the strategy hypothesis generation. The above example shows that the radiation of the pain indicates sciatica.

7) *Relations between action slots and diseases*

As explained above (point #5), in some cases, it is useful to link action and observation to differentiate between patient observation and physician observation. Likewise, an action slot can be linked with a hypothesis to make the same distinction. The relation becomes:

```
                  causes(Action, Hypothesis).
e.g.              causes(ask_radiation_pain, trapped_nerve).
and not           causes(no_radiation_pain, trapped_nerve).
```

For example, the house officer asked the patient if the pain radiates. She linked *radiation of the pain* to *trapped nerve* and not *no radiation of pain* to *trapped nerve.*

8) Relations between observations and anatomical information

The database of medical knowledge does not contain a descriptive representation of the anatomy of the back. Instead it includes some anatomical information about the back. Each piece of anatomical information is linked with a patient observation in the form of:

```
        anat(Observation, Anat).
e.g.    anat(no_aggravating_relieving_factors,disc_going_into_hole_in_spine).
```

This knowledge is used by the anatomically based strategy.

9) Knowledge for confirmation and elimination strategies

As mentioned in chapter seven, subjects did not always make explicit in their protocol whether they were trying to confirm or rule out a hypothesis. DEMEREST models confirmation and elimination strategies when it is explicit in the protocol. The confirmation and elimination strategies are based upon some evidence, that is, some observation that is used to confirm or rule out a hypothesis. DEMEREST recognises whether confirmation and elimination strategies have been used by finding in the medical database some particular facts which are explained below.

Confirmation:

In the system, confirmation can occur in two instances i) after a hypothesis generation (this is the interaction between HGN and CONF) and ii) after a specialisation (this is the interaction between SPEC and CONF). The facts through which the system will search to confirm a hypothesis contain observations that are used in confirmation and in addition, the negatives of these observations that may also be used to confirm a hypothesis. For example, one may want to confirm (correctly or not) *sciatica* if there is *radiation in the leg* or to confirm *prolapsed disc* if there is *no radiation.*

The facts associated with the interaction HGN CONF are of the form[3]:

```
        negative_rs6(Observation, Negative_observation).
e.g.    negative_rs6(past_waterwork_infection, no_past_waterwork_infection).
```

In this example, the 5th year student tries to confirm *urinary_infection* with the piece of evidence *past_waterwork_infection.*

The facts associated with the interaction SPEC CONF are of the form:

```
        negative_rs3(Observation, Negative_observation).
e.g.    negative_rs3(similar_pain, no_similar_pain).
```

In this example, the GP tried to confirm (i.e. that the patient had a *similar disc problem* to the one she had previously) with the evidence that the *pain was similar.*

Elimination

In the system, elimination can occur in two instances i) after a hypothesis generation, (this is the interaction between HGN and ELIM) and ii) after a specialisation, (this is the interaction between SPEC and ELIM). The facts

---

[3] In the term negative_rs3, rs stands for reasoning strategy, the number 3 corresponds to the strategy confirmation. Each strategy was originally numbered in the program: 1 for generalisation, 2 specialisation, 3 confirmation, 4 elimination, 5 problem refinement, 6 hypothesis generation and 7 anatomical.

through which the system will search to rule out a hypothesis consist of observations that are used for elimination and the negatives of these observations may also be used to rule out a hypothesis. For example, one may want to rule out (correctly or not) a *vertebral problem* if there is *no tenderness over the vertebrae* or to rule out *a mechanical problem* if there is *tenderness over the vertebrae.*

The facts associated with the interaction HGN ELIM are of the form:

```
      negative_rs(Observation, Negative_observation).
e.g.  negative_rs(results_positive_urine_sample, results_negative_urine_sample).
```

In this example, the 5th year student tried to rule out *kidney_condition* based on the evidence that the *result of a urine sample is negative.*

The facts associated with the interaction SPEC ELIM are of the form:

```
      negative_rs4(Observation, Negative_observation).
e.g.  negative_rs4(flexion_ok, flexion_not_ok).
```

In this example, the 5th year student tried to rule out a *hip problem* as evidence of the fact that the *flexion was ok.*

As mentioned in this section, the action slot allows one to differentiate between patient observation and physician observation. Hence, the action slot may also be found in the negative facts. For instance, the physician may not have the results of the urine sample for the given patient but knows that if they are positive it could mean a kidney problem.

10) Details about the action slots of the goals

The action slot of a goal describes an action that the physician had taken, such as asking the patient to locate the pain. In some cases, more information or other information about the action is required to apply a reasoning strategy. Since this extra information needed is not present in the other slots of the goal, details about the action are added. Detail of an action slot of a goal is of the form:

```
      details_action(Action, Details_action).
e.g.  details_action(ask_about_occupation, ask_occupation_heavy_lifting).
```

In the above example, the 3rd year student asked the patient about her occupation. It happens that the patient's occupation was office secretary, but the student was nevertheless interested in whether or not her occupation required her to lift heavy objects; since physical occupations such as carrying and lifting heavy objects, may also cause back pain. In order to model, in this case, a hypothesis generation strategy the fact details_action was used, to generate the hypothesis *disc problem* from the observation *occupation heavy lifting*.

Details about actions are used by generalisation, specialisation and hypothesis generation strategies, and if necessary could be extended to confirmation, elimination, problem refinement and anatomically based strategies.

9.1.2 Inputs for DEMEREST

As indicated in section 5.2.1, the physician's protocol is transformed manually into a set of goals that can be understood by the system. The following illustrates how this process was carried out by hand analysis.

The task of extracting a goal from the protocol involves filling in the different slots of a goal structure (section 5.3.3) that will be the input for the system. The physician has been interviewed and her protocol recorded and analysed. An interaction between the patient and the doctor is chosen, that is, one question/answer (or a set of questions/answers) is being considered.

Each goal has a number of slots to be filled in:

i) a name that corresponds to the topic of the interaction
ii) a precursor that corresponds to the effect of the previous goal
iii) one or more subgoals of the given goal
iv) an action name that corresponds to the action of the doctor
v) an effect that is a patient observation and corresponds to the answer of the interaction between the patient and the doctor.

Let us take the example of the 3rd year student who asked the patient about the location of the pain and its severity. The excerpt of the protocol is shown below (lines 1 to 4 for the location of the pain and 5 to 6 for its severity):

1 "Student: whereabout is the pain?
2 Patient: it's about from here to here
(the patient shows the right side of the back)
3 Student: and it is just on the right side ?
4 Patient: yes.
5 Student: can you describe the type of pain to me ?
6 Patient: well it's quite a bad pain, it's sharp really, it's here all the time."

Two goals may be constructed from the excerpt of this procotol: *check location of the pain* and *check severity of the pain*. The slots of the goal *check location of the pain* are filled up as follows:

Name of the goal: check location of the pain
Precursor of the goal: none
Subgoals: none
Action of the goal: ask about the location of the pain
Effects of the goal: right sided back pain

The name given is *check location of the pain*. There is no precursor since it is the first goal. There is also no subgoal for *check location of the pain*. The action name is *ask about location of the pain* and the effect *right sided back pain* because it is the patient's response to the question.

The slots of the goal *check severity of the pain* are filled up as follows:

Name of the goal: check severity of the pain
Precursor of the goal: right sided back pain
Subgoals: none
Action of the goal: ask about severity of the pain
Effect of the goal: sharp pain

The name given is *check severity of the pain,* the precursor is *right sided back pain* because it is the effect of the previous goal *check location of the pain*. There is no subgoal for *check severity of the pain,* the action name is *ask about severity of the pain* and the effect is *sharp pain* because that is what the patient had said about the severity of the pain.

9.1.3 Recognising the reasoning strategies

This section examines the mechanisms to instantiate a reasoning strategy from a given goal. The inputs are a goal from the physician's plan which is derived by hand analysis and the set of possible reasoning strategies. The output is an *instantiation* of one or more of these reasoning strategies, that is, the context in which the strategy has been applied (e.g. the goal, the hypotheses generated etc). It should be pointed out that there may be more than one strategy associated with a goal which will correspond to interactions between strategies.

The following sections describe the implementation of each reasoning strategy. For each strategy, an example taken from the protocols of the empirical study is given as illustration. It should be pointed out that some of the strategies have more than one definition because of the use of details_action and action (discussed in section 9.1.2) in addition to the observation slot of a goal.

GENERALISATION:

The predicate that describes the generalisation strategy is given a hypothesis H2 (hypothesis child), then looks for another hypothesis H1 (hypothesis parent) that is more general than H2 in the database of medical facts. The predicate also checks that the hypothesis H1 is present in the database. For example, the general practitioner applied generalisation in the following context: the GP generated the hypothesis *acute back strain,* from a list of observations *(no aggravating relieving* *factors, similar pain, sharp pain, right sided back pain, onset this morning)* and then generalises to *muscularskeletal.* The corresponding goal is[4]:

Name of the goal: check aggravating relieving factors
Precursor of the goal: none
Subgoals: none
Action of the goal: ask aggravating relieving factors
Effect of the goal: no aggravating relieving factors

The generalisation strategy cannot be applied on its own but through the interaction with the strategy of hypothesis generation, as results of the study showed (see section 8.1.2). In this example, the effect *no aggravating relieving factors* is linked to the knowledge base of diseases and the system generates the hypothesis *acute back strain* by applying the hypothesis generation. Incidentaly, in this case, the GP applied a repetition of HGN by combining *no aggravating relieving factors* with other observations. Then, the system succeeds in applying generalisation by finding another hypothesis *muscularskeletal* to *acute back strain.* In the database of medical facts both hypotheses are presented as follows:

kinds(hypothesis, acute back strain).
kinds(hypothesis, muscularskeletal).

and their hierarchical connection between hypothesis parent and hypothesis child:

kinds(muscularskeletal, acute back strain).

---

[4] For reasons of clarity, it is assumed in this example and the following ones that there are no precursors and subgoals.

The system will output a generalisation strategy that has been applied in the following format:

FORMAT	EXAMPLE
STRATEGY: GEN	STRATEGY: GEN
GOAL: goal name	GOAL: check aggravating relieving factors
HYPOTHESIS CHILD: hypothesis	HYPOTHESIS CHILD: acute back strain
HYPOTHESIS PARENT: hypothesis	HYPOTHESIS PARENT: muscularskeletal
EVIDENCE: observation/action	EVIDENCE: characteristics of the pain

SPECIALISATION:

The predicate that describes the specialisation strategy is defined in a similar way to the predicate for generalisation: given a hypothesis H1 (hypothesis parent), the system searches for another hypothesis H2 (hypothesis child) that is more specific than H1 in the database of medical facts. The predicate also checks that the hypothesis H2 exists in the database. For example, the 5th year student applied specialisation in the following context: the student asked the patient about past waterwork infections and specialised from *infection* to *urinary infection* given the evidence of *waterwork infections in the past* when the patient answered that she had a number of urinary infections in the past. The corresponding goal is :

Name of the goal: check past waterworks infection
Precursor of the goal: none
Subgoals: none
Action of the goal: ask about past waterworks infection
Effect of the goal: waterwork infections in the past

In a similar way to generalisation, specialisation cannot be applied on its own but after hypothesis generation as the results of the study showed (see chapter eight). In this example, the effect *had urinary infections in the past* is linked to the knowledge base of observations, and the system generates *infection* by applying hypothesis generation. Then, the system succeeds in finding a relation to *urinary infection* by applying specialisation. As in the generalisation strategy, both hypotheses are presented in the database of medical facts as follows:

```
kinds(hypothesis, infection).
kinds(hypothesis, urinary_infection).
```

and their hierarchical connection between hypothesis parent and hypothesis child:

```
kinds(infection, urinary_infection).
```

The system will output a specialisation strategy that has been applied in the following format:

FORMAT	EXAMPLE
STRATEGY: SPEC	STRATEGY: SPEC
GOAL: goal name	GOAL: check past waterworks infections
HYPOTHESIS PARENT: hypothesis	HYPOTHESIS PARENT: infection
HYPOTHESIS CHILD: hypothesis	HYPOTHESIS CHILD : urinary infection
EVIDENCE: observation/action	EVIDENCE : waterwork infections in the past

CONFIRMATION:

The predicate that describes the confirmation strategy is given a hypothesis which is to be added to the differential diagnosis. As explained in section

9.1.1, DEMEREST recognises whether elimination strategy has been used by finding some particular facts of the form negative_rs#(Observation, Negative_observation) in the medical database. Confirmation may occur in the interactions HGN CONF and SPEC CONF.

For example, the 5th year student tries to confirm the hypothesis *urinary infection* given the evidence that the patient has a history of such infections. In this example, the goal is the same as the one given in specialisation. Once the system had applied the specialisation strategy, it succeeded in applying confirmation by finding in the database the fact:

```
negative_rs3(past_waterworks_infection, no_past_waterwok_infection).
```

and the hypothesis *urinary infection* is added to the list of differential diagnosis:

```
differential_diagnosis[urinary_infection].
```

The system will output a confirmation strategy that has been applied in the following format:

FORMAT	EXAMPLE
STRATEGY: CONF	STRATEGY: CONF
GOAL: goal name	GOAL: check past waterworks infections
HYPOTHESIS CONFIRMED:	HYPOTHESIS CONFIRMED:
hypothesis	urinary infection
EVIDENCE: observation/action	EVIDENCE is : past waterwork infections
DIFFERENTIAL DIAGNOSIS:	DIFFERENTIAL DIAGNOSIS:
hypothesis	urinary infection

ELIMINATION:

The predicate that describes the elimination strategy is given a hypothesis which is to be subtracted from the differential diagnosis. As explained in section 9.1.1, DEMEREST recognises whether an elimination strategy has been used by finding some particular facts of the form negative_rs#(Observation, Negative_observation) in the medical database. Elimination may occur in the interactions HGN SPEC ELIM and HGN ELIM.

For example, the 5th year student tried to rule out the hypothesis *hip problem* given the evidence that the *flexion is ok.* The corresponding goal is:

Name of the goal: check flexion
Precursor of the goal: none
Subgoals: none
Action of the goal: ask the patient to flex
Effect of the goal: flexion ok

As with confirmation, elimination cannot be applied on its own but through interaction with hypothesis generation as the results of the study showed (see chapter eight). Once the system has generated the hypothesis *hip problem* linked with the effect of the goal, it succeeds in ruling out that hypothesis by finding the fact:

```
negative_rs(flexion_not_ok, flexion_ok).
```

and the hypothesis is subtracted from the list of hypotheses in the differential diagnosis, which then becomes empty, unless there was already another hypothesis which had been confirmed previously.

The system will then output an elimination strategy that has been applied in the following format:

FORMAT	EXAMPLE
STRATEGY: ELIM	STRATEGY: ELIM
GOAL: goal name	GOAL: check flexion
HYPOTHESIS RULED OUT:	HYPOTHESIS RULED OUT :
hypothesis	hip problem
EVIDENCE: observation/action	EVIDENCE is: flexion is ok
DIFFERENTIAL DIAGNOSIS:	DIFFERENTIAL DIAGNOSIS is : empty
hypothesis	

PROBLEM REFINEMENT:

The predicate that describes the problem refinement strategy is given an observation which can be a patient observation or a physician observation and generates either another observation in the case of experience based refinement, or a routine protocol in the case of protocol based refinement (see section 6.3.5). Observations (and routine protocols) are connected by kinds_obs(observation, observation-protocol) in the database of medical facts.

For example, the house officer refined *movements limited* to *bony pain.* The corresponding goal is:

Name of the goal: check aggravating and relieving factors
Precursor of the goal: none
Subgoals: none
Action of the goal: ask for aggravating and relieving factors
Effect of the goal: movements are limited

The system succeeds in applying problem refinement by finding the fact:

kind_obs(movements_limited, bony_pain).

The system will output a problem refinement strategy that has been applied in the following format. The first example shows a patient observation being refined to another observation. The second example shows a physician observation and the use of a routine protocol.

FORMAT
STRATEGY: PREF
GOAL: goal name
OBSERVATION/ACTION: observation/action
OBSERVATION REFINED TO or
ROUTINE PROTOCOL USED: observation

EXAMPLE 1	EXAMPLE 2
STRATEGY: PREF	STRATEGY: PREF
GOAL:check aggravating relieving factors	GOAL: check radiation pain
OBSERVATION/ACTION:	OBSERVATION:/ACTION
movements limited	ask radiation pain
OBSERVATION REFINED TO or	OBSERVATION REFINED TO or
ROUTINE PROTOCOL USED:	ROUTINE PROTOCOL USED:
bony pain	routine protocol for refining the pain

HYPOTHESIS GENERATION:

The predicate that describes the hypothesis generation strategy is given an observation, Obs, which can be either a sign, a symptom or a test result, and it then generates a hypothesis H. The observation and the hypothesis are connected by causes(Obs, H) in the database of medical facts. The system also checks for causes(Action, H) in the case that the physician has used her own observation. For instance, the 4th year student asked about the location of the pain and being told that the pain was on the right side, generated the possible hypothesis kidney problem. The corresponding goal is :

Name of the goal: check location of the pain
Precursor of the goal: none
Subgoals: none
Action of the goal: ask about location of the pain
Effect of the goal: right sided back pain

The system succeeds in applying hypothesis generation by finding the fact:

causes(right sided back pain , kidney problem).

DEMEREST keeps track of the hypotheses which have been generated, not just the ones put in the differential diagnosis. Each time a hypothesis has been generated, the system first checks that the hypothesis is not already in the list of hypotheses, and if it is not, the hypothesis is added:

list_of_hypotheses([kidney_problem]).

The list of hypotheses contains all the hypotheses that have been generated, and not necessarily confirmed or ruled out. In contrast, the list of the differential contains only the hypotheses that have been confirmed or ruled out.

The system will output a hypothesis based strategy that has been applied in the following format. The observation is either a patient's observation (example 1) or a physician's observation (example 2).

FORMAT
STRATEGY: HGN
GOAL: goal name
OBSERVATION: observation/action
HYPOTHESIS: hypothesis
LIST OF HYPOTHESES: list of hypotheses(hypotheses)

EXAMPLE 1
STRATEGY: HGN
GOAL: check location of the pain
OBSERVATION/ACTION:
right sided pain
HYPOTHESIS: kidney problem
LIST OF HYPOTHESES:
list of hypotheses(kidney problem)

EXAMPLE 2
STRATEGY: HGN
GOAL: check occupation of the patient
OBSERVATION/ACTION:
ask lifting heavy objects
HYPOTHESIS: disc problem
LIST OF HYPOTHESES:
list of hypotheses(disc problem)

ANATOMICAL:

The predicate that describes the anatomically based strategy is given an observation Obs which can either be a sign, a symptom or a test result and it then generates an anatomical information. The observation and the anatomical information are related by anat(Obs, ANAT) in the database of medical facts. For example, the 4th year student asked about any aggravating or relieving factor of the pain, and gave an anatomical information i.e. that the disc may be going into the hole in the spine. The corresponding goal is:

Name of the goal: check aggravating and relieving factors
Precursor of the goal: none
Subgoals: none
Action of the goal: ask about aggravating and relieving factors
Effect of the goal: no aggravating and relieving factors

The system succeeds in applying the anatomical based strategy by finding the following fact in the medical database:

anat(no_aggravating_relieving_factors, disc_going_into_hole_in_spine).

The system will output an anatomically based strategy that has been applied in the following format:

FORMAT	EXAMPLE
STRATEGY: ANAT	STRATEGY: ANAT
GOAL: goal name	GOAL: check aggravating and relieving factors
OBSERVATION: observation	OBSERVATION: no aggavating relieving factors
ANATOMICAL INFORMATION:	ANATOMICAL INFORMATION:
anatomical information	the disc is going into the hole in the spine

9.1.4 Recognising the interactions of strategies

The system contains a model of changes of strategies at different levels of expertise (which was constructed from the empirical study, see chapter eight). As mentioned in the beginning of the chapter, the definitions of the strategies and their interactions are grouped. In other words, the reasoning strategies recogniser mentioned in chapter five is in fact included in the developmental modeller. The inputs are the physician's set of goals, the set of reasoning strategies and the models of interactions of reasoning strategies (derived from the empirical study). The outputs are the instantiations of reasoning strategies and their interactions and a possible level of expertise for the given protocol.

The developmental modeller takes one goal at a time, and runs through each strategy and the various interactions of reasoning strategies. Each interaction corresponds to a level of expertise and each level of expertise is associated with one or more types of interactions. For instance, the interaction HGN ELIM SPEC and the interaction HGN SPEC CONF ELIM both belong to the level of expertise #3. When an interaction is found, the program sets the corresponding level of expertise for that goal.

Once a level of expertise has been determined, it is only changed if a higher level of expertise can be achieved. For example, the house officer had first applied the interaction PREF HGN which put the level of expertise to #1, then applied the interaction of PREF's which put the level of expertise one step higher to #4. Her level of expertise will remain at #4 even if she applies an interaction of a lower level. The assumption made here is that a physician who applies an interaction of strategies at level #3 for instance, will be able to apply the interactions of levels #1 and #2.

The reader will recall that level 2 of interactions of strategies contained the same kinds of interactions of strategies as level 1 (section 8.2). In other words, level 2 is similar to level 1. This means that the system will not be able to determine a level 2. In order to do so, it would be necessary to have a new interaction that will correspond to a change from a 3rd year student level to 4th year student level.

The following is an example of how the system generates an interaction of strategies. Let us look at the 5th year student who applied the interaction of reasoning strategies HGN SPEC. She first generated the hypothesis *infection* from the evidence *past of waterworks infection* and then specialised into *urinary infection* with the same evidence. The goal for this context is:

Name of the goal: check past waterworks infections
Precursor of the goal: none
Subgoals: none
Action of the goal: ask about waterworks infections
Effect of the goal: history of past waterwork infections

The system starts by trying to apply ANAT since it is the first strategy to be applied. It fails to apply the anatomical strategy because it cannot find a fact of the form anat(past_of_waterwork_infections, X) where X is some anatomical information. The system then goes on to try to apply PREF and again fails because there is no fact of the form kinds_obs(past_of_waterwork_infections, X) where X is some other observation or routine protocol.

The system then tries to apply HGN and succeeds in it since it finds a fact of the form causes(past_of_waterworks_infections, infection). It has also checked that *infection* is in the database by searching for the fact kinds(hypothesis, infection). The system generates the hypothesis *infection* and adds it to the list of hypotheses. The system attempts to apply one interaction from HGN i.e. HGN SPEC. Hence, the program tries to apply SPEC by using some piece of information generated by the previous strategy HGN. In this case, the information is the hypothesis *infection*. The system searches for a fact of the form kinds(infection, X) where X is another hypothesis (a child hypothesis) and finds kinds(infection, urinary_infection). The system generates the specialised hypothesis *urinary infection*. Every time the interaction HGN SPEC is found in a physician's protocol, it puts her or his level of expertise to #1 (level of a 3rd year student).

9.1.5 Generating plans

The plan generator generates a physician's plan containing the goals and their associated reasoning strategies that the physician has applied. The inputs are the goals and the output is a physician's plan.

Novices and more experienced physicians have similar goals to achieve; however they may not achieve these goals necessarily in a similar order. In order to generate a physician's plan that would indicate whether a goal was applied in the same order in another plan of a physician with a different level of expertise, a plan was manually constructed from the protocols (referred to as a plan of reference). A graphical representation of the plan of reference can be found in appendix A8. The plan of reference is global since it includes all the goals that were constructed from the empirical study and is considered to be the default plan. By constructing a plan from the most novice physician (3rd year student) onwards, the order of the goals in the default plan is closer to what novices do. Thus, a deviation of the default order of the goals is expected with intermediate and more experienced physicians. More experienced physicians would be expected to produce other deviation plans. It should be clear that using a deviation plan does not necessarily signify improper reasoning. It means that the goals of the plan have been manipulated in a different way.

The plan generator generates a plan for each level of expertise. Graphical representations of the plans can be found in appendix A7. Each plan contains the physician's goals and the strategies and interactions of strategies associated with them.

Generating a deviation plan

The plan generator generates the global plan for each level of expertise by searching for the effect/precursor link between two goals explained in section 5.3.5. That is, the effect of one goal is the precursor of another one. The names of the goals (of a given protocol) have been input in the program according to the order of the default plan. Each default goal is of the form:

default_goal(name of the goal).
e.g. default_goal(check location of the pain).

A default goal G is characterised by the fact that no other goal whose precursor is the same as the effect of G can be found. In this case, the system generates the strategies associated with the goal and takes the next goal in the list of default goals. In contrast, a deviation goal G1 is characterised by the fact that the system can find another goal G2 whose precursor is the same as the effect of G1. This other goal would not be the next goal in the list of default goals. The system generates the reasoning strategies associated with G1 and then takes G2 as the next goal to apply.

The following example shows how the 5th year student had deviated from the default plan for the goals related to characteristics of the pain. The default order of the goals for characteristics of the pain is shown below under the heading 'Default set of goals'. In the default plan, *check duration* is first applied then *check sudden onset* and so on until *check aggravating factors*. However, the student did not always follow the default order of these goals. The deviation to the default plan is also shown below under the heading 'Deviation set of goals'.

DEFAULT SET OF GOALS	DEVIATION SET OF GOALS
default_goal(check_duration_pain).	default_goal(check_duration_pain).
default_goal(check_sudden_onset).	*default_goal(check_similar_pain).*
default_goal(check_gradual_onset).	default_goal(check_sudden_onset).
default_goal(check_radiation_pain).	default_goal(check_gradual_onset).
default_goal(check_relieving_factors).	*default_goal(check_radiation_pain).*
default_goal(check_aggravating_factors).	default_goal(check_relieving_factors).
	default_goal(check_aggravating_factors).

The student first asked about the duration of the pain, and being told about the duration along with the additional information that the patient had a previous event of back pain, then asked whether the pain was similar to the previous pain rather than asking whether it was of sudden onset. Afterwards, the student returned to the default plan by asking whether the pain was of a sudden onset or a gradual onset, but then deviated again by asking about relieving and aggravating factors and finally asking whether or not the pain radiate anywhere.

This example shows two uses of deviations: i) in the first case, the student used the goal *check similar pain* that is not a subgoal of *check characteristics of the pain* but a subgoal of *check previous back pain.* In the second deviation, the student used the goal *check radiation pain* in a different order from the default plan.

Generating a default plan

The plan that is being generated by the system as explained above corresponds to the order in which the goals were found in the protocols. The system can also generate a plan which follows the sequence of the default plan. The system does not look for default_goal. It simply takes a goal and the data structures associated with it (name, percursor, action and effect) at a time from the list of goals and instantiates any strategy or interactions of strategies associated with a goal.

The default and deviation plans provide two different views of how the goals have been manipulated for a given level of expertise. Unlike the deviation plan, the default plan generates only the goals that have strategies and interactions associated with it. This point is explained further below.

Abstractions of the plans

A physician's plan (as represented in appendix A7) contains goals which are more abstract than others. For instance, *check characteristics of the pain* is a more abstract goal than *check duration of the pain.* Such a goal may not necessarily have a reasoning strategy associated with it. Hence, the ordered global plan contains all the goals, even those which do not have any strategy associated with them. As explained in section 5.3.3, in the system, each goal has the same level of abstractions and thus it does not generate abstraction spaces of a plan. However, each goal structure has a subgoal slot which contains the possible subgoals of that goal. The slots of subgoals could be useful in generating abstractions of a plan. This is an issue for further work discussed in chapter eleven.

The abstractions of goals is related to the issue (referred in section 8.1.2) regarding indirect interactions of strategies. Indirect interactions between strategies correspond to abstractions of the goals being used. For example, a physician may generate a strategy HGN for the goal *check radiation of the pain* and another strategy PREF for the goal *check location of the pain.* Both goals are related since they both contribute to achieve the goal *check characteristics of the pain.*

9.2 DEMEREST: Lessons Learned

This section discusses the system DEMEREST in terms of its planning features and the use of planning in a tutoring environment. The strengths and weaknesses of the system and its current implementation are also examined.

9.2.1 DEMEREST and Artificial Intelligence Planning

The use of planning techniques in intelligent tutoring systems has already been investigated (e.g. Peachey & McCalla 1986, Bretch 1988). However, research in this direction has considered planning from the teacher's point of view and not from the student's point of view. That is, the focus has been on the goals and plans of the teacher rather than on the goals and plans of the student.

For instance, tutoring systems that combine CAI programs and AI planning techniques, are able to plan teaching strategies tailored to the particular student being taught (Peachey & McCalla 1986). The use of planning can also be extended to the whole tutoring system by viewing the instructional session as a planning process (Bretch 1988). The system is referred to as an instructional planner, using a global plan of instructional goals to achieve. The system interacts with the student, reacts to input from the student and replans its course of actions if necessary.

In contrast to these two examples, DEMEREST is concerned with the goals and plans of the student. DEMEREST could provide the tutoring system with information about the student's current state of knowledge and development. The role of the tutoring system is to help the student to be aware of her reasoning and mistakes made and to provide the student with new reasoning strategies to learn according to her developmental state.

As mentioned ealier, the main focus of this research work has not been to develop new AI planning mechanisms, but to exploit AI planning techniques in the design of DEMEREST. Planning has been used as a way to analyse and decompose the diagnostic process into a set of goals and associate the goals and reasoning strategies to form a plan. DEMEREST incorporates several features that are found in other AI planners. These features include representation of operators and plans, interactions between goals and the planning process.

1) Representation of operators

As in other planners such as STRIPS (Fikes and Nilsson 1971), the operators in DEMEREST contain a set of information specifying aspects of the operator such as name, precursors (referred to in other planners as preconditions) actions and effects. However, as explained in section 5.3.5, this set of information does not convey all the knowledge necessary to diagnose medical problems in general and back problems in particular. That is, the structure of a goal does not contain enough medical information for a goal to be achieved. Hence, the goals/operators are extended and complemented by two knowledge bases; one containing knowledge about observations and the other one knowledge about diseases/hypotheses.

2) Interactions of Goals

DEMEREST is similar to most planners in that it constructs a plan from a single goal. This goal is usually decomposed into subgoals which may interact with one another in complex ways. The problem of interacting goals arises whenever there are conjunctive goals, that is, there is more than one condition to be satisfied. A number of solutions have been proposed to deal with goal interactions, such as constraint posting (MOLGEN, Stefik 1981) whereby constraints represent interactions between subproblems, and critics (HACKER, Sussman 1975). Goal interaction has also been used to guide the evolution of the problem (Tate 1975).

The interaction of goals in DEMEREST is based on Wilensky's idea of positive goal interaction (Wilensky 1983). Wilensky refers to one kind of positive goal interaction called goal overlap where it is more efficient to have a single plan for two goals, if the goals are similar enough, than a plan considered for each goal separately. In the case of medical diagnosis, it is indeed better to have a plan that tries for instance to achieve the goals *check history* and *check physical examination* as they contribute towards the same goal *check diagnosis of the patient.* Wilensky also discusses negative goal interactions. Negative goal interactions are goal interactions which cause difficulties to the achievement of plans and have not been implemented in DEMEREST. The system only deals with positive goal interactions.

3) Planning Process

The emphasis in most AI work on planning has been on producing plans that are correct and complete. Once these plans are generated, they can be executed. Although, this approach is well suited for tasks such as robot planning, it is not always appropriate in other cases. More recently, however, work in planning in the context of HCI - human-computer interaction - (Young & Simon 1987) has addressed this issue. The very nature of interaction with a computer, demands that the planning process be intertwined with the execution process. Likewise, planning in the context of medical diagnosis also requires that the generation of the plan and its execution be interlinked.

Young and Simon argue that the very nature of interactive settings means that all task-related behaviour of the user can be classified as situated actions (Suchman 1985). That is, the user is not purely performing a goal directed activity. Rather, the user interplays between such goal driven activity and the actual physical and functional setting, that includes but is not limited to the state of the computer system. Similarly to planning in HCI, planning in the context of medical diagnosis may involve a mixture of goal driven activities and task-related behavior of the physician which can be classifed as 'situated action' (Suchman 1985). A physician may be at one time goal-driven (e.g. the physician plans to find out about the location and severity of the pain). At other times the physician may act in response to a concrete context, e.g. the patient, without being asked, gives the physician some information about her condition.

Suchman has questioned the role of planning in cognitive science and offered an alternative of situated action. As Elsom-Cook (1989) points out, many of the criticisms which Suchman makes of AI planning can be answered with the use of opportunistic planning systems such as (Hayes-Roth and Hayes-Roth 1979). Their system makes use of a blackboard architecture and a set of planning specialists to facilitate multi directional planning. The planner can establish models of planning and evaluation criteria to be achieved in a particular context and can support the simultaneous following of multiple plans. The research presented here has focused on the goal-directed activity of the physician and as mentioned in section 5.3.3, opportunistic planning as in (Hayes-Roth and Hayes-Roth 1979) has not been investigated. However, one may argue that some of the goals derived from the physicians' protocols are the result of opportunistic decisions made by the physicians.

The planning process and goal interactions are related. In most AI planners, a search space is defined and the planner seeks a point in that search space, which is defined as a solution. In other words, given some operators, the planner tries to produce a plan that achieves the goal state from the initial state. This is done by using methods such as means-ends analysis, depth-first backtracking etc. In the context of medical diagnosis, the goal state is the diagnosis of the patient and the initial state is the patient case. However, the physician does not always know the exact nature of the diagnosis. The goal state is usually uncertain and the initial state has incomplete information. In DEMEREST, the search space is defined in relation to the search space of the reasoning strategies. The initial search space of the back pain domain is fairly large. By identifying which reasoning strategy has been applied and the goals related to it, the system can define points in the search space which are partially elaborated plans. The search space is transformed when another reasoning strategy (or interaction of strategies) is applied. Hence, the planning process is entirely controlled by the reasoning strategies used. In this approach the construction of the plan and its execution are intertwined, that is, the applied reasoning and the goals associated with them not only form a plan, but also reflect the execution of the plan.

9.2.2 DEMEREST and Planning for Tutoring

By having a plan of the student's protocol, the tutor is given the appropriate information to help the student understand her plan. The mapping of the student's reasoning strategies into a plan provides a representation of the student's view of the clinical problem. The student has the possibility to reflect on her current plan and the tutor can assist her in manipulating and adjusting goals of her plan in various ways.

Since a plan is a product of applying one or more strategies, the tutor can present to the student the instantiation of a strategy which has been applied, that is, the goal upon which the student was focussing, the hypothesis generated, the symptom or sign considered etc. Teaching what is a strategy is not enough; what is also important to the student is to understand the context in which a strategy was applied.

Moreover, by combining different viewpoints of the diagnostic process (for a given patient case), that is, different levels of expertise, the student's understanding of that process might be improved. The tutor can help the student compare her plan with other plans from different levels of expertise. For example, the student might want to view how the history taking was conducted at the house officer level and relate her plan with the house officer's plan. The tutor might not teach planning to the student but rather might use planning as a way to analyse and decompose the diagnostic process.

9.2.3 DEMEREST and Developmental Models

This section discusses the system with respect to some of the developmental models reviewed in chapter four, that like DEMEREST, attempt to model the development of expertise.

Integrating various models of expertise

The approach to build a series of models to model the progression of expertise is not new and has already been investigated. For instance, QUEST (White and Frederiksen 1986) contains successions of mental models that correspond to increasing levels of expertise about electrical principles. Unlike those in DEMEREST, the causal model progressions in

QUEST are more complex and complete, and each model in QUEST contains a tutor, a student model and the domain simulation. However, in both systems a similar view is shared - that the transition from a novice to expert is regarded as a process of model evolution. In the case of DEMEREST, the evolution is constrained to a set of reasoning strategies for carrying out a medical diagnostic task.

Reasoning strategies as part of the developmental process

The genetic graph contains similar reasoning processes found in DEMEREST. These are generalisation, specialisation and refinement. However, in the genetic graph these reasoning processes are represented as genetic links between procedural rules. In comparison, in DEMEREST, *interactions* between these processes are considered part of the development of the (medical) diagnostic process.

Development of medical expertise

As discussed in chapter four, Lesgold's work has focused on the development of a perceptual skill in medicine and on the importance of organised knowledge in the development of medical expertise. In contrast, the design of DEMEREST has focused on the changes of reasoning strategies which seemed important for the development of medical expertise. But while Lesgold's work has stayed at a descriptive level, DEMEREST is an attempt to implement computationally some changes of medical reasoning that occur at various levels of expertise. As the next section will discuss, the implementation of the system is meant to demonstrate the feasibility, its usefulness and the potential of modelling various levels of medical reasoning.

9.2.4 DEMEREST: A Prototype

In this section, strengths and weaknesses of the system, and its current state of implementation are discussed. The system DEMEREST is a *prototype* which has allowed us to investigate the issue of development of expertise and implement development of medical reasoning strategies in terms of interactions of these strategies from one level of expertise to another.

Strategic knowledge in DEMEREST

In DEMEREST, strategic knowledge (i.e. reasoning strategies) is made explicit. The reasoning strategies are descriptions of what physicians do, and are represented in a declarative way. The focus on the reasoning processes in DEMEREST relates to other work (Cohen 1987, Gruber 1989) on generic tasks (also referred to as abstract tasks). The idea of a generic task is that classes of problems are characterised by the kinds of knowledge and strategies they require. Problem solving is viewed in terms of abstract tasks. For example, generic tasks for diagnostic problem solving describe how the diagnosis is made e.g. by hierarchical refinement and have specific knowledge requirements that depend on how the problem is solved.

Knowledge is not just required in problem solving context, strategies are important as well. Reasoning strategies are part of expertise and are as important as substantive medical knowledge. It is not enough to know everything about back pain problems, unless one also knows how to use this knowledge in efficient and useful ways. The move towards generic/abstract tasks emphasises how tasks are solved; hence it has implications for knowledge acquisition as demonstrated with the system

ASK (Gruber 1989). It also has implications for the development of expertise.

In DEMEREST, a number of task level primitives i.e. reasoning strategies and their interactions have been modelled. Each interaction corresponds to a change in the reasoning. Each interaction of reasoning strategies can be viewed as an abstraction of the diagnosis task from a developmental perspective. For example, the abstraction related to the interaction between hypothesis generation and elimination is that "in order to eliminate a hypothesis, one needs to know which evidence is needed to generate that hypothesis in the first place". Associated with each interaction of strategies is a goal which conveys the context in which the interaction took place. Therefore, for each abstraction of the diagnostic task, the system knows which decision has been considered.

Common knowledge

As mentioned in chapter five, common knowledge among the levels of expertise was assumed since some pieces of information may be used at more than one level. However, the use of common knowledge was found to be a problem with the strategies SPEC and GEN because in some cases the application of these strategies triggered more strategy than was required. For instance, some knowledge e.g. kinds(hypothesis, infection) may be used in levels 1 and 3. At each level, this knowledge may be used differently e.g. for a hypothesis generation strategy at level 1 and for a specialisation strategy at level 3. The use of common knowlege means that SPEC will be applied at the two levels whereas it should be applied at level 3 only. An ad hoc solution which makes the knowledge specific to the strategy to be applied was used to deal with this situation.

Generating a plan

The system always succeeds in achieving a goal either via the default or the deviation routes. Hence the system does not incorporate a mechanism of backtracking to find another goal if the current goal cannot be achieved. Moreover, in the system one goal is to be satisfied at a time.

Applying a strategy correctly

DEMEREST does not determine whether the physician has applied a strategy correctly. The system assumes that the evidence used is correct. An incorrect use of a strategy will result from the physician's misuse of the evidence. Different kinds of evidence that a physician might take into consideration were discussed in chapter six. It would be useful for an intelligent tutor to know whether or not the student applied a strategy properly in order to tutor the student accordingly.

Another shortcoming of the evidence used in an interaction of strategies is related to the equivalence of evidence within an interaction. That is, the evidence of one single strategy may also be used for the other strategy which belongs to the interaction. For example, in the interaction HGN and SPEC, the observation that is used is the same to apply HGN and SPEC individually. However, there may be cases where a different evidence needs to be used to generate a hypothesis and then specialise.

Interactions of reasoning strategies

The models of interactions of strategies (described in section 8.2) sometimes do not match exactly the order in which the two strategies interact (as found in the protocol). For example, in her protocol, the house officer applied hypothesis generation by generating the hypothesis *inflammatory cause* and then problem refinement to refine *palpation of* *the vertebrae* Given the model of changes of strategies, the system first applies the problem refinement strategy and then the hypothesis generation strategy.

9.3 Summary

In this chapter, the implementation of the system and the knowledge sources that it uses have been described. The system was constructed using half the protocols of the empirical study. Each protocol has been decomposed into a set of goals which correspond to diagnostic decisions about the patient case. Taking each goal at a time, the system checks whether there is one or more strategies and /or interactions of strategies associated with the goal and generates the instantiations of the strategies. A combination of strategies for one goal corresponds to an interaction of strategies. The set of goals for each protocol forms a plan. At each level of expertise, the system generates a global plan with its goals and coresponding reasoning strategies and interactions of strategies.

The prototype system DEMEREST is a demonstration of a first step towards a possible developmental user model for medical reasoning. In the following chapter, an evaluation of DEMEREST that was carried out with the remaining protocols of the empirical study is reported.

Chapter Ten

TESTING DEMEREST

Chapter nine discussed how the system DEMEREST diagnoses a physician's reasoning strategies and their interactions, determines her level of expertise and generates a physician's plan. This chapter reports on a testing of the system using the other half of the data collected from the empirical study. The methodology of the testing phase is first presented and the results are then discussed.

10.1 Methodology

The purpose of the testing has been to investigate whether the reasoning strategies and the model of interactions of strategies incorporated in the system capture the data from the empirical study. The model of interactions of strategies was implemented on a small number of subjects (n=5) and each level of expertise was constructed from a single subject. The system is being evaluated on its performance 1) for a given protocol, to determine a level of expertise, 2) model the reasoning strategies, 3) model reasoning strategies interactions and 4) generate plans corresponding to the protocols. The following sections report on the system performance given these criteria as well as on problems that arose during this testing phase.

10.1.1 Inputs to DEMEREST for the testing

The inputs to the system for its testing are the five other protocols from the study. Due to limitations of memory space using LPA Prolog on a Macintosh SE, it was difficult to include the evaluation protocols in

DEMEREST. Consequently, DEMEREST was duplicated into another program called EVALUATION. The medical knowledge and the goals specific to the protocols that were used to build the system were removed. The programs to generate the strategies and interactions of strategies and the plan were retained. The goals of the protocols to test and the medical knowlege specific to these protocols were input into EVALUATION.

In the following sections, the protocols that were used for the testing phase are presented, their consistency as data are discussed and finally, the predictions of performance of the system given.

10.1.2 Protocols used for the testing

The testing of the system was carried out using the other half of the data from the empirical study, that is, the protocols of the second 5th year student, the senior house officer, the trainee in general practice and the consultant in orthopaedics. An additional protocol of a 3rd year student was collected to have enough levels of expertise in the students' protocols for the testing. This extra protocol was obtained and analysed following the same methodology used with the other protocols. The protocols used for the testing correspond to various levels of expertise i.e. novices, intermediates and more experienced physicians.

10.1.3 Consistency of the protocol categories

The same independent assessor that had validated the first half of the data (Data 1) also validated the protocols for the testing (Data 2). The role of the independent judge was once again to assess the low level categories (i.e signs, symptoms, hypotheses etc) in the same way that it was done with Data 1. She had in fact coded the protocols that were used for the testing at the same time as she coded the protocols that were used to build the system. She then coded the new protocol of a 3rd student at the time of the testing phase. In addition, she was given a list of goals and the drawing of the global plan (appendix A8) and was asked to associate goals from the list of goals with part 1 of the protocol i.e. consultation between doctor and patient (see appendix D1 for the instructions). That list of goals was constructed from Data 1 and the purpose of validating the goals was to assess the plausibility of these goals (see appendix A3 for the list of goals).

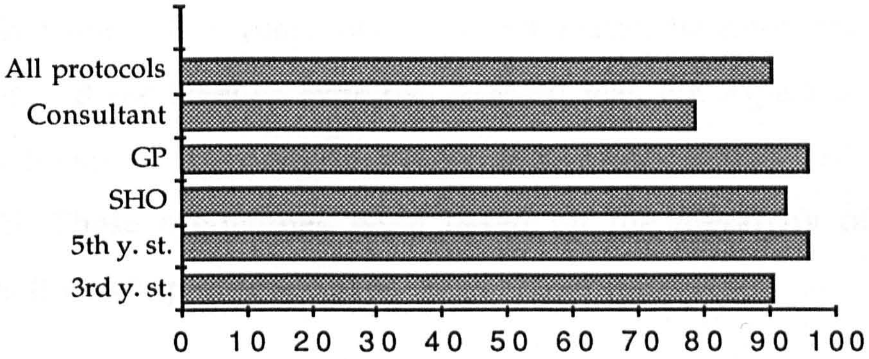
The differences between my assessment and the independent judge's assessment regarding the goals were based on the following:

1) new goals added to the plan by myself only
2) new goals added to the plan by the independent assessor only (found in appendix D2)
3) new goals that the independent assessor and I added to the plan, but that were placed in a different part of the plan
4) goals selected by myself only
5) goals selected by the independent assessor only

In some cases the independent assessor created a new goal e.g. *check recent activity* while there is an already existing goal *check recent occupation* describing it. In other instances, a goal was mistaken for another one. For example, the goal *check location pain* was used instead of *check radiation pain*. The goals that were renamed by the assessor can be found in appendix D3.

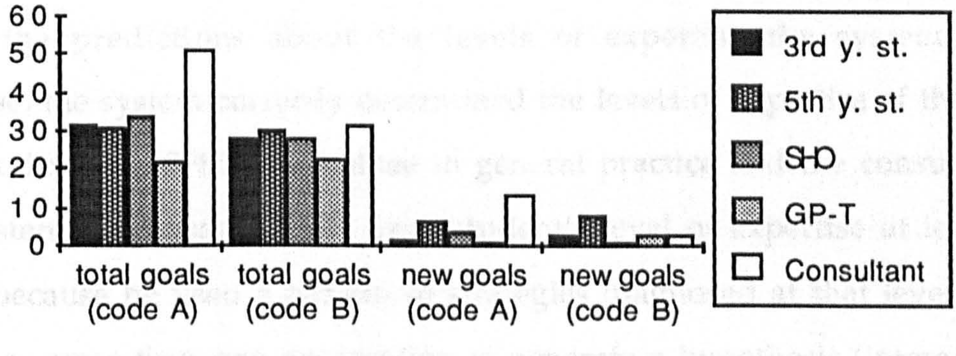
The percentage of agreement per protocol and for all the protocols is quite high (see figure 10.1) and is taken as evidence of consistency of the goals. This result also demonstrates the plausibility of these goals.

Figure 10.1: Agreement in %



For each protocol, the percentage of agreement of the assessment was calculated by adding the total of goals and new goals that I selected with the total of goals and new goals that the independent assessor selected and then dividing that total into the number of differences resulting in the percentage of difference. The percentage of agreement was then calculated by subtracting 100% from the percentage of differences. Figure 10.2 shows how each protocol was checked in terms of the goals by myself (code A) and by the independent judge (code B). The major difference is with the consultant's protocol. This difference may be explained by the fact that the independent assessor had less experience of medical protocols.

Figure 10.2: Independent Assessement of the Goals



10.1.3 Predictions

Since the model of changes of strategies upon which the testing is made, was built from a few protocols, an exact match between the protocol evaluated and the level of expertise selected was not expected. Rather a range of levels of expertise in which each protocol may fall into was predicted. These predictions were based on the hierarchy of medical expertise. It was hypothesised that:

i) The 3rd year student's protocol stays at novice levels (levels 1 or 2).
ii) The 5th year student's protocol be either at a student's level (level 3) or at the intermediate level (level 4).
iii) The protocols of the SHO and the GP trainee would fall into the intermediate or more experienced levels (levels 4 or 5).
iv) The protocol of the consultant would reach the highest level of expertise modelled (level 5).

10.2 Results and Discussion

In discussing the results of the testing, one needs to make the distinction between implementation problems and conceptual problems inherent in the design of the system.

10.2.1 Diagnosing of the levels of expertise

Given the predictions about the levels of expertise the system will diagnose, the system correctly determined the levels of expertise of the 5th year student, the SHO, the trainee in general practice and the consultant. The system diagnosed the 5th year student's level of expertise at level 4 (HO) because he used a pattern of strategies diagnosed at that level that combines more than one observation to generate a hypothesis (interaction of HGN's). The student also applied an interaction of strategies from his level i.e. HGN ELIM. The SHO, the trainee in general practice and the consultant applied an interaction of strategies specific to level 5 i.e. HGN CONF.

DEMEREST made one mistake in diagnosing a level of expertise with the 3rd year student's protocol. The 3rd year student had a level of expertise 3 and not 1 or 2. However, the student's level of expertise remained at the level of student, and not a higher level.

10.2.2 Modelling of the reasoning strategies

DEMEREST was able to model all the reasoning strategies except for the anatomically based strategy in the case of the 3rd year student. The way ANAT was defined allows one observation to be used i.e. from one observation an anatomical information is given, and not a list of observations. In the case of the student's protocol, a list of observations *(continuous_pain, dull_pain)* was necessary to generate the anatomical information *something pressing against the spinal cord.* The definition of ANAT could be easily altered to include a list of observations by adding a new definition to the predicate describing the strategy.

10.2.3 Modelling of the reasoning strategies interactions

Results on modelling the interactions of strategies are as follows:

i) The system was able to generate successfully the following interactions of strategies. Examples of these interactions generated by the system can be found in appendix D4.

- Interaction between HGN and SPEC
- Interaction between PREF and HGN
- Interaction between HGN and ELIM
- Interaction between PREF's
- Interaction between HGN's

ii) The system did not generate the interaction between ANAT and HGN since (as explained in the above section), ANAT was used in a way that the system could not handle.

iii) The system generates the following interactions successfully. However, in a few cases, the system could not generate the interaction of strategies accurately because it could not access the correct or necessary information e.g. observations or hypotheses.

- Interaction between HGN ELIM SPEC
- Interaction between HGN SPEC ELIM
- Interaction between HGN SPEC CONF ELIM

The following section examines the problems that the system had in generating the above interactions of strategies.

Interaction HGN ELIM SPEC

In this interaction, the pattern is (HGN obs1 h1 ELIM h1 SPEC h1 h2), where the specialisation is always on the hypothesis (h1) generated with HGN. In the SHO's protocol, the use of this interaction was found to be applied in a slightly different way with regards to the specialisation strategy, which the current implementation of the system cannot handle. The interaction HGN ELIM was applied twice i.e. (HGN obs1 h1 ELIM h1

HGN obs2 h2 ELIM h2), where from the observation *pulses_not_ok* the hypothesis (h1) *vascular problem* was generated. From another observation *pedal movements* the hypothesis (h2) *referred problem* was generated. In each case both the hypotheses were to be ruled out. The specialisation that occured afterwards was to be between h1 and h2. However, the system cannot do a specialisation between h1 and h2. The only specialisation that the system can do is with h2 and another hypothesis h3 e.g. *arteriosclerosis.*

Interaction HGN SPEC ELIM

In one instance, the application of this interaction triggered an extra strategy i.e. an elimination that was not in the physician's protocol. The consultant asked about the general health of the patient, and generated from the observation *not being perfectly well* the hypothesis *referred problem* (HGN), then specialised into kinds of *referred problem* such as *gastroinstestinal problems* (SPEC) and again specialised from *gastrointestinal problems* to *pancreatitis* (SPEC). The system applies ELIM twice whereas in reality the consultant only eliminated *gastrointestinal problems.* The system was able to generate this additional elimination strategy because the fact 'negative' containing the evidence *not being perfectly well* was present in the data base to trigger the first specialisation.

Interaction HGN SPEC CONF ELIM

The system failed to properly apply this interaction by producing redundant strategies. In the consultant's protocol the following interaction was found: (HGN obs h1 SPEC h1 h2 CONF h2 ELIM h2) where from the observation *indication of gyne problem* the hypothesis (h1) *gynaecological problem* was generated, and then specialised into the hypothesis (h2) *retroverted uterus* which was to be confirmed or disconfirmed. The system wrongly generated the same interaction twice.

Combining two interactions

The system can generate two interactions of strategies if one interaction is an extension or a part of the other interaction. For example, in the SHO's protocol, for the goal *check_pedal_movements,* the system generated the interaction (HGN obs1 h1 ELIM h1) where the hypothesis (h1) *vascular_problem* was generated from the observation *pulses_not_ok* and then was eliminated. The system applied the second interaction (HGN obs2 h2 ELIM h2 SPEC h2 h3) where the hypothesis (h2) *referred_problem* was generated from the observation (obs2) *pedal_movements* and then the hypothesis (h3) *arteriosclerosis* was generated from *referred_problem.* The second interaction HGN ELIM SPEC is an extension of the first interaction HGN ELIM. In contrast to the above case, for a given goal, the system cannot generate two interactions if they come from a different path (see figures 8.4 and 8.5 for reference). In the current implementation of the system, this problem was handled by creating a new e.g. *goal_more.* For example, in the SHO protocol, the interaction HGN CONF and the interaction HGN ELIM SPEC were associated with the same goal *check treatment previous back pain.* That goal was kept to generate the first interaction, and the goal *check treatment previous back pain more* was created to trigger the second interaction. The interaction HGN CONF and the interaction HGN ELIM SPEC are not in the same path of interactions.

Combining similar interactions

In some cases, the system generated successfully the same interaction more than once for a given goal. For instance, in the SHO's protocol for the goal *check_pedal_movements,* the system applied HGN ELIM twice. The two interactions were of the form (HGN obs1 h1 ELIM h1 HGN obs1 h2 ELIM h2) were the hypothesis *vascular_problem* (h1) was generated from the observation *pulses_not_ok* (obs1) and then eliminated. Then, the hypothesis *referred_problem* (h2) was generated from the same observation *can_hardly_move* and eliminated.

In other cases, the system could not generate the same interaction more than once for a given goal. This problem may be due to the fact that, in some cases, different kinds of knowledge need to be accessed for each run of the interaction. For example, an interaction may be applied a first time using the action slot of the goal e.g. details_action(Action, Details_action) and a second time the observation slot of the goal e.g. causes(Observation, Hypothesis).

For example, the program could not generate HGN CONF twice in the SHO's protocol. In this example, associated with the goal *check aggravating relieving factors,* two hypotheses were generated and confirmed *acute problem* and *mechanical backache* from the observation *aggravating relieving factors.* The system only generated and confirmed the first hypothesis. In SHO's protocol, HGN CONF was to be applied once using the fact causes(no_aggravating_relieving_factors, disc problem) and a second time using the fact causes(ask_aggravating_relieving_factors, mechanical_backache). The first fact was accessed by the system but not the second one. In other words, once the system has found an interaction for a goal, it does not backtrack to look for another one. In this particular case, causes(Observation, Hypothesis) is defined before causes(Action, Hypothesis) in the program and hence applied first.

Combining same interactions more than once for a given goal was also found to be a problem in the protocol of the 3rd year student. In this case, the system could not generate SPEC CONF twice for the goal *check_xrays*. The system generated only the hypothesis *slight displacement of the spine* and not the hypothesis *slight displacement of the vertebrae.* The problem also occurred in the protocol of the GP-T's protocol. The system could not generate the interaction HGN CONF twice for the goal *check_back_pain_history.* The system generated the hypothesis *disc_problem* but not the hypothesis *urinary_infection.*

Generating an interaction of strategies and single strategies

In most cases, the system could generate an interaction of strategies and a single strategy for the same given goal. For example, in the consultant's protocol for the goal *check_ultra_sound_kidney,* the following set of strategy and interaction of strategies (HGN obs1 h1 ELIM h1) (HGN obs1 h2) was found where the hypothesis (h1) *stone* was generated from the observation (obs1) *results_of_ultra_sound* and then the hypothesis was eliminated. Then, the hypothesis (h2) *partial_obstruction_problem* was generated from the observation *results_of_ultra_sound.*

In some cases, the system was not able to generate an interaction of strategies *and* other single strategies associated with the same goal. For example, in the consultant's protocol, the interaction (HGN SPEC ELIM) was applied for the goal *check_general_health_more.* The interaction was of the form (HGN obs1 h1 SPEC h1 h2 ELIM h2) and the single strategies of the form (SPEC h2 h3 SPEC h2 h4 SPEC h2 h5 SPEC h2 h6) where the hypothesis (h1) *referred_pain_problem* was generated from the observation (obs1) *not being perfectly well.* Then, the hypothesis (h2) *gynaecological_problem* was generated from *referred_pain_problem* and eliminated. From the hypothesis h2 other hypotheses (h3 to h6) were also generated. These are respectively *gynaecological problem, retroverted uterus, tumour, pelvic inflammatory disease* and *ectopic pregnancy.* The system failed to generate the hypotheses h4 to h6, possibly because it could not backtrack to search for these hypotheses.

Combining more than two interactions

Let us examine the following complex interaction that was found in the 3rd year student's protocol for the goal *check_xrays:*

i) HGN obs1 h1 SPEC h1 h2 CONF h2 SPEC h1 h3 CONF h3
ii) HGN obs1 h4 HGN obs1 h5 HGN obs1 h6 HGN obs1 h7
iii) PREF obs1 obs2
iv) ELIM h7
v) PREF obs1 obs4

where firstly from one observation (obs1) *results of xrays* the hypothesis (h1) *structural problem of spine* was generated and then specialised (into h2 and h3) into *slight displacement of spine* and *slight displacement of vertebrae,* in an attempt to confirm each of these.

Secondly from the same observation (obs1) *results of xrays* once more a number of hypotheses were generated (h4 to h7) *titling pelvis,* and *tumour, constipation problem* and *osteoporosis* Thirdly, from the same observation (obs1) *results of xrays* another observation (obs2) was produced to *refine degree of calcification.* Fourthly, one of the hypotheses *osteoporosis* generated was ruled out and finally from the observation (obs1) *results of rays* another observation (obs4) was produced to *refine full rectum.*

The way the model of changes of strategies is contructed indicates that the system cannot generate the combination of interactions as found in the protocol. For the same goal PREF needs to be applied before HGN. In addition, one of the hypothesis generation is combined with elimination since ELIM cannot be applied on its own. As a result this complex interaction was decomposed into three goals:

i) the first goal *check xrays* is associated with the strategies (PREF obs1 obs4) which *refines full rectum;* (HGN obs1 h1 SPEC h1 h2 CONF h2 SPEC h1 h3 CONF h3) which generate and confirm *slight displacement of spine* and *slight displacement of vertebrae.* As mentioned earlier on, the system failed to apply the specialisation and confirmation of h3.

ii) The second goal *check xrays more* is associated with the strategies (PREF obs1 obs2) which refines *degree of calcification;* (HGN obs1 h4 HGN obs1 h5 HGN obs1 h6) which generate the hypotheses *titling pelvis,* and *tumour, constipation problem.*

iii) The third goal *check xrays once more* is associated with the strategies HGN obs1 h7 ELIM h7 which generates and rules out *osteoporosis.*

This example clearly illustrated the difficulties in combining a number of interactions from one single goal. The model would have to be restructured if it was to contain combinations of interaction from one single goal. Recombining the interactions as it was done is not the proper solution since one looses the meaning of physician's intentions in the first place. However, it showed how it is possible to decompose a combination of interactions into single ones.

10.2.4 New reasoning strategies interactions

Interaction HGN CONF ELIM

In the process of modelling the interactions of reasoning strategies, the system could not generate interaction of strategies for which it did not have a description. One new pattern of interactions was found in the 5th year student's protocol. In the model of changes of strategies over time, confirming *and* eliminating a hypothesis h2 occurs after h2 has been generated using specialisation; the pattern is (HGN obs h1 SPEC h1 h2 CONF h2 ELIM h2). In the case of the student, the interaction occured for the goal *check urine sample,* with the observation *urine sample positive sugar,* and the hypotheses *diabetic condition* (h1) and *persistent infections related to diabetes.* (h2). In the student's protocol, the confirmation and elimination of the hypothesis was of h1 and not of h2 Thus, the new interaction of strategies required is (HGN obs h1 CONF h1 ELIM h1 SPEC h1 h2). Although the model of changes of strategies contains the interactions HGN CONF and HGN ELIM, they cannot be used for the same goal because they come from two separate paths. It should be pointed out that the confirmation and elimination of the hypothesis h2 is not necessarily incorrect since h2 is a child hypothesis of h1. Thus, by trying to confirm or eliminate h2, one may also try to confirm and eliminate h1.

This new interaction of strategies HGN CONF ELIM was also found in the GP-T's protocol and applied a couple of times. In the first instance, the goal was *check_episodic_continue_pain,* the observation *constant pain* and the hypothesis (h1) *musculo skeletal.* In the second instance, the goal was *check_aggravating_factors,* the observation *aggravating factors* and the same hypothesis h1.

Interaction HGN CONF ELIM SPEC

This interaction is an extension of the previous one in which the confirmation and elimination is of the general hypothesis rather than of the specialised one, or of both hypotheses. When faced with this interaction, the assumption was made that by generating a specific hypothesis, the physician also tries to confirm or rule out that hypothesis, and thus the interaction HGN SPEC CONF ELIM was applied instead. For example, in the consultant's protocol the physician generated the hypothesis *degenerative problem* from the observations *history of back pain* and *age forties;* and specialised to *tear in the anulous of the disc.* He tried to confirm or rule out both the general and the specific hypotheses. The system only applied confirmation and elimination to the hypothesis *tear in the anulous of the disc.*

Interaction HGN ELIM GEN

In the consultant's protocol, the new interaction HGN ELIM GEN that was found cannot be modelled since the generalisation is only triggered from the path of SPEC. In this example, from the observation *appetite habits* the consultant generated a number of hypotheses *ulcer, gall bladder disease, obstruction of gut, and pancreatitis,* and then generalised all these hypotheses as *gastrointestinal problems.*

10.2.5 Generating plans

Chapter nine discussed how the system generates a global plan for each protocol. Each plan includes the goals and their associated interactions of strategies. Similarly, the system produced a global plan for each protocol testing without any problem.

10.2.6 Some problems and limitations

In this section, some other problems which were identified during the testing phase are reported.

Order of the strategies in an interaction

In chapter nine, it was mentioned that in some cases the system does match exactly the order of an interaction of the protocol. A typical example that recurred in the 3rd year student's protocol is the interaction between PREF and HGN. In the protocol, the interaction is HGN PREF and in the model it is PREF HGN. Given a goal, the system first looks for an instantiation of PREF and then of HGN. In the case of the student, she generated from the observation *results xrays* a set of hypotheses such as *osteoporosis, tumour,* and then applied problem refinement by refining from the observation *results xrays.*

Differential Diagnosis

Each time a new hypothesis is generated, the system checks whether that hypothesis is in the list of hypotheses and if not, it is added to the list. In the case of the differential diagnosis, the system always adds to the differential diagnosis without checking if the hypothesis is already in the differential. This is why, in the case, of the GP-T the differential contains the hypothesis *disc problem* twice. This alteration could easily be done by using the same checking procedure for list of hypotheses. In one case the differential diagnosis was not well maintained. In the consultant's protocol, the differential wrongly contains the hypothesis *retroverted uterus* because the interaction HGN SPEC CONF ELIM associated with it was not applied properly (see section 10.2.4).

Using a goal just once

Once a goal has been active and its corresponding reasoning strategies applied, the system cannot call that goal again since the precursor slot of the goal is already filled in. For example, in the consultant's protocol the goal *check_previous_back_pain* was used twice. The first time at the beginning of the consultation the physician asked the patient about back pain she had before: "and you had never had back pain before" (part 1 p.1). The goal produced for that context is:

> Name of the goal: check previous back pain
> Precursor of the goal: onset pain this morning
> Subgoals: check_similar_pain, check_waterwork_infections_problems,
> check_treatment_previous_back_pain
> Action of the goal: ask about the previous back pain
> Effects of the goal: has history of previous back pain

The effect slot contains *history of previous back pain* since the patient told the doctor that she had back pain before. The second time, later in the consultation, the physician asked the patient again if she had that pain once before "... and you had only had this pain once before" (part 1 p.4). Although the physician did not ask exactly the same question he did however refer to previous back pain. By probing the physician regarding his repetition of similar questions to the patient, it was found that it is not an unusual technique that experienced doctors use. Patients do not always want to say things about themselves or they have forgotten about a particular event, and by probing them more than once the physician may eventually obtain the information she or he is looking for.

The temporary alternative to this problem has been to create an additional goal e.g. *check_history_previous_back_pain_more* for the goal *check_history_previous_back_pain*. A possible solution to this problem is to have the slot precursor containing a list of precursors rather than just one, to allow the system to call the goal more than once.

Grouping goals together

In the consultant's protocol, it was found that the physician grouped two goals together *check_past_illnesses* and *check_occupation* by asking the patient the reason for losing her job a couple of years ago:

> "Expert: was this because of illness or absence from work ?"
> "Patient: no, it was just redundancy. I was a secretary "
> (part 1, page 3)

The system cannot group goals together and instead it considers the two goals separately. One can either 1) model one of the goals to generate the corresponding strategy, or 2) model both goals, each generating the associated strategy. In this particular example, both goals *check past illnesses* and *check occupation* were input, each generating a problem refinement strategy to refine if the absence of work was related to the patient's back problems.

Details of the actions of the goals

In chapter nine, the need to include in the data base of medical knowledge, details about the action slots of the goals (in the form of details_action(Action, details_action)) was discussed. During the testing phase, two problems were uncovered using this kind of knowledge:

1) The use of details_action should be extended to include a list of observations, and not just a single observation. For example, in the 3rd year student's protocol, HGN was applied with the goal *check_palpation*

to generate the hypothesis *combination of trapped nerve and slipped disc* from the observations *[tenderness in left side, degree of leg elevation is less on left side]*. The system could not handle a list of observations and thus the two observations were combined as one.

2) The interaction (HGN obs h1 CONF h1) does not handle details_action since details_action is not included in the definition of CONF. Thus, in the case that a physician observation is defined through details_action (i.e. details_action(Action, Details_action)), the confirmation strategy cannot be applied. This is illustrated in the consultant's protocol where the physician observation *sciatic pain* defined by *details_action(ask radiation previous pain, ask previous sciatica pain)* would prevent CONF to be applied. In this particular case, an extra goal *check radiation previous pain more* was created with the effect slot as *previous_sciatica_pain*.

Evidence

The problem of dependence of evidence within one interaction of strategies was discussed in chapter nine. That is, with the interaction HGN SPEC the same evidence is used which in reality may not always be the case. For instance, the physician may used evidence e1 to generate a hypothesis h1 and then evidence e2 to specialise from h1 to h2.

This problem occured during the testing of system. For example, in the 5th year student's protocol, HGN was applied with the goal *check_kidney_problem* generating the hypothesis *abnormality of urinary tract* with the evidence *no left kidney*. SPEC was then applied generating *double ureters, ectopic ureters* and *kidney material lower down* using the same evidence *no left kidney*. The assumption made here is that the same evidence was used for applying HGN and SPEC. While this assumption may not be incorrect, there may be in fact other evidence that the student had in mind while specialising.

10.2.7 Outcomes of the testing

The aim of this testing was to assess whether the reasoning strategies and the model of interactions of strategies and incorporated in the system captures the data from the empirical study. As was outlined in section 10.1, the system was tested in terms of 1) determining a level of expertise for each protocol, 2) modelling the reasoning strategies applied 3) modelling interactions of strategies and 4) generating a plan for each protocol.

The results reported in the previous section demonstrate that as a prototype, the system can perform reasonably well. It should also be pointed out that the data (Data 2) used for the testing includes more protocols from experienced doctors (GP-T, SHO and consultant) than the data (Data 1) that was used to build the model of interactions of strategies. Data 1 contained the protocol of a GP as its most experienced level and the protocol of a HO level as its intermediate level. The fact that with a few exceptions the system generated correct strategies and interactions of strategies from data of higher levels of expertise, is considered a positive and encouraging result in designing a system that contains various levels of expertise embedded within one another.

The problems and limitations of the system that were identified during the testing phase not only helped in testing the system but also served in debugging the prototype version. Some of the problems discussed were conceptual related to the design of the system such as the order of strategies in an interaction, the use of a goal more than once and grouping goals together. Others were implementation considerations such as the maintenance of the differential and the use of details_action facts.

10.3 Summary

In this chapter, the testing of the system DEMEREST was reported. The protocols used for this testing came from the empirical study, specifically, half of the data was used to build the system and the other half for its testing. Protocols for the testing correspond to various levels of expertise i.e. medical students, senior house officer, trainee in general practice and consultant in orthopaedics. The aim of this testing was to assess whether the reasoning strategies and the model of interactions of strategies incorporated in the system capture the data from the empirical study. The system was tested in terms of 1) determining a level of expertise for each protocol, 2) modelling the reasoning strategies applied, 3) modelling interactions of strategies and 4) generating a plan for each protocol. Given the results of this assessment, the overall performance of the system was judged to be successful. In the following, each of the testing points is summarised:

1) *Determining level of expertise:* The system gave a level of expertise for each protocol similar to the ones that were predicted. In one case only the system underestimated the level of expertise of the 3rd year student. The student had a higher level of expertise than the one the system diagnosed.

2) *Modelling reasoning strategies:* The system was able to model all the reasoning strategies as defined in the system, except in the case of the anatomically based strategy since the current definition of this strategy does not deal with a list of observations.

3) *Modelling interactions of strategies:* The interactions of strategies that have been defined in the system were well applied. The interactions (HGN SPEC CONF ELIM), (HGN SPEC ELIM) and (HGN ELIM SPEC) need further testing, as in a few instances it was not applied properly. In addition, the testing showed that, for a given goal, the system can handle a combination of interactions only if they belong to the same path of interactions. A number of new interactions were identified at various levels of expertise: one new interaction (HGN CONF ELIM) was found in the protocols of the 5th year student and of the GP-T. Two additional interactions were found at the consultant level: HGN ELIM GEN and HGN CONF ELIM SPEC.

4) *Generating plans:* For each protocol of the testing, the system generated a global plan that corresponds to the reasoning strategies applied during the consultation.

The testing of DEMEREST ends the research work which has been presented in this thesis. The next and concluding chapter examines what has been achieved in this research work. In particular, the contributions of modelling medical reasoning processes from a developmental perspective are discussed. Some research directions in this area which could be taken for further work are also proposed.

Chapter Eleven

CONCLUSIONS

The aim of the research work reported in the thesis was to investigate the development of medical reasoning strategies for student modelling for an intelligent medical tutor. A prototype system called DEMEREST was implemented to illustrate how this could be achieved. The system analyses a physician's reasoning strategies and their interactions, determines the physician's level of expertise and produces a plan corresponding to the application of these strategies. The reasoning strategies considered in the thesis were identified in the medical problem solving literature whereas changes of these strategies over time were examined from an empirical study. This last chapter summarises the achievements of the research work, outlines the contribution of the research in various areas, discusses the limitations of the research work and indicates some directions for further work.

11.1 Achievements

The achievements of this research are summarised in the following list:

- A literature review bringing together three research areas and an extensive and varied selection of research work was prepared.
- Specifications of an initial set of medical students' reasoning strategies were drawn up.
- A study of development of medical reasoning strategies was made.
- An empirically-based model of interactions of strategies over time was built.

- A set of plans based on patient/doctor consultation at various levels of expertise was constructed.
- A prototype system for modelling changes of medical reasoning strategies over time was implemented.
- A testing of the prototype was conducted.

Each of these achievement is described more fully in the following section:

A literature review bringing together three research areas and an extensive and varied selection of research work

The research work pursued in the thesis has taken an interdisciplinary approach by combining three research areas - ITS in medicine, medical problem solving and the development of expertise. Each of these research areas contains a number of relevant papers and a comprehensive review was undertaken for each area. Each research area review focussed on a specific issue: student modelling for medical tutors, students' medical problem solving and development of medical expertise. The combination of these reviews led to the specification of the design considerations required for a developmental student model for a medical tutor.

Specifications of an initial set of medical students' reasoning strategies

A set of seven strategies were identified in the medical problem solving literature. These are strategies applied by medical students. The literature only provided a general description of these strategies which is not sufficient for a system to recognise and analyse the reasoning strategies used by the physician (whether novice or experienced) during a consultation. Hence, the next step after identifying the strategies was to formalise these strategies with regards to their features.

Results of the empirical study showed that this set of strategies formed a coherent system of reasoning processes for medical diagnosis. In fact, the greater part of students' protocols (and of more experienced physicians) could be analysed given these strategies. There is no claim that these strategies are the only possible ones that students apply. It may be that students apply other strategies which have not yet formed part of the existing literature. This is an issue for further research which will be discussed in section 11.4.

A study of development of medical reasoning strategies

The empirical study of the development of medical reasoning strategies was carried out as the literature did not provide any experimental results or hypotheses about changes of these strategies over time. The empirical work not only identified the predefined strategies but also looked for interactions of these strategies at various levels of expertise.

An empirically-based model of interactions of strategies over time

The construction of a model of combinations of strategies at various levels of expertise is seen as an important achievement in the thesis. First, it is empirically based. Although one may argue that the model was built using data from a small number of subjects, the testing of the system indicated that the model is reasonably sound and identified a few constraints of the model. Secondly, this model shows that reasoning strategies are not applied in an independent manner. Rather interactions between these reasoning strategies are important in carrying out a medical diagnostic task.

Constructions of a set of plans based on patient/doctor consultation at various levels of expertise

A plan corresponding to the physician's medical diagnostic process was constructed for each level of expertise considered in the research work. Adopting a planning approach for medical diagnosis is not an innovation as indicated by related research in section 5.3.1. However, what has been achieved is the decomposition of the medical diagnostic task (at various levels of expertise) into a set of goals associated with the reasoning strategies. Each single plan constructed for a given level of expertise corresponds to a viewpoint of the diagnostic process (for a given patient).

Implementation of a prototype system

The system was built as a prototype. The implementation of DEMEREST shows the feasibility of modelling certain medical reasoning strategies and changes of these strategies over time.

Testing of the prototype

The testing helped in assessing how well the system captured the data from the empirical work and in identifying conceptual as well as implementational constraints of the prototype.

11.2 Contributions

This section examines the contributions of these achievements to various research areas - ITS in medicine, medical problem solving, medical education and the development of expertise.

11.2.2 A contribution to ITS in medicine

There are four areas in which a contribution has been made:

- modelling medical reasoning strategies,
- progress towards a developmental-based student modelling,
- planning for tutoring and
- specifications for a medical tutor.

Modelling medical reasoning strategies

Modelling medical reasoning processes have been achieved in other systems such as NEOMYCIN (Clancey 1985). In fact, some of the strategies considered in this research are similar to the ones in NEOMYCIN (as discussed in section 6.5). However, the particularities of the set of reasoning strategies in DEMEREST are i) their sources in the literature and their descriptions complemented through discussions with a medical doctor and ii) their formalisations. The strategies modelled in DEMEREST were identified as strategies applied by medical students whereas strategies in NEOMYCIN were expert physicians strategies. In addition, the formalisations of the strategies in DEMEREST does not contain a meta-level like the strategies in NEOMYCIN.

Progress towards developmental-based student modelling

The notion that student models in ITS should capture developmental processes is an important current research issue. The case of medical tutors is a clear example where an expert-based approach to student modelling has been predominant. In addition, research on the development of medical expertise for student modelling has been very limited and has been confined to a theoretical level. Hence, while modelling development of expertise for student modelling is not new and its application to ITS in medicine has been poorly investigated, its computational implementation is innovative. Moreover, the focus on the development of medical expertise in terms of a set of reasoning strategies is also novel.

As mentioned before, DEMEREST is considered a prototype system. In that perspective, the system is viewed as a step towards a developmental-based student model for medical tutors. Section 11.4 will discuss future directions in which to extend the prototype.

Planning for tutoring

The role of planning in intelligent tutoring systems has already been investigated as the discussion in section 5.3.1 showed. However, planning in ITS has usually considered the goals and plans of the teacher rather than the goals and plans of the students, as is achieved in DEMEREST. Planning for tutoring involves not only the student model component but also the whole tutoring system and hence is related to the specifications for a medical tutor which are discussed below.

Specifications for a medical tutor

The thesis has reported on the implementation of a student model for a medical tutor. As Self (1988) points out, any proposed feature of a student model should specify how the student model would be linked with the tutoring system. The thesis has outlined the role that the proposed student model could play in a medical tutor, and in particular, has examined how the planning approach could be used for tutoring medical diagnosis.

The integration of the proposed student model into a medical tutor is linked to the specifications for designing the medical tutor. In the case of DEMEREST, the medical tutor should relate to reasoning strategies and the decomposition into goals of the diagnostic process for teaching (as elaborated in sections 5.4 and 9.3.4).

Additional specifications that a medical tutor should have are related to feedback and self reflection on one's problem solving processes. A review of the teaching of medical diagnosis (chapter three) indicated that teaching methods have usually been oriented towards helping students in the learning of factual knowledge rather than providing feedback on their reasoning processes. A system like DEMEREST generates a detailed analysis of reasoning strategies that the student has applied. A medical tutor could use this information to provide the student with some feedback and help the student reflect on her performance regarding the applied strategies and the goals associated with them.

11.2.3 A contribution to medical problem solving

The thesis has made two contributions to medical problem solving:

- a focus on the form of medical problem solving, and
- interactions of medical reasoning strategies.

A focus on the form of medical problem solving

The medical problem solving literature has showed that the majority of research has explored the contents of medical problem solving (that is, the medical knowledge) rather than on its form (that is, the reasoning processes that supports that knowledge). Moreover, one assumption which prevails in the medical problem solving literature is that novice as well as more experienced physicians use the same reasoning processes and differ in the medical knowledge they possess. While it is clear that both aspects are important for understanding medical problem solving, the thesis has focussed on its form. By doing so, an attempt has been made to demonstrate that the above assumption may not always be true. The thesis has showed that while medical students as well as experienced physicians may use similar reasoning strategies, they do not always combine them in a similar way. This point is discussed further below.

Interactions of medical reasoning strategies

The interactions of reasoning strategies that have been identified are viewed as an important contribution to the the field of medical problem solving. These interactions are unheard of, and they not only demonstrate that strategies are not applied in an independent manner but also that combining reasoning strategies is a critical factor in the diagnostic process. The thesis reported on direct interactions as well as pair and multiple interactions. It is most probable that additional interactions exist. This point is discussed in section 11.4.

One result from the empirical study was that no monotonic development of these strategies was found. One may then suggest that the interactions of reasoning strategies may be in fact an alternative to a monotonic development since each strategy alone may not be enough to trigger a developmental change in the medical diagnostic process.

11.2.4 A contribution to medical education

The thesis has made one contribution to medical education:

- the role of reasoning strategies in teaching medical problem solving.

A review of the teaching of medical diagnosis (chapter three) has indicated that two of the problems in medical schools are i) inadequate feedback to medical students on their performance and ii) lack of time for students to reflect on and discuss their reasoning. These two problems are related to the kinds of reasoning that students adopt to carry out diagnoses. The suggestion put forward for medical education is of integrating in the curriculum teaching of reasoning strategies in an explicit manner.

One possible medium for teaching medical diagnosis is computers. The usefulness of computers in medical schools has already been recognised (e.g. Chard 1988). A system like DEMEREST has demonstrated that reasoning strategies applied by a student can be identified and analysed and a medical tutor could use this information to teach these strategies to the student. It is clear that changes in the medical curriculum towards teaching of reasoning processes would require clarification of aspects of the curriculum such as the course aims, and course design. Nevertheless, what is suggested here, is a line of direction for a possible change in the medical curriculum either as a complementary or alternative course to the already existing courses.

11.2.5 A contribution to the study of development of expertise

The thesis makes a contribution to the study of development of expertise in the following way:

- a construction of an initial model of the development of some medical reasoning strategies.

The research work reported in the thesis makes a contribution to the field of development of expertise in several ways: firstly, it has put emphasis on the process of becoming expert. As Leventhal and Instone (1988) pointed out, this is an issue where relatively limited work has been done. The interactions of reasoning strategies at different levels of expertise correspond to changes in the medical reasoning. This is viewed as an initial model of development of reasoning strategies as it reports on some aspects of this development and provides a conceptual basis for the developmental picture, rather than a full descriptive model of a development of these reasoning strategies.

Secondly, the research work offers a new direction for examining the development of medical expertise by adopting a different focus from what was found in the literature. The review of the development of medical expertise (chapter four) indicated that research in that area has remained at a theoretical level and has focussed on the role of medical knowledge in the acquisition of medical diagnostic expertise. The developmental model proposed in the thesis has not only been implemented but also has focussed on the reasoning processes of medical diagnosis. Furthermore, while work reviewed on medical expertise has been concerned with the *acquisition* of the medical diagnostic skill, the present research has been concerned with its *development,* that is, the research work has not investigated how a physician acquires the reasoning strategies, but rather how she applies and combines them.

11.3 Limitations

The section examines some limitations of the research work. Suggestions for dealing with these limitations are taken up in the further work section.

Empirical study

i) The empirical study was carried out with a small number of subjects (n=10). Moreover, since half of the data were used to construct the system and the other half for the testing, the model of changes of strategies over time was built from a small number of protocols.

ii) The scale of expertise considered in the study was not complete, as no registrars or senior registrars were available for interview.

iii) The analysis of the protocols centered around the set of pre-defined reasoning strategies and did not, for example, focus on the misuse of the reasoning strategies by the physicians. Also it did not aim to clarify the kinds of evidence associated with the strategies.

Modelling development of medical reasoning strategies

Some of the levels of expertise in the model of changes of strategies over time incorporate very few interactions of strategies which indicate the changes from one level to another. For instance, the GP level includes two interactions that differentiate that level from the level below. In addition, the level of the 4th year student does not contain any new interaction.

Implementation

Some limitations of DEMEREST reported here are described in details in chapter ten. In particular, the system could not successfully:

- generate some interactions of strategies,
- generate new interactions of strategies,
- combine interactions of strategies if they do not belong to the same path of interactions,
- use a goal more than once and
- group goals together.

In addition, in the current version of DEMEREST, the system does not generate abstractions of plans since all the goals have the same level of abstraction.

11.4 Further work

The preceding sections have recorded the achievements of the thesis, the contributions of the thesis in a number of areas and outlined the limitations of the work. This section will present five possible directions for future work:

- extensions of the empirical study,
- extension of the model of changes of strategies over time,

- an improved implementation of DEMEREST,
- some issues for investigation arising from the prototype and
- application to other medical domains.

Extensions of the empirical study

1) All the descriptions of the strategies were found to be satisfactory except for the anatomically based strategy. There were two problems with this strategy: i) subjects did not verbalise anatomical information as much as was expected during the interview and ii) the description of the strategy was not satisfactory to account for complete anatomical reasoning.

2) The analysis of the protocols did not take into account the kinds of evidence physicians apply. That is, no differentation was made on whether the evidence used was correct. It would be useful, particularly in the context of tutoring, to know the kinds of evidence that are associated with the applied strategies. The classification of evidence discussed in chapter six could be a starting point to carry out this investigation.

3) The collected protocols contain very rich data. They include a large amount of information (e.g. the doctor's reasoning, the dialogue between the doctor and the patient and so on). The analysis of the data centered only on the predefined set of reasoning strategies. A further analysis of the data could, for instance, reveal additional reasoning strategies that subjects apply which would extend the set of strategies already considered.

4) The study has not focussed on the kind of medical knowledge that subjects use when they apply strategies. One research direction would be to investigate

i) the medical knowledge associated with the application of strategies and combinations of strategies and ii) whether the mis-application of these strategies and combinations of strategies reflects an inability to use such strategy (or combinations of strategies) or an insufficient medical knowledge to be able to do so. For instance, a student may not apply combined strategy such as HGN ELIM either because she does not know how to use these strategies or this combined strategy, or because she does not have enough medical knowledge to apply this interaction.

5) Another possible extension of the empirical study would be to interview additional physicians so as to incorporate the levels of medical expertise that are missing (i.e. registrar and senior registrars) and also to help in altering the model of changes of strategies. This point is discussed further below.

Extension of the model of changes of strategies over time

1) Some of the levels of the model of changes of strategies over time do not contain a significant number of interactions of strategies. Collecting additional protocols would help to refine and elaborate each level with more interactions of reasoning strategies. One starting point would be to focus on the level which does not contain a significant number of interactions of strategies such as the GP level or the 4th year student. By interviewing more 4th year students and general practitioners, one could possibly identify new interactions of strategies for that level.

2) The research has focussed on direct interactions between strategies. In section 8.1.2, it was reported that some indirect interactions of strategies were also observed. Further research on indirect interactions could be to examine

the kinds of concepts that link two goals to a goal at a higher level of abstraction such as the characteristics of the pain.

An improved implementation of DEMEREST

Some possible improvements to the current version of the system are described below.

1) **Refining the implemented description of the anatomically based strategy.**

Two improvements for ANAT could easily be made. The first one, mentioned in chapter ten, concerns the extension of the strategy to handle a set of observations rather a single observation. The second improvement would be to include details_action in the definition of ANAT so that the distinction between patient observation and physician observation can be made when applying this strategy.

2) **Evidence used in applying a strategy.**

The system should be improved such that the interaction of strategies (i.e. HGN and SPEC) does not depend on the same set of evidence of individual strategy.

3) **Implementing the new interactions of strategies that were identified in the testing phase.**

This would be a first step towards the refinement of the model of changes of strategies. One of the new interactions is (HGN CONF ELIM). There are a number of possibilities for adding the new interaction to the model. One possibility is to add the new interaction at the junction of the strategy HGN (see figures 8.4 and 8.5 for reference). A second possibility is to integrate the new interaction into an existing path of interactions of HGN. One may notice that the interaction (CONF ELIM) is already present in the interaction (HGN SPEC CONF ELIM GEN). In the current version of DEMEREST, the system applies HGN and then applies *only* one of the paths e.g. (ELIM SPEC) or (SPEC CONF ELIM). In order to have the system apply (HGN CONF ELIM) in the path of (HGN SPEC CONF ELIM GEN), the control of some interactions would need to be altered. Integrating the new interaction within an existing path of interactions is not always possible. For instance, one may notice that the interaction HGN CONF already exists at level 5. Extending HGN CONF with the new interaction will produce a conflict in setting up the level of expertise since the new interaction HGN CONF ELIM belongs to level 3 and the interaction HGN CONF to level 5.

The two other new interactions belong to a non existent level i.e. the consultant level. One new interaction (HGN CONF ELIM SPEC) is an extension of the previous one. The other new interaction is (HGN ELIM GEN). These interactions could form the basis of modelling the new level of expertise.

4) **Combining interactions.**

As discussed in chapter ten, for a given goal, the system cannot combine interactions if they do not belong to the same path of interactions. The short term solution that was adopted during the testing of the system was to create an additional goal e.g. *check_flexion_more* to associate the second interaction. This solution did not involve changes in the coding. An alternative could be to alter the ways the strategies interact so that the system would generate all the possible interactions. This could be achieved, for instance, by *not* preventing backtracking once an interaction has been found.

5) The use of a goal more than once.

Possible changes to enable a goal to be used more than once would include i) deleting the USED predicate that assured that a goal is used only once and ii) adding the repetitive goal in the list of 'default goals' that indicates the order of the goals.

6) Abstractions of plans.

As mentioned in chapters five and nine, all the goals have the same levels of abstraction and thus the system does not generate abstractions of plans. A possible extension would be to have the system generate abstractions (simplifications) of the plans. The slot subgoal in the goal structure contains the subgoals of the goal. These subgoals could be used to generate the abstractions of the plans.

7) Phases of the consultation.

The phase of the consultation in which a strategy has been applied was recorded in the analysis of the protocols but not implemented. The extension of the program to include the phases of the consultation could be achieved by, for instance, adding to the goal structure a slot 'phase' and inserting the corresponding phase in each goal.

Some issues for investigation arising from the prototype

1) In the current implementation of the system, the interpretation of the protocols is done manually. As mentioned in chapter five, this issue is related

to the problem of plan recognition. In the context of DEMEREST, the problem addressed would be to implement a protocol interpreter that would automatically generate the goals of the physicians' protocols. A starting point in carrying out this task would be to examine the existing plan recognition systems (e.g. Woodroffe 1988 for a review) and to draw from the techniques used in these systems. One could, for instance, consider each question/answer in the protocol (doctor/patient interaction) as the input unit and encode it as an act schemata as in BELIEVER (Schmidt et al 1978).

Another possibility would be to examine the work of Jansweijer et al (1982), which describes a protocol diagnostic program (PDP) for problem solving in physics. The interesting feature of PDP is that it has been developed as a tool to be used in the analysis of think aloud protocols of subjects solving elementary physics problems. There may be some analogue possible for medical diagnosis.

2) Adding a graphical representation of the plans that the physician has used along with the goals and their associated strategies would be very useful. Other systems such as GUIDON-WATCH (Richer and Clancey 1987) makes extensive use of graphics which has led to the enhancement of the learning process of the student (see chapter two).

Application to other medical domains

The domain of orthopaedics was chosen to construct the system. However, since the strategies that were identified in the medical problem solving literature were not specific to this medical domain, it is reasonable to believe that the system could be equally applicable in other medical domains. This is a research direction that could be further investigated. A number of medical systems which are built perform well in their domain of expertise, but not beyond other domains. If one aims to tutor the diagnostic process a student has used, it seems important to be able to do so across various medical fields.

11.5 Summary

Medical problem solving is a complex skill which has been extensively studied (as the review in chapter three showed). However, there are currently no formal paradigms of this task and as Evans (1989) points out

> "no uniform models of medical problem solving and none to explain ... the transformations in ability that characterize the progression from novice to intermediate to expert [physicians]. " (p10).

The research reported in this thesis represents an investigation of this progression in the context of intelligent tutoring systems. In particular, the research has focussed on reasoning strategies associated with medical diagnosis, and has demonstrated how these strategies and their changes at various level of expertise can be achieved for addressing the problems of student modelling. A prototype system called DEMEREST was implemented. The system can analyse a physician's reasoning strategies and their interactions, determine the physician's level of expertise and produce a plan corresponding to the application of these strategies.

Evans goes on to say

> "The obvious differences - such as amount of specific knowledge related to a problem- clearly play a role. But there are other differences - in approaching problems, in organizing information, in inference strategies - that must be understood." (p10).

It is clear that further research needs to be undertaken in order to better understand medical problem solving. This thesis represents a step towards understanding medical problem solving with regard to the applied strategies. If the aim is to build intelligent medical tutors that can teach medical students the process of medical diagnosis, one needs to understand the process itself as carried out by novice as well as expert physicians.

REFERENCES

Alpay, L. 1988a. A survey and examination of intelligent tutoring systems in medicine. **CITE Research report no. 42.** The Open University, U.K.

Alpay, L. 1988b. Medical problem solving and intelligent tutoring systems: proposed research directions. **CITE Research report no. 47.** The Open University, U.K.

Alpay, L. 1989a. Acquisition of reasoning strategies in medical diagnosis. **CITE Research report no. 81.** The Open University, U.K.

Alpay, L. 1989b. Developmental user models. **Interactive Learning International,** 5, 2, 79-86.

Alpay, L. 1990a. Investigating the development of reasoning strategies in medical diagnosis: a report on an empirical study. **CITE Research report no. 102.** The Open University, U.K.

Alpay, L. 1990b. Development of expertise and teaching of medical diagnosis. **Proceedings of 7th International Conference on Technology and Education** (ICTE), Brussels, Belgium, March 20-22. 1, 367-369.

Alpay, L. (in press). Alternate approaches to medical reasoning and diagnosis. In Moyse, R. and Elsom-Cook, M. (Eds) Knowledge Negociation. Chapman Publishers.

Anderson, J.R. 1982. Acquisition of cognitive skill. **Psychological Review,** 89, 4, 369-406.

Anderson, J.R. and Kosslyn, S.M. (Eds). 1984. **Tutorials in Learning and Memory.** New York: W.H. Freeman.

Anderson, J., Conrad, F. and Corbett, A. 1989. Skill acquisition and the LISP tutor. **Cognitive Science,** 13, 467-505.

Balla, J.I. 1988. Changing perceptions of clinical education. **Supplement To The Gazette, University of Hong Kong.** Vol. XXXV, no.1.

Balla, J.I, Gibson, M. and Chang, A. (Eds). 1989. **Learning in Medical School.** Hong Kong University Press.

Barrows, H.S. and Tamblyn, R.M. 1980. **Problem-Based Learning: An Approach to Medical Education.** New York: Springer.

Berner, E.S. 1984. Paradigms and problem-solving: a literature review. **Journal of Medical Education,** 59, 625-633.

Bordage, G. and Zacks, R. 1984. The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. **Medical Education,** 18, 406-416.

Bordage, G. 1987. The curriculum: overload and too general?. **Medical Education,** 21, 183-188.

Bretch, B.J. 1988. Deciding what needs to be said: Instructional planning for intelligent tutoring systems. **Research Report 88-6. ARIES Laboratories,** University of Saskatchewan, Canada.

Bretch, B.J. and Jones, M. 1988. Student Models: The Genetic graph approach. **Research Report 88-2. ARIES Laboratories,** University of Saskatchewan, Canada.

Britton, B.K. and Black, J.B (Eds). 1985. **Understanding Expository Text: A Theoretical and Practical Handbook For Analyzing Explanatory Text.** Hillsdale: Lawrence Erlbaum Associates.

Buchanan, B. and Shortliffe, E. 1984. **Rule-Based Expert Systems.** Addison-Wesley.

Buchanan, B. G. 1990. Personal communication during a visit in February 1990 at Intelligent Systems Laboratories (ISL), Pittsburgh, USA.

Burton, R.R. 1982. Diagnosing bugs in a simple procedural skill. In Sleeman, D. and Brown, J.S. 157-183.

Burton, R.R. and Brown, J.S. 1982. An investigation of computer coaching for informal learning activities. In Sleeman, D. and Brown, J.S. 79-98.

Carbonnel, J. 1970. AI in CAI: an artificial intelligence approach to computer-assisted instruction. **IEEE Transactions on Man-Machine Systems,** 11, 4, 190-202.

Chard, T. 1988. **Computing For Clinicians.** London: Elmore-Chard.

Clancey, W. 1979. Tutoring rules for guiding a case method dialogue. **International Journal of Man Machine Studies,** 11, 24-49.

Clancey, W.J. Methodology for building an intelligent tutoring system. 1984. In Kintsch, W., Miller, R. and Polson, P. 51-83.

Clancey, W. and Shortliffe, E.H. 1984. **Readings in Medical Artificial Intelligence: The first decade.** Reading, Mass.: Addison-Wesley.

Clancey, W.J. 1985. Acquiring, representing, and evaluating a competence model of diagnostic strategy. **Report STAN-CS-85-1067.** Department of Computer Science, Stanford University, USA.

Clancey, W.J. 1986. From GUIDON to NEOMYCIN and HERACLES in twenty short lessons: ORN final report 1979-1985. **AI Magazine,** 40-60.

Clancey, W. 1987. **Knowledge Based Tutoring: The GUIDON program.** Cambridge, Mass: MIT Press.

Cohen, P.R. 1987. The control of reasoning under uncertainty: a discussion of some programs. **COINS Technical report 87-81.** Department of Computer and Information Science, University of Amherst, USA.

Culter, P. 1985. **Problem solving in clinical medicine: from data to diagnosis** (second edition). London: Williams and Wilkins.

Diaper, D. and Winder, R. (Eds). 1987. **People and Computers III,** Cambridge: Cambridge University Press

deDombal, F.T. 1988. Computer-aided diagnosis of acute abdominal pain: the British experience. In Dowie, J. and Elstein, A. 190-199.

Dowie, J. and Elstein, A (Eds). 1988. **Professional Judgment.** Cambridge University Press.

Dreyfus, H. and Dreyfus, S. 1986. **Mind over machine.** Oxford: Basil Blackwell.

Elithorn and Banerjit (Eds). 1983. **Human and Artificial Intelligence.** London: North Holland.

Elsom-Cook, M. 1987. MADILDA and IMPART: Lisp tools. **CITE Research report no. 12.** The Open University, U.K.

Elsom-Cook, M. 1989. Dialogue and teaching styles. **CITE Research report no. 62.** The Open University, U.K.

Elstein, A.S., Shulman, L.S. and Sprafka, S.A. 1978. **Medical Problem Solving: An Analysis of Clinical Reasoning.** Cambridge, Mass.: Harvard University Press.

Ericsson, K.A. and Simon, H.A. 1984. **Protocol analysis: Verbal reports as data.** MIT Press, Cambridge, Mass.

Evans, D.P. 1982. **Backache: its Evolution and Conservative Treatment.** England: MTP Press Limited.

Evans, D. 1989. Issues of cognitive science in medicine. In Evans, D and Patel, V. 1-18.

Evans, D and Patel, V. (Eds). 1989. **Cognitive Science in Medicine,** MIT Press.

Feltovitch, P.J., Johnson, P.E., Moller, J.H. and Swanson, D.B. 1984a. LCS: the role and development of medical knowledge in diagnostic expertise. In Clancey, W.J. and Shortliffe, E.H. 275-319.

Feltovitch, P.J. and Patel, V.L. 1984b. The pursuit of understanding in clinical reasoning. **Paper presented at the Annual Meeting of the American Educational Research Association,** New Orleans.

Fieschi, D. 1984. Contribution au system expert SPHINX: application a l'enseignement medical. **These de 3eme cycle Informatique.** L'Universite Pierre et Marie Curie, Paris 6, France.

Fieschi, M. 1984. Le systeme SPHINX. In Fieshi, M. 75-127.

Fieshi, M. (Ed). **Intelligence Artificielle en Medecine.** Paris: Masson.

Fikes, R. , Hart, P. and Nilsson, N. 1972. Learning and executing generalized robot plans. **Artificial Intelligence,** 3, 4, 251-288.

Fox, J. 1988. Uncertainty plus logic and control: towards a theory for flexible decision making. **Paper presented to the Royal Statistical Society Medical Section,** London School of Hygiene and Tropical Medicine.

Friedlander, E.R. and Gillespie, H.D. 1987. Concept mapping: making learning meaningful. **Proceedings of the Annual Conference on Research in Medical Education.**

Gale, J. 1980. The diagnostic thinking process in medical education and clinical practice. A study of medical students, house officers and registrars with special reference to endocrinology and neurology. **Ph.D dissertation,** London University Institute of Education.

Gale, J. and Marsden, P. 1983. **Medical Diagnosis From Students to Clinicians.** Oxford: Oxford University Press.

Gale, J. and Marsden, P. 1984. The role of routine clinical history. **Medical Education,** 18, 96-100.

Gale, J. and Marsden, P. 1986. Medical students as adult learners: implications for an innovative short course in the clinical curriculum. **Medical Teachers,** 8, 3, 243-249

Goldstein, I.P. 1982. The Genetic Graph: A Representation for the Evolution of Procedural Knowledge. In Sleeman, D. & Brown, J.S. 51-78.

Gorry, G.A., Kassirer, J.P., Essig, A. and Schwartz, W.B. 1973. Decision analysis as the basis for computer aided management of acute renal failure. **American Journal of Medicine,** 55, 473-484.

Grant, J. and Marsden, P. 1987. The structure of memorized knowledge in students and clinicians: an explanation for diagnostic expertise. **Medical Education,** 21, 92-98.

Gruber, T.R. 1989. **The Acquisition of Strategic Knowledge.** Academic Press Inc.

Hartley, J.R. 1973. The design and evaluation of an adaptive teaching system. **International Journal of Man-Machine Studies,** 5, 2.

Hayes-Roth, B. and Hayes-Roth, F. 1979. A cognitive model of planning. **Cognitive Science,** 3, 275-310.

Hewson, M. 1986. Teaching and learning in medicine: a cognitive science approach. **Research Report No. 143,** University of Wisconsin.

Hunter, J. Cookson, J. and Wyatt, J. (Eds). 1989. **Lecture Notes in Medical Informatics.** Germany: Springer-Verlag.

Irby, D. 1986. Clinical teaching and the clinical teacher. **Journal of Medical Education,** 61, 9, 35-45.

Jansweijer, W., Konst, L., Elshout, J. and Wielinga, B. 1982. PDP: a protocol diagnostic program for problem solving in physics. **Proceedings of ECAI,** 278-280.

Jayson, M.I.V (Ed). 1980. **The lumbar spine and back pain.** Great Britain: Pitman Medical.

Jayson, M.I.V. 1981. **Back Pain: The facts.** Oxford: Oxford University Press.

Jayson, M.I.V. 1983. Back pain and sciatica. In Jayson, M.I.V. and Million, R. 159-179.

Jayson, M.I.V. and Million, R. (Eds). 1983. **Locomotive Disability in General Practice.** Oxford: Oxford University Press.

Jeffries, R. 1982. A Comparison of the Debugging Behavior of Expert and Novice Programmers. **Paper presented at the AERA Annual Meeting.**

Johnson, P.E., Duran, A.S., Hassebrock, F., Moller, J., Prietula, M., Feltovitch, P.J. and Swanson, D.B. 1981. Expertise and error in diagnostic reasoning. **Cognitive Science,** 5, 235-283.

Kintsch, W., Miller, R. and Polson, P. (Eds). 1984. **Methods And Tactics In Cognitive Science,** London: Lawrence Erlbaum Associates Publishers.

Kodratoff, Y. 1988. **Introduction to Machine Learning.** Pitman.

Kuipers, B., Moskowitz, A.J. and Kassirer, J.P. 1988. Critical Decisions under Uncertainty: Representation and Structure. **Cognitive Science,** 12, 177-210.

Langoltz, C.P., Fagan, L.M., Tu, S.W., Sikic, B.I. and Shortliffe, E.H. 1987. A therapy planning architecture that combines decision theory and artificial intelligence techniques. **Computers and Biomedical Research,** 20, 279-303.

Larkin, J., McDermott, J., Simon, D.P. and Simon, H.A. 1980. Expert and Novices Performance in Solving Physics Problems. **Science,** 208, 1335-1342.

Laurillard, 1988. The pedagogical limitations of generative student models. **Instructional Science,** 17, 235-250.

Laurillard, D. 1989. Understanding medical students' problem solving. In Balla, J. Gibson, M. and Chang, A. 119-129.

Lesgold, A.M., Feltovitch, P.J., Glaser, R., and Wang, Y. 1981. The acquisition of perceptual diagnostic skill in radiology. **Technical report no.PDS-1.** LRDC, Pittsburgh.

Lesgold, A.M. 1984. Acquiring Expertise. In Anderson, J.R. and Kosslyn, S.M. 31-60.

Leventhal, L.M. & Instone, K. 1988. Becoming an Expert?: The process of acquiring expertise among highly novice computer scientists (Pilot and Proposal). **Research Report.** Department of Computer Science, Bowling Green University, U.S.A.

Lindley, D.V. 1985. **Making Decisions** (2nd ed.) London: John Wiley & Sons.

London, B. and Clancey, W. 1982. Plan recognition strategies in student modelling: prediction and description. **Proceedings of the Second National Conference of the American Association for Artificial Intelligence,** 335-338, Pittsburgh.

Macnab, I. 1977. **Backache.** Baltimore: The Williams and Wilkins.

Manfred, E. and Rutheld, W. 1981. **Laws of The Game: How The Principles of Nature Governs Chance.** Allen Lane.

Marton, F., Hounsell, D.J. and Entwistle, N.J (Eds). 1984. **The Expertise of Learning.** Edinburgh: Scottish Academic Press.

Marton, F. and Saljo, R. 1984. Approaches to learning. In Marton, F., Hounsell, D.J. and Entwistle, N.J.

McGregor, J., Mansour, A., Franklin, M. and Poyser, J. 1988. An intelligent tutoring system for the instruction of medical students in techniques of general practice. **Research Report.** University of Sheffield, Department of Computer Science, U.K.

Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. (eds). 1983. **Machine Learning. An Artificial Intelligence Approach,** volumes I and II. Los Altos: Morgan Kaufmann.

Miller, P.B. 1975. Strategy selection in medical diagnosis. **Master Dissertation, MAC TR-153.** MIT, Cambridge.

Miller, P.L. 1984. **A Critiquing Approach To Expert Computer Advice: ATTENDING.** Research Notes in Artificial Intelligence. London: Pitman Publishers

Miller, R.A., Pople, H. and Myers, J. 1984. INTERNIST-1, an experimental computer based diagnosis consultant for general internal medicine. In Clancey, W. and Shortliffe, E.H. 190-209.

Miller, R.A., Massarie, F. and Myers, J. 1986. Quick Medical Reference (QMR) for diagnostic assistance. **Clinical Computing,** 3, 5, 34-48.

Miller, R.A. and Massarie, F. 1989. Use of Quick Medical Reference (QMR) program as a tool for medical education. **Methods of Information in Medicine,** 28, 4, 340-345.

Morgenstern, L. 1987. Knowledge preconditions for actions and plans. **Proceedings of the Tenth International Joint Conference on Artificial Intelligence,** Milan, Italy, 867-874.

Newble, D.I. and Clarke, R.M. 1986. The approaches to learning of students in a traditional and in an innovative problem-based medical school. **Medical Education,** 20, 267-273.

Norman, G.R. 1984. The role of knowledge in teaching as assessment of problem solving. Based on a paper entitled Cognitive Psychology and Medical Problem Solving presented at **The 68th Annual Meeting of the American Educational Research,** 1-17.

O'Shea, T., Bornat, R., DuBoulay, B., Eisenstadt, M. and Page, I. 1983. Tools for designing intelligent computer tutors. In Elithorn and Banerjit. 181-199.

Patel, V.L. and Frederiken, C.L. 1984. Cognitive processes in comprehension and knowledge acquisition by medical students and physicians. In Schmidt, H.G. and DeVoder, M.L. 143-157.

Patel, V.L and Groen, G.J. 1986. Knowledge based solution strategies in medical reasoning. **Cognitive Science,** 10, 91-116.

Patil, R.S., Szolovits, P. and Schwartz, W.B. 1982. Modelling knowledge of the patient in acide-base and electrolyte disorders. In Szolovits, P. 191-226.

Pauker, S.G. and Szolovits, P. 1984. Towards the simulation of clinical cognition: taking the present illness by computer. In Clancey, W.J. and Shortliffe, E.H. 131-159.

Park, Y-T., Tan, K-W., and Wilkins, D.C. 1989. ProHC: a knowledge based system with declarative representation and flexible control. **Technical Report.** Department of Computer Science, University of Illinois, USA.

Parker, R.C. and Miller, R.A. 1989. Creation of realistic appearing simulated patient cases using the INTERNIST-1/QMR knowledge base and interrelationship properties of manifestations. **Methods of Information in Medicine,** 28, 4, 346-351.

Peachey, D.R. and McCalla, G.I. 1986. Using planning techniques in intelligent tutoring systems. **International Journal of Man-Machine Studies,** 24, 77-98

Polson, M. and Richardson, J. (Eds). **Foundations for intelligent tutoring systems.** Lawrence Erlbaum.

Ramsden, P., Whelan, G. and Cooper, D. 1988. Some phenomena of medical students' diagnostic problem-solving. **Medical Education,** 22, 799-809.

Richardson, J. 1988. Directions for research and applications. In Polson, M. and Richardson, J. (Eds). 243-252.

Richer, M. and Clancey, W. 1987. GUIDON-WATCH: a graphic interface for viewing a knowledge based system. In Yazdani, M. 1, 373-412.

Rodolitz, N. and Clancey, W. 1989. GUIDON-MANAGE: teaching the process of medical diagnosis. In Evans, D. and Patel, V. 313-348.

Sacerdoti, E. 1974. Planning in a hierarchy of abstractions spaces. **Artificial Intelligence,** 5, 115-135

Schmidt, C.F., Sridharan, N.S. and Goodson, J.L. 1978. The plan recognition problem: an intersection of psychology and artificial intelligence. **Artificial Intelligence,** 11, 45-83.

Schmidt, H.G. 1983. Problem-based learning: rationale and description. **Medical Education,** 17, 11-16.

Schmidt, H.G. and DeVoder, M.L. (Eds). 1984. **Tutorials in Problem Based Learning.** Assen, Holland: Van Gorcum.

Self, J. 1974. Student models in computer assisted instruction. **International Journal of Man Machine Studies,** 6, 261-276.

Self, J. 1979. Student models and artificial intelligence. **Computation and Education,** 3, 309-312. Great Britain: Pergamon Press.

Self, J. (Ed). 1988. **Artificial Intelligence and Human Learning,** London: Chapman and Hall Computing.

Self, J. 1988. Bypassing the intractable problem of student modelling. **Proceedings of the conference on Intelligent Tutoring Systems,** 18-24, Montreal, Canada.

Sharples, M. and duBoulay, B. 1987. Knowledge representation, teaching strategy, and simplifying assumptions for a concept tutoring system. **Cognitive Science Research Report no.102.** University of Sussex, U.K.

Shijven, R., Talmon, J., Ermers, E., Penders, R. and Kitslaar, P. 1989. Machine Learning as a knowledge acquisition tool. In Hunter, J. Cookson,.J. and Wyatt, J. 37-44.

Shortliffe, E.H. 1976. **Computer based medical consultation: MYCIN.** New York: American Elsevier.

Shortliffe, H., Buchanan, B.G. and Feigenbaum, E.A. 1979. Knowledge engineering for medical diagnosis making: a review of computer-based clinical decision aids. **Proceedings of the IEEE,** 67, 9, 1207-1221.

Sleeman, D. and Brown, J.S. (Eds). 1982. **Intelligent Tutoring Systems.** London: Academic Press.

Stefik, M. 1981. Planning with constraints (MOLGEN: Part 1). **Artificial Intelligence,** 16, 111-140.

Suchman, L.A. 1985. Plans and situated actions: the problem of human-machine communication. **Xerox Parc Technical Report P85- 00005.** Xerox Corporation, Palo Alto Research Centers, California. U.S.A.

Szolovits, P. (Ed). **Artificial Intelligence in Medicine.** Westview Press.

Tate, A. 1975. Interacting goals and their use. Advance paper of **4th International Joint Conference on Artificial Intelligence,** Tbilisi, USSR, 4, 215-218.

Waddel, G. and Hamblen, D. 1983. The differential diagnosis of backache. **The Practitioner,** 227, 1167-1175.

Welbank, M. 1983. A review of knowledge acquisition techniques for expert systems. **British Telecom Research Laboratories.** Martlesham Heath, Ipswish, England: Martlesham Consultancy Services.

Wenger, E. 1987. **Artificial intelligence and tutoring systems.** Los Altos: Morgan Kaufmann.

Wilensky, R. 1983. **Planning and Understanding: A Computational Approach to Human Reasoning.** Reading, Mass.: Addison-Wesley.

Wilkins, D.C. 1988. Knowledge based refinement using apprentiship learning techniques. **Proceedings of AAAI.**

White, B.Y. and Frederisken, J.R. 1986. Progressions of qualitative models as a foundation for intelligent learning environments. **BBN Laboratories Research Report no. 6277.** Cambridge, U.S.A.

White, B.Y. & Frederisksen, J.R. 1987. Causal model progressions as a foundation for intelligent learning environments.**BBN Laboratories Research Report no. 6686.** Cambridge, U.S.A.

Woodroffe, M. 1988. Plan recognition and intelligent tutoring systems. In Self, J. 212-225.

Yazdani, M. (Ed). 1987. **Artificial Intelligence and Education.** Ablex.

Yob, G. 1975. Hunt the Wumpus. **Creative Computing,** September/October, 51-54.

Young, R.M. & Simon, T. 1987. Planning in the context of human-computer interaction. In Diaper, D. and Winder, R. 363-370.