MODELLING OF EXPERT NURSES' PRESSURE SORE RISK ASSESSMENT SKILLS AS AN EXPERT SYSTEM FOR IN-SERVICE TRAINING

by

Alan Hyslop MA, RMN, RGN

Being a thesis submitted for the degree of Doctor of Philosophy in the Faculty of Social Sciences, Department of Psychology. University of Glasgow, June 1988.

TABLE OF CONTENTS

CHAPTER 2 KNOWLEDGE ACQUISITION

CHAPTER 3 ANALYSIS AND MODELLING OF EXPERT KNOWLEDGE

CHAPTER 4 EVALUATION OF COGNITIVE MODEL PERFORMANCE

ACKNOWLEDGEMENTS

Several individuals and institutions have contributed to this project. I would wish to give acknowledgement ....

to Scottish Home & Health Department, Chief Scientist Office, for sponsorship through the award of the Nursing Research Training Fellowship and to Dr. Margaret Clark of SHHD,

to Dr. B.T. Jones, Superviser at base (Department of Psychology, University of Glasgow),
to Miss W. Logan, Superviser (formerly of Department of Health and Nursing Studies, Glasgow College of Technology),
to Dr. J.B. MacIntosh, Adviser (Greater Glasgow Health Board),

to the many nurses who have co-operated in various experimental tasks,

to Greater Glasgow Health Board and Ayrshire & Arran Health Board for permitting access to wards and departments,

to fellow student Iain Ritchie for writing the computer program which ran the Process Tracing experiment,

and not least to my wife Alison and children Jennifer and Robin for tolerance and encouragement.

DECLARATION

I declare that this project and thesis is my own work carried out within the normal terms of Supervision.

SUMMARY

In the nursing literature to date there have been no reported applications of 'cognitive simulation' nor of Intelligent Computer Assisted Learning. In Chapter 1 of this thesis a critical review of existing nurse education by computer is used to establish a framework within which to explore the possibility of simulation of thinking processes of nurses on computer. One conclusion from this review which is offered concerns the importance of firstly undertaking reliable study of nursing cognition. The crucial issue is that an understanding must be gained of how expert nurses mentally represent their patients in order that a valid model might be constructed on computer.

The construction of a valid computer based cognitive model proves to be an undertaking which occupies the remainder of this thesis. The approach has been to gradually raise the specificity of analysis of the knowledge base of expert and proficient nurses while seeking concurrently to evaluate validity of the findings.

Reported in Chapter 2, therefore, are the several experimental stages of a knowledge acquisition project which begins the process of constructing this knowledge base. Discussed firstly is the choice of the skill domain to be studied - pressure sore risk assessment. Subsequently, the method of eliciting from nurses top-level and micro-level descriptors of patients is set out. This account of knowledge acquisition ends with scrutiny of the performance of nurse subjects who performed a comprehensive simulated patient assessment task in order that two groups might be established - one Expert and one Proficient with respect to the nursing task.

In Chapter 3, an extensive analysis of the data provided by the simulated assessment experiment is undertaken. This analysis, as the most central phase of the project, proceeds by degrees. Hence, the aim is to 'explain' progressively more of the measured cognitive behaviour of the Expert nurses while incorporating the most powerful explanations into a developing cognitive model. More specifically, explanations are sought of the role of 'higher' cognition, of whether attribute importance is a feature of cognition, of the point at which a decision can be made, and of the process of deciding between competing patient judgements. Interesting findings included several reliable differences which were found to exist between the cognition of subjects deemed to be proficient and those taken as expert.

In the final part of this thesis, Chapter 4, a more formal evaluation of the computer based cognitive model which was constructed and predictions made by it was undertaken. The first phase involved analysis in terms of process and product of decision making of the cognitive model in comparison to two alternative models; one derived from Discriminant Function Analysis and the other from Automated Rule Induction. The cognitive model was found to most closely approximate to the process of decision making of the human subjects and also to perform most accurately with a test set of unseen patients. The second phase reports some experimental support for the prediction made by the model that nurses represent their patients around action-related 'care concepts' rather than in terms of diagnostic categories based on superficial features.

The thesis concludes by offering some general conclusions and recommendations for further research.

INTRODUCTION

An apposite passage from the Report of the Nursing Process Evaluation Working Group (Hayward 1986) can be offered to introduce this thesis .....

> "The thinking processes of nurses seems rarely to have been put forward as one possible contribution to success or failure in using the nursing process ..... what are the thinking processes used by nurses? ...... can a computer model be developed which will simulate effectively critical aspects of clinical problem-solving in nursing? ..... can such a model be used for teaching?" (p103-104)

The nursing literature would seem, at first sight, to have adequately addressed these questions. There are, for example, several reports of 'simulation programs' within the extensive literature on nursing Computer Assisted Learning (CAL). Furthermore, there has been considerable writing on the 'thinking processes used by nurses'. It might therefore be the case that these questions have already been addressed and that there is a clear route toward construction of a computer-based model of expert clinical thinking.

There is, however, a large catch which is paradoxically brought about by the 'stupidity' of the computer in that each step in the problem solving process requires complete specification if the model is to actually work. A computer makes no sense of an arrow on a diagram between, for example, 'assessment' and 'diagnosis'. If the assessment 'irregular pulse' is entered into the machine then it requires to have previously been told in precise detail what to conclude and how to proceed from that point. Considerable discipline is therefore imposed on the researcher by the goal of modelling on computer.

With this discipline in mind, then, it becomes clear that nursing CAL and knowledge of nursing cognition is currently inadequate for the construction of a computer simulation model. The justification of this statement will occupy many pages in the present thesis - the point for now, however, is that if the route toward model construction is not clear then it becomes necessary to begin with 'first principles' and to proceed in a stepwise manner until such a model is achieved.

This project, then, sets out to achieve this goal by choosing to model the cognitive processes of nurses who are assessing the risk of a patient developing pressure sores. Several factors influence this choice of 'problem domain'. These are discussed in more detail later; however it fits the purpose of this introduction to explain that the principal factor influencing choice was the apparent 'specifiability' of pressure sore risk assessment knowledge. Perhaps due to the prevalence of the problem or because of the relatively well-understood 'concrete' nature of pressure sores, it can be seen that choice of this problem confers considerable advantages for an exploratory project.

For similar reasons of specifiability, cognitive psychologists have chosen 'toy' problems for exploratory study of the cognitive processes used by subjects attempting to solve such teasers as the Chinese Tea Ceremony. Pressure sore risk assessment, however, is demonstrably not a 'toy' problem. In choosing a 'real world' problem to study, it can seen that advantage is confered in terms of greater liklihood of professional acceptibility of findings. Moreover, it can be argued that there is a more realistic generalisation of findings to other nursing problems - although it remains to seen if the methods and findings gained from study of pressure sore risk assessment can generalise to the many nursing cognitive problems which are of a considerably more 'nebulous' nature.

Early decisions such as choice of problem domain to study can therefore be seen to affect several aspects of a study. It quickly becomes apparent, for example, that the chosen 'real world' problem is considerably more complex than typical 'toy' problems. A further issue is that validity and acceptability of findings will be determined to a degree by showing that the model built has not been based solely on a single nurse whose cognition may be unrepresentative. These early decisons, and the points which arise from them, result in the need to explore alternate methodology to that used more commonly in cognitive psychology. It would not be possible, for example, to depict the many thousands of 'moves' made by several nurses assessing several patients using techniques such as State Space Analysis. The use of innovative methodology is an additional factor which demands that efforts are frequently made to demonstrate validity and reliability of findings.

This introduction, then, has sought to raise some the key issues which will figure large in the coming pages. Put simply, a detailed review of the CAL literature will be followed by a comprehensive report of the method whereby the knowledge of pressure sore risk assessment held by several nurses is studied. Concurrently, the narrative will cover the way in which a computer simulation was constructed of an 'average' nurse assessing one of her patients. It will be useful, however, to introduce not only the project in outline but also one of the key findings - that the cognitive component of nursing patient assessment is characterised by the goal of planning care rather than by the goal of making a categorisation (or diagnosis) of that patient.

To find that a nurse assesses in order to plan care seems intuitively correct and rather unremarkable. Nevertheless, if the decision making process is 'care-driven' then this is a finding which stands in contrast to the prevailing North American model where 'patient diagnosis' has become an assessment goal *in its own right*. Yet, if patient information is being collected to form a diagnosis then the nature and processing of this data will differ compared to the situation where care-planning is the goal. The two contender explanations can be formed into an empirical question - although it is interesting that the diagnosis model has been largely derived from theoretical transposition of the medical profession literature. Within the forthcoming pages, several lines of evidence will be brought together to support a reformulation of the understanding of nursing cognition which re-emphasises its care-driven nature.

CHAPTER 1 REVIEW OF LITERATURE

INTRODUCTION AND FRAMEWORK FOR REVIEW

To take a rather fanciful analogy, imagine for a moment a world without books – without, even, the written word. Up until now the spoken word only has been used to communicate information and ideas. All education has been effected by demonstration and explanation. In such circumstances the development of the ability to set words to paper (and the consequent ability of being able to read) would rightfully be hailed as a revolutionary educational medium in the sense that words could now be passed on from expert to novice and back again via paper.

The proclaimed advantages of this discovery for learners would doubtless include that the printed word could be used for passing on factual knowledge and for modelling the real world using a variety of different teaching strategies flexibly tailored to the needs of students. Furthermore, students would be able to learn independantly and at their own pace while being encouraged to develop critical thinking or even while being counselled. The advantages for teachers, on the other hand, would be that the printed word could be used to ensure uniformity of standards both in terms of what they teach and in terms of the written feedback they receive on what has been learnt by their students. Freed of the necessity of so much face to face contact, teachers would be able to develop better teaching packages, do research, or offer individual remedial work.

Precisely such claims and assertions have recently been made for the 'new' educational medium, not of learning nursing via the printed word but of learning nursing via the computer. See, for examples, the respected work of Bitzer & Boudreaux (1969), Conklin (1981), Hannah (1983), Mirrin (1983), Norman (1983). The nursing profession is being urged to acknowledge the claimed benefits of the 'computer as teacher' while funders of nursing education are being asked to sponsor development of this potential.

It could be that these claims and assertions are valid. On the other hand, it could be that they are not. It could be that there are damning criticisms to match each one of the claims – to return for a moment to the analogy of the invention of the written word, it could be that exponents of the previous educational system would have serious misgivings about altering their role. Perhaps the very 'newness' of the proposed media confers on it spurious virtue in a situation where high kudos surrounds the innovatory and the modern – certainly an observation which seems apposite to computers. Two points can be made. The first is that all of, or some of, or none of these things could be true for either position. The second point is that the onus must reasonably be on the innovators to provide evidence for the usefulness of their ideas. It is toward an elaboration on these points that the discussion must eventually turn.

A useful exercise prior to examination in depth of the literature is to construct a rather more general overview. Nevertheless the task of trying to achieve a grasp of the literature is by no means straightforward. The problem is not one of quantity - quite recently Norman & Townsend (1982) identified 50 relevant papers and since then there has been a steady growth in published work particularily consequent to the IFIA-IMIA Workshop on the Impact of Computers on Nursing (Scholes et al 1983).

The problem, rather, is one of specificity. Perhaps due to the relative novelty of computers in nursing education, the impression gained is that authors of papers feel it necessary to cover all aspects from historical development to implications for the future. The result is a succession of introductory papers - a phenomenon of which Hawkins (1978) similarily complained with regard to the literature on CAL in tertiary education where it is seldom that fundamental issues such as theoretical base and evaluation are addressed. Only recently, for example, have there been papers specifically pointing to the 'rather scant' literature on evaluation of CAL in nursing education (eg Koch & Rankin 1985).

What aspects of CAL, then, are covered by the nursing literature? With varying degrees of detail, the following areas are addressed and can be arranged in roughly hierarchical order according to amount of coverage........

- (most covered) The perceived importance and potential of CAL. (eg Hannah 1983)
- Prescriptions for producing CAL programs. (eg Grobe 1983)
- Assuring quality of CAL program design and presentation. (eg Parsonage 1986)
- The role of the nurse tutor *vis a vis* CAL. (eg Sweeney 1983)
- The place of CAL in nursing curricula. (eg Hassett 1984)
- Technological considerations. (eg Hoy,R. 1983)
- (least covered) Evaluation of CAL effectiveness. (eg Koch & Rankin 1987)

This list, while neither exhaustive nor systematically derived, will serve to communicate the 'flavour' of the literature. The most notable absentee from the list is any serious work on theoretical concepts in that only rarely is a passing reference made to a psychological theory of learning which a given program seems apparently to fit. This in fact is the key to the entire CAL issue:

it seems to be the available technology which drives the intitiative rather than a developed concept of learning being applied via an appropriate medium.

It is intended that this thesis, offered at an early stage in this review, will underscore much of the discussion below. For the moment it suffices to point out that if, in fact, the theoretical concept came first and the implementation subsequently then a much greater research-based emphasis on the literature could be expected with regard to assessment of learner needs and assessment of needs of the profession. In addition, much greater emphasis could be expected on model-driven implementation and testing of theory generated hypotheses. In reality, however, these emphases are virtually absent from the nursing literature. While general education CAL literature has recently begun to attend to underlying theoretical concepts (eg Jonassen 1985), the CAL in nursing work seems still to be characterised by a rather unquestioning pioneering excitement.

It is in fact almost impossible to find any dissenting voices in the nursing literature - the single exception being that of Townsend (1983). With the benefit of having witnessed the rise and fall of Programmed Instruction, Townsend offers observations which should be sobering but seem little heeded. Turning to the literature on CAL in non-nursing education, however, it is possible therein to find more commentators who are prepared to sound cautionary notes. Whiting (1985), for example, shows that CAL is in fact enjoying its second 'life cycle' of enthusiasm (due to cheaper equipment) and stresses the imperative of paying attention to mistakes made in the past. Diem (1982) futhermore suggests that the disappointment experienced by teachers who were promised much of first generation CAL is a strong contributory factor to reluctance in becoming 're-excited' this time round.

An overview of the nursing literature on ICAL is, simply enough, not possible since there are no fully reported applications of this development within the nursing educational field. The only possible contender, outside the present project, would seem to be the COMMES system (Evans,1983). This program will be reviewed below, for the moment, however, it suffices to note that COMMES might better be described as a computer-based dictionary of medical and nursing facts rather than a learning program *per se*. For references to ICAL, therefore, it will be necessary for this review to look to reported applications in other disiplines.

Despite the absence of Intelligent Computer Assisted Learning in nursing, however, there are several indications that the concept would be a useful one for the profession to adopt. During the review of these indications below, the distinction will be made between ICAL programs which are Artificial Intelligence based rather than cognitive science based. Hence, it will be argued, ICAL programs in general seem to offer a way round the problems of CAL in nurisng - but ICAL programs based on *a cognitive model of human nursing expertise* offer the greatest potential.

In conclusion, this review will discuss aspects and problems of CAL in nursing and aspects and problems of ICAL. It will be concluded that an ICAL program based on a cognitive model of nursing expertise seems to offer potential for avoiding these problems. Such a program will itself have many aspects. It will be argued that of these aspects the valid construction of that model is paramount. The remainder of the project, therefore, will concentrate on achieving that goal.

This overview of the CAL in nursing literature has been undertaken with the intention of introducing the critical issues which can now be explored in more detail. The scene has been set with much having been promised but with little in the way of scholarly back-up to these promises. The task now becomes one of searching for such underpinning as exists in the literature and, in the case of fruitless search, inferring from wider knowledge what that underpinning might be. In short, a critical review is required. Throughout this intended review the discussion of ICAL will continue in a compare-and-contrast fashion. When all of these aims have been achieved an overall evaluation can be offered on the value and role of the use of computers in nursing education.

A framework is required within which the literature on the computer as teacher of nurses can be reviewed. A search for such a framework within the literature is, however, unhelpful since little if anything in the way of review seems to have been published. Completely absent, apparently, is any work which could be classified as meta-review (review of reviews). Perhaps this is not in itself a criticism since an innovatory area might need time to mature into this level of scholarship - although computers have been used to teach nurses since at least 1963 (Bitzer & Boudreaux 1969). A framework, then, will have to be constructed from the starting point of 'tabula rasa' and open mind.

Several issues will need to be explored on this blank slate before a conclusion can be offered on the validity of any claims and counter-claims of both CAL and ICAL in nursing. These issues, which would be common to a review of any educational innovation, will take the form of questions to which one would 'need to know' the answers before offering conclusions. Issues such as .....

1. What form do the innovations take and what are their characteristics?
2. What theoretical constucts underlie each innovation?
3. What model of the learner and what model of the teacher is employed by each innovation?
4. What role will each innovation have within the existing teaching system?
5. What evidence is there for the effectiveness of each innovation?

To an extent these questions overlap. For example, the underlying theoretical constructs will come into models of learning and of teaching. It is intended, nevertheless, to examine how adequately the literature itself can provide the answers to these five 'need to knows' - thereby using the questions as a framework for organising the material to be covered. As will become plain, however, the literature is rarely sufficiently detailed to permit an adequate review-based exposition of these questions. It will be necessary, therefore, for the discussion to adduce what some of the answers may be.

FORM AND CHARACTERISTICS OF CAL AND ICAL

Describing the form that CAL takes could be achieved in a 'technical' fashion or in a 'blank slate' fashion. The former, more usual, product approach would rather endlessly lead to explanations of computer construction, computer functioning and, of course, to translation of the many and various terms in use. The latter approach would take its perspective more along the lines of examination of the process whereby the product came about. The perspective adopted in this discussion will lean more toward this 'process' approach since, quite simply, the concern of this review is in the product of the medium rather than the medium *per se*.

Straightforwardly, then, what does the observor see when sitting in front of a CAL program? Most obviously, a screen resembling a television and an interface comprising a QUERTY-style keyboard, although some of the more expensive American systems have keyboards arranged in a more logical fashion, or 'light-pens' which can be used to touch parts of the screen. The screen, which displays material to the user in the form of text or graphics, is commonly not of very high resolution - the result being that the text is of somewhat fuzzy quality. Drawings fare rather worse in that photograph quality is impossible, with the optimal standard being comparable to a somewhat amateurish version of the outline drawings found in anatomy books. Colour is nowadays possible but of a disappointingly garish nature. Sound also is possible but usually limited to robotic buzzes and beeps. Often there are supplementary instructions or teaching on paper which must be read in order to run the program.

What is being described is the current equipment in use such as the BBC series of computers which have been adopted as standard by the National Boards for Nursing, Midwivery and Health Visiting throughout the UK. The literature regularly extols the exciting new developments in computer technology, but this is the present and the foreseeable future. The stark picture sketched above is not, however, designed to downplay the media but rather to emphasis that there is little of a 'magical' nature in the presentation of CAL material and that consequently any evaluation should focus on content.

One further observation about CAL programs is apposite - their location. At best, the equipment is sited in a separate room within a College of Nursing and at worst in a corner of the library or even, in one instance in Scotland, at the back of a lecture theatre. However, the logic for siting these machines exclusively within Colleges in the first place is not articulated in the literature. Given that more and more hospitals have computer-based Patient Administration Systems in the wards, it might seem reasonable for CAL programs to have been made available as an extra facility on ward terminals.

CAL programs, especially the 'simulations' (see below), are after all *models* of the real world of nursing. Sheehan (1986) takes the definition that a model is a means of transferring a relationship from its actual setting to one in which it can be more conveniently studied. Furthermore, there is good evidence to suggest that adult learners learn best when they can relate theoretical concepts to the work situation (Manpower Services Commission 1986). Given that many programs set out to teach, for example, drug dosage calculation or patient assessment, it would seem sensible to make such programs available as near to the actual practice setting as possible.

CAL Defined and Types of CAL

To add to the definition of CAL offered earlier there are several other terms in common use in the literature. The main terms in use include....
Computer Aided Instruction, used by eg. Hoffer et al (1983),
Computer Assisted Instruction, used by eg. Collart (1973),
Computer Based Instruction, used by eg. Bitzer et al (1973),
Computer Based Training, used by NHS Training Authority.

Hannah (1983) feels that these terms, along with Computer Simulated Instruction, Computer Based Education, Automated Teaching and Computerised Instruction, are used interchangably and makes no more of it. Nevertheless it is perhaps worth making the point that such loose use of terminology can be taken at least to testify to lack of groundwork in the area. There is, for example, quite a difference between 'assisted' and 'based'. Moreover, it is not just the semantic differences which separate learning and education from training and instruction in that these various terms imply real variations in the degree of *passivity* expected of the student.

Perhaps what is being suggested here is that these four terms each accurately belong <u>either</u> to a Process <u>or</u> to a Product Model of education. The distinction between these models, recently elucidated by Sheenan (1986), can arguably be characterised by differences in the degree of active exploration of the knowledge domain expected from the student. Nevertheless, it is clear from matching the definitions used by workers in the field (eg 'learning' or 'training') to type of program constructed that there is no consistent distinction in practice which relies on a Process or a Product philosophy of learning.

The other major classificatory term used in the literature is Computer Managed Instruction (CMI). Seemingly interchangable with the term Computer Managed Learning, CMI is seen as related to but distinguishable from CAL by the additional <u>monitoring</u> role assigned to the computer (Hannah 1983). Monitoring is used here in the sense that the computer records and analyses the performance of students on computer-administered tests such as multiple choice questions. It is important to bear in mind that performing of this evaluative function was found by Norman (1983 b) to be the most common use of computers in nurse education in the USA. In this respect the student sits at the computer in order to be tested only after having first received traditional teaching on a given topic.

Two points arise from the finding that CMI is the most common use of CAL in the country currently most committed to the idea. Firstly, assessment of learning is of course integral to education but not the primary *raison d'etre* of education. Secondly, it would seem to be stretching the already loosened term 'instruction' to suggest that administration of a test is itself an educating experience in the sense that knowledge of the domain is improved. It is important to bear in mind that when CMI is defined (eg by Hannah 1983) as providing functions such as monitoring a student's progess, diagnosing learning needs or prescibing remedial work, what is in fact meant is that it is the teacher who is provided break-downs of test scores by the computer. The impressive-sounding functions are up to the teacher to provide or not to provide, making it perhaps more accurate to entitle this form of 'CAL' as Computer Assisted Assessment.

With respect to types of CAL, there are two influential classificatory schemes used with respect to CAL. The most commonly cited (originator unknown) is a five category format which Meadows (1977) and others have utilised to descibe and classify CAL programs in nursing. These categories are ......

1. **Page Turner** - the computer presents textual or graphical information to the learner. When one frame has been read a key is pressed for the next in a fixed series. Also known rather dismissively as an 'Electronic Book'.
2. **Drill & Practice** - the learner is given the opportunity to practise repetitive previously learned material, eg drug dosage calculations.
3. **Tutorial** - new material is presented in small units with the student being required to input responses to demonstrate understanding or mastery before receiving the next unit. Correct responses are reinforced before moving on while incorrect responses lead to hints or guidance being offered until a correct response is ultimately received.
4. **Simulation** - the computer offers a graphical (eg ECG trace) or textual (eg patient case history) model of a 'real life' situation to which the student responds. In some programs, the effect of the student's decisions are modelled (eg "you have just killed the patient!") and feedback is given as appropriate.
5. **Inquiry & Discovery** - similar to simulation except that the student is required to be more active in eliciting the information felt necessary to complete the task before making

decisions and receiving feedback (eg in a simulated patient assessment the student types "what age is patient?" and ultimately might be informed that age was not relevant in this case).

Two points stand out from this five-category classification. Firstly, there is a gradual shift from the more rigid computer-orientated approaches used in 1. and 2. toward the more individualistic learner-orientated approaches of 4. and 5. Roughly, this trend is a chronological one which follows the development of programming techniques over time. Nevertheless, Jonassen (1985) quotes figures which suggest that Drill & Practice programs account for as much as 70% of existing CAL courseware. The second and more important point about this classification scheme is its fundamental failure as a classification scheme *per se*. This failure, which reduces the usefulness of the scheme to the level of descriptive framework, is due to the fact that it is difficult to identify particular programs which can be exclusively classified by a single category.

To illustrate this difficulty, the examples alluded to in the Simulation category above actually refer to programs described by NBS (1986) and Sweeney et al (1982) respectively. ECG traces and patient case histories may indeed be simulations of 'real life', yet the task facing the student is very different. In the ECG program (NBS 1986) the student is firstly shown traces with the patient conditions they represent and subsequently presented with a trace and asked to 'diagnose' it in a Drill & Practice as well as in a Tutorial fashion - correct responses lead to the next trace, incorrect leads to correction. In the patient case history program (Sweeney et al 1982), on the other hand, a simulation of a patient experiencing a myocardial infarction is offered in that the student must type in actual questions in order to elicit information they will need in order to devise a nursing care plan (eg "Where is your pain ?"). Ultimately the student writes up this care plan on paper. This program may indeed be a Simulation but it is clearly also Inquiry & Discovery which, while not offering any Tutorial facility, seems to be an elaborate form of Page Turner.

The second classificatory scheme used in the literature seems to offer a possible avoidance of the difficulties outlined with respect to the more common five-point framework above. This scheme, developed by the non-nursing author Rushby (1979), seeks to apply a framework of four paradigms - instructional, revelatory, conjectural and emancipatory CAL. To expand, with the help of the writings of MacDonald et al (1975) and Hartley (1981):

Instructional CAL - the principal features here involve the careful selection and sequencing of stored material into small logical steps. The learners' progress through the material is dependent on mastery of previous levels. Feedback with guidance if appropriate is usually available. As such this paradigm can be taken as having evolved from Page Turners to encompass Drill & Practice and Tutorial facilities.

Revelatory CAL – is distinguished by a greater degree of student control of the interaction with the stored base of predefined knowledge. Simulations with Inquiry & Discovery facilities would fall into this category in that it is the student who decides on how to build up knowledge of the model represented within the machine. An additional feature falling into this paradigm (but virtually absent in nursing CAL programs) is the facility whereby the student can alter variables within the knowledge base and subsequently observe the effect of this alteration on the model constructed eg by altering certain vital signs in a 'patient' while witnessing the effect on other vital signs.

Conjectural CAL – more properly than within revelatory CAL, this paradigm encompasses the explicit facility of students altering variables in order to set up and test 'microworlds'. The distinction seems to rely on the degree of control invested in the student.

Emanicipatory CAL - although at first difficult to see how this paradigm constitutes learning, the rationale here is that the 'in-authentic' labour (Kemmis et al, 1977) involved in learning can be reduced by the machine offering word-processing or information retrieval facilities. These facilities, it is argued, will enhance the product of the interaction with the machine while reducing the time spent on activities indirectly related to learning. This type of CAL would tend to be available alongside other paradigms - as an example, a well set-out care plan would be printed out after a student had interacted with a simulated patient.

This classification scheme is evidently more broad than the five-point framework above. The other chief difference is that no reliance is set on either characteristics of the program or on descriptions of the students' task to achieve the classification. As such, the framework could be taken as a functional one which attempts to classify programs on the basis of intended educational purpose - to instruct, to reveal, to permit conjecture or to emancipate non-educational tasks. Nevertheless, there remain problems with adopting Rushby's framework for the purposes of reviewing specific programs since once again the categories cannot be taken as mutually exclusive. Moreover, the categories lack theoretical rationale to account for how it is that the intended educational purposes are achieved. The importance of underlying theoretical constructs will be a theme for much more detailed discussion below.

ICAL Defined and Types of ICAL

Perhaps because of the relative youth of Artificial Intelligence as a discipline, there are fewer terms used to describe learning programs than were identified with respect to traditional CAL.

The common specific term in use is Intelligent Computer Assisted Instruction (ICAI) as used by eminent practitioners such as Clancey (1979), while commentators such as Yazdani (1986) prefer the umbrella term of Intelligent Tutoring System (ITS). At the risk once more of delving into unjustifiable semantics, the observation could be made that 'instruction' and 'tutoring' imply differences in students' pre-existing knowledge of the subject - for this reason the more generic term 'learning' as used by O'Shea & Self (1983) is being used throughout this discussion. If any word deserves emphasis, however, it is 'intelligent' since it denotes the defining concept which differentiates this approach to learning via computer from the traditional approach outlined earlier. An elaboration is required, therefore, of what exactly 'intelligent' is taken to mean.

The underlying aim of Artificial Intelligence (AI) is that computer programs might be constructed to perform activities which could normally only be performed by humans - such as language understanding or problem solving. There is, however, a subtle but important distinction between AI and its partner Cognitive Science which rather depends on how the word 'intelligent' is used. The difference is that Cognitive Science, a psychology-based disipline, is more concerned with constructing computer programs to perform 'human' activities in a manner which emulates humans.

This sometimes elusive distinction has strong implications for ICAL - an AI program such as MYCIN might be successful in its task of diagnosing blood disorders and an argument might therefore be put forward for the student to 'organise your knowledge as this program does' (O'Shea & Self 1983). On the other hand, a Cognitive Science program to perform the same task would be based, to the best of the researcher's knowledge, on ways in which a human expert would actually diagnose blood disorders - asking students to model themselves on this latter approach is very different to asking them to adopt a method of organising their knowledge which happens to be computationally effective. An example of the latter approach could be found in diagnostic programs such as MYCIN (Shortliffe, 1976) which are based on combinations of the probabilities that the observed symptoms in a patient are diagnostic of a particular disease.

This important point will be returned to later, but for the moment it has served to introduce what is possibly the most important component within a ICAL program - the *domain knowledge*. While it is self evident that a traditional CAL program contains knowledge of the subject matter, the difference with respect to ICAL is that the program 'knows' how to utilise that knowledge. It is not just an 'idiot savant'. As will be seen later, this facility lends itself to the teaching of knowledge processing rather than the passing on of more straightforward factual knowledge.

The other two defining features of ICAL programs also involve the concept of 'intelligence'. These features involve firstly the intelligent construction of a *student model*, and, secondly, the intelligent delivery of teaching to the student - a *tutoring module*. Fox (1984) defines a student model as a representation of the student's understanding of the subject expertise. The tutoring module can be more simply taken to be that part of the program which contains the strategies used for teaching. To the extent that all teaching media have some sort of underlying 'theory of instruction', the explicit provision of a student model can be considered more definitive of ICAL. Nevertheless, as will be discussed in later sections, the manner in which teaching is delivered via the tutoring module is often radically different to, for example, traditional CAL.

When discussing ICAL it is easy to forget that the innovation is relatively vestigial and at what Yazdani (1986) terms a 'pre-technology phase' when compared to CAL. Fox (1984), in a review of four of the best known ICAL systems, underlines this immaturity by pointing out that there were at that time no ICAL which had progressed beyond the experimental stage to a point of commercial usefulness. However, although this reminder serves to rather devalue any attempt to construct a typology, it is nevertheless possible to try to fit the programs which do exist into the two classification schemes discussed above. Just as for CAL in the foregoing section, the task of classifying types of ICAL might be made easier by looking for commonalities in different applications.

It was mentioned above that the focus on knowledge processing in ICAL leant the medium to types of programs which seek to educate within this area. Consequently, such ICAL programs as exist are predominantly of the Simulation, and Inquiry & Discovery type. For example, the GUIDON program (Clancey 1979) will simulate a diagnositic encounter with a patient suffering from a blood disorder. Other programs such as SOPHIE (Fox 1984) are similarly of the 'revelatory' problem solving type, in this instance fault-finding in electrical circuitry. Nevertheless, the problems identified aerlier with this functional classication can be seen to continue to apply in that repeated practice, in this case at solving a simulated problem, is definitive of Drill & Practice. It would be useful, therefore, if ICAL could be classified using its own features and approaches rather than having to rely on the rather flawed typologies that have been borrowed from traditional CAL.

Despite the small number of reported projects, Yazdani (1986) has detected what could be the beginnings of an 'intra-ICAL' classification scheme which nevertheless can be extended to include types of traditional CAL. The approach can be taken as following on from the observation made by Mills (1985) that CAL need not only be categorised according to degree of learner control, rather it is possible to see that different implementations reflect different models of learning. This discussion has already bemoaned the lack of a theoretical rationale to the two classification schemes reviewed earlier; it is therefore worthwhile exploring further Yazdani's 'spectrum' of types of CAL program – whether Intelligent or traditional.

Mills (1985) and Yazdani (1986) each begin with the premise that different theories of learning, as implemented on a computer-based teaching medium, should be used for the teaching of different skills. A spectrum is suggested based on the amount of *appropriate structure* which different programs have in their teaching approach. The amount of appropriate structure is in turn dictated by the nature of the skill to be learned. Some skills, for example abstract skills such as analogical problem solving, are possibly better taught using 'discovery' learning principles as implemented in unstructured 'learning environment' programs such as Papert's (1980) LOGO program. At the other end of the spectrum it might even be appropriate to have a highly structured 'behaviourist' type program to teach much more concrete and specific information.

It is possible to find support for this proposal in different ICAL programs – the approach of O'Shea & Self (1983) is one of making more intelligent the tutorial feedback component of a program to teach knowledge which, typically, is the province of Tutorial CAL. In the ACT problem solving program of Anderson (1983) the teaching of this somewhat more abstract skill is accomplished via an emphasis on getting the knowledge domain to be able to solve the task – the student will 'discover' how it is done by being able to extract the principles through observation of the expert (an apprentice – master approach). The usefulness of this 'theoretical spectrum' approach to classification of all types of computer learning programs seems therefore to hold considerable promise. The scheme is, however, still tentative – as witnessed by the absence as yet of any labelling of points along the spectrum. Nevertheless, in the schematic representation of the spectrum offered in Figure 1.1 below, the opposing poles on the continuum have been labelled 'exploratory' and 'didactic'.

Figure 1.1 Theoretical Learning Spectrum (adapted from Yazdani 1986)

Learning environments	Problem solving (eg Anderson)	ITS systems (eg O'Shea & Self)	Tutorial (traditonal)	Drill & Pract. (Skinnerian)

<--->

<- exploratory approach **didactic -> approach**

Conclusions.

The bulk of this section was taken up with typological discussion. The consideration of the various classification schemes offered by the literature of CAL was forced to conclude that no single scheme in use offered a sufficiently adequate framework within which specific CAL programs can be reviewed. The key omission, it is argued, lies in the failure to take into account the theoretical learning base. When ICAL was examined the beginnings of a useful typology were identified which addressed this failing. This 'learning spectrum' approach recalled the earlier discussion of Product versus Process models of education (Sheahan 1986). There is much more to the theory of learning through computer, however, than has been covered in this section. A later section will therefore offer a considerably more detailed exposition of theory.

THEORETICAL CONSTRUCTS UNDERLYING CAL AND ICAL

In order to organise the material to be covered in this section a division will be made between CAL and ICAL. As previously indicated, the absence of examples of ICAL in nursing effectively makes this an additional division - 'nursing' applications versus 'non-nursing' applications.

Theoretical Constructs Underlying CAL in Nursing

The aim of this section is to search for an adequate theoretical model which can then be put to work for the purpose of helping address the overall issue of evaluating the CAL innovation. Various contender theories will be outlined and applied in a general fashion in order to assess their suitability for carrying forward to the concluding section of this chapter - the in depth analysis of selected CAL programs.

The literature on CAL in nursing does not apparently include a general exposition of the theoretical constructs which underlie the various types of CAL programs or paradigms. Nevertheless, there are some studies which discuss specific CAL programs in the light of the preferred learning theories of the authors. For example, Huckaby et al (1979) find support in the findings based on evaluating their CAL program for the learning theories of Ellis (1965) and Gagne (1977). Prior to undertaking a detailed review of some of these papers, it would be useful firstly to outline the major classes of educational theory and, secondly, to adduce which educational theories underlie the various types of CAL program.

Two rather broad 'churches' of educational theory seem most directly pertinent to CAL in nursing - the behavourist models and the cognitive models (including Artificial Intelligence work). In the light of historical antecedents it might seem at first sight that a behaviourist analysis will possess the greater explanatory power of the two. Perhaps cognitive models explain the innovation only partially. Here indeed are statements which require considerable elaboration. Firstly, however, it is necessary to discuss the learning theories which do not easily fit into these two categories but which nevertheless are being deemed here as being of less pertinence to this exposition of CAL learning theories. The two most evident omissions, then, are the Humanistic theories and Social Learning Theory.

Humanistic theories of learning, as developed notably by A H Maslow and C Rogers, at first sight seem eminently worthy of inclusion in an explanation of how it is that a CAL program acts a learning medium. Essentially the humanistic theories can be taken as relating principally to the affective component of the learning event. Much is made, for example in Rogers (1969) ten principles of learning, of the benefits of self-initiated, independent, and meaningful learning experience. Maslow (1971) suggests that self-actualisation in learning will occur when the individual experiences 'fully, vividly, selflessly, with full concentration and with total absorbtion' (p 44). In the 'book' analogy at the opening of this chapter the paraphrased benefits of CAL for students would seem to fit rather well with these ideas given that much is made in the nursing literature both of the individualisation of learning and of the exciting novelty of CAL for students (eg Mirrin 1983).

To the extent that there is no external reward to be had by students who often voluntarily spend time in front of a CAL program, the motivation must be internally generated. Deci (1975) suggests that intrinsic motivation such as this must be based on the human need to be competent and self-determining. However, the need for self-determination places a strict condition on deeming CAL experience to be intrinsically motivating. The experience should facilitate truly autonomous learning in that the 'individualised' programs should be just that. It is insufficient for individualisation to refer to freedom of choice about when and how often the student wishes to run a program.

On closer inspection, therefore, the 'humanistic' aspects of CAL become rather facile since even the novelty value of spending time in front of a computer is perhaps more accurately characterised as an experience of general wariness or caution (Sweeney 1983). In Rogers' own terms, learning experiences should not be threatening to the self. Moreover novelty (or wariness) is temporary - what then remains can by no means be automatically taken as individualised self-directed learning of an accepting and facilitating nature. Instead it could just as well be argued that all too often the mechanistic and poorly-designed programs in more common use are very clearly non-individualised in that all students are treated identically with rarely any concession given to individual learning styles or difficulties. The respected nursing author Grobe (1984) suggests that programs should be written in such a way as to ensure "no-fail" encounters by the student. To an extent, this may point to awareness of student reticence with CAL. It certainly does not point to idealised notions of intrinsically motivating activity.

Furthermore, given that the overwhelming majority of CAL programs are of a Drill & Practice or Tutorial nature, it is necessary to transcend the superficial level of analysis of the experience of CAL for students and look instead to the implicit message conveyed by these types of programs - which Kochar & McLean (1985) neatly summarise as .....

I know what you need

I decide the training aims

I decide the training content
I decide the training sequence
I know how and when to test you
I know when you have learned.

Social Learning Theory (SLT) (Bandura 1977) is the other major omission from further elaboration of the theoretical basis of CAL learning in nursing. CAL program content, most obviously, does not extend to teaching psychomotor skills - a truism worth stressing in the nursing context. However the justification for omitting SLT does not rely on restriction of learning via modelling processes to psychomotor skill learning. Naturally a learner cannot observe a computer expertly giving a bedbath. The restriction, rather, is within the teaching modes offered by state-of-the-art CAL programs in nursing in that only the very palest shadow of a model is ever really available to the student.

To anticipate briefly the section below on teaching strategies, it is exceedingly difficult to identify programs in nursing which include the facility of the program being able to demonstrate and explain (ie model) any but the most trivial of tasks set before the student. For instance, while a drug dosage program might display "No, the answer is 2.5 mgs." (or just as likely display "Try again"), even the more complex simulation programs do not seem to provide much in the way of modelled demonstrations of how an expert nurse, for example, would assess and plan the care of a patient presented in case history form. As will be discussed more fully below, modelling of skills can be regarded as a strength rather than an omission in Intelligent Computer Assisted Learning (ICAL). If and when, therefore, there are nursing applications of ICAL programs then Social Learning Theory will become more relevant.

Behaviourist models

The discussion can now return to an elaboration of the two principal classes of learning theory which here are taken to be relevant. The first of these, the behavourist model, must be taken as potentially the major theoretical model underlying CAL in nursing. This statement is justified by the fact that the majority of CAL programs currently available - Drill & Practice and Tutorial - seem at first sight to fit most economically with a behaviourist model without much necessity for recourse to 'higher' cognitive models of learning. Without any further prejudging of the adequacy of the model, however, a behavioural analysis must now be offered of the mode of operation of these programs in the light of their Programmed Learning antecedents.

Drill & Practice programs are fundamentally linear. The material is presented to the student in a step-wise sequence toward attainment of the desired behaviour, eg being able to name all the parts of the heart. After a frame of material has been presented, the student is required to respond, eg by typing in "left ventricle", and is immediately informed whether she is right or wrong. If the response is correct the program moves to the next frame (in a sequence predetermined by the author) while if the response is wrong then usually little is made of it except perhaps to offer another chance. Wrong answers may even be ignored if the program has been prepared along theoretically 'pure' Programmed Instruction lines, although in nursing CAL it is perhaps only the 'knowledge test' programs which follow these lines.

Using a behavioural model, such programs are almost wholly explained by operant conditioning in that correct responses are reinforced and incorrect ones largely ignored or reinforced in successive approximations toward desired response. It is therefore important to sequence material in order to maximise the chances of obtaining correct responses from the learner. The pedagogy of Programmed Instruction, as advocated principally by Skinner (1968) but now considerably devalued by educational technologists (eg O'Shea & Self 1983), would nevertheless look askance on the development of feedback delivery after incorrect learner responses. It is this 'broadened feedback' aspect of nursing Tutorial programs which requires the discussion to turn to the cognitive learning models in order that a more complete exposition of the theoretical basis of CAL can be constructed - although there are many levels of feedback. As will be argued below, it is by no means clear that even the broadened version of feedback typically offered by Tutorial programs is what the cognitive theorists have in mind.

Perhaps because of the fall in popularity of 'connectionist' theories of learning, programs which are fundamentally Drill & Practice have been 'upgraded' to Tutorial status through the provision of broadened feedback. Thus programs which owe their theoretical allegiance to behaviourism become underpinned by more recent and more fashionable models of learning. The literature provides evidence of this. Bratt & Vockell (1986), for example, describe their suite of Drill & Practice / Tutorial programs (reviewed in detail below) as fitting with the 'mastery' learning paradigm of Gagne (1977). The rationale being that there is a steady progression through successive levels of difficulty which is accomplished principally through sequenced material with feedback on performance.

In passing it is worth mentioning that Gagne is a cognitive theorist who considers stimulus-response learning to be of a very basic type; however, the main point is that the feedback in these programs is a highly degraded version of the feedback held to be important by Gagne and also by Ausubel et al (1978). In Tutorial programs a guidance comment is at best a pre-programed hint. 'Broadened feedback' is therefore far from individually-tailored feedback designed to meet specific student difficulties. This feedback issue will be more fully discussed below, but for the present it must be concluded that a behaviourist analysis of Drill & Practice or Tutorial programs in nursing would be that the student learns either by being told the correct answer or by strengthening the S-R connections between required answer and desired answer.

Within the behavioural framework, therefore, Tutorial programs are contended here to be different yet similar to Drill & Practice programs. Similar in the respect that material is carefully sequenced and predetermined by the author; similar also in that knowledge is acquired (rather than gained through experience) by reinforcement of correct connections. Where Tutorial programs differ, however, is with respect to their acceptance of the necessity to feedback at some level when responses are incorrect. A student may, for example in the program described by Richards et al (1986), be presented with five possible actions she might take when a patient develops pyrexia. The task is to choose the correct actions in correct priority - if an incorrect action is choosen then an explanation is displayed of why this was wrong. The value of this type of feedback, sometimes delayed, has been shown by Tait et al (1973) to transcend the straightforward reinforcement principle, but, as argued above, even this type of feedback is not wholly individual.

So far the discussion has focused on the more common types of nursing CAL program since these are the ones apparently most suited to behaviourist explanation. Left out of the account so far have been the Simulation, and Inquiry & Discovery type of programs which are very difficult to fit to a 'connectionist' model - although as previously stated these programs are rarely devoid of Drill & Practice or Tutorial elements. It is toward an elaboration of the theoretical models which might account for these programs (and also 'broadened feedback') that the discussion can now turn. The interim conclusion on the adequacy of a behavioural analysis of nursing CAL is that the model has rather low explanatory power and that, furthermore, nursing CAL programs are shown in a poor light by such an analysis.

Cognitive models

It is worth repeating that theoretical work within the literature on CAL in nursing seems virtually absent. If this is true for the behaviourist theories, however, it is not so strongly the case with respect to the cognitive theories of learning. A few studies mention that their programs fit in with the work of cognitive theories such as those of Gagne (1977), cited by Bratt & Vockell (1986) and Huckaby et al (1979). In view of the apparent absence of theoretical work, then, it is necessary to once again adduce which cognitive theories can be taken as apposite.

The first task is to identify which learning theorists of the cognitive tradition are those whose insights might be utilised with respect to understanding the theoretical basis of CAL in nursing. Subsequently the task becomes one of searching for a common 'cognitive' framework within which specific CAL programs can be analysed. Most obviously the work of Bruner (1960), Ausubel (1978) and Gagne (1977) merits attention. In addition, the broader field of information processing theory (eg Lindsay & Norman 1977) deserves also to be looked at. Finally the discussion will turn to the field of Artificial Intelligence learning models in order to construct an account of nursing CAL programs which fall into the 'problem solving' category.

Bruner, firstly, has made a notable contribution to the psychology of learning. Interested principally in the mental representation and acquisition of conceptual categories, Bruner's model can be characterised as one where a hierarchial ordering of knowledge categories exists. Conceptual categories are acquired and represented through coding systems being applied to the attributes which relate to the events or objects being learned. The learning strategy advocated is one of 'discovery' where a situation is set up for the learner to explore, discover and master inductively the principles underlying a specific instance. Bruner's model can be taken as similar to other cognitive models in so far as the essential idea is one of learning being the incorporation of new information to existing cognitive structures.

Although associated with recent cognitive psychology, the incorporation of new information to existing cognitive structures is a model which can be traced back to Piaget (1970) - for Piaget 'assimilation' and 'equilibration' are terms denoting the processes of integration of information into structures and modification of structures to compensate for the disturbances caused by the new information. More recent support for the same essential idea of incorporation can be found in the writings of Artificial Intelligence workers such as Schank (1982) who has developed a theory of dynamic memory which stresses the importance of a context existing into which the learner can place new information in order to make sense of it.

Moving more toward general information processing theory, Gagne can be taken not so much as having contributed a theory of learning but rather as having offered influential thoughts relating to the factors which influence instruction. Learning, classified by Gagne into five 'capabilities', is said to be facilitated when subordinate levels of the various hierarchies of conditions are satisfied. These levels are often akin to the components of the information processing system (eg sensory register, working memory, long-term memory) which have been thoroughly established by cognitive psychologists such as Lindsay & Norman (1977). Searching for a model which offers a more direct application of the principles of information processing theory to learning brings to the attention the work of Norman & Rumelhart (1975). Their model, which possesses the added attraction of simplicity, categorises several aspects of learning into three modes:

1. Accretion - the addition of new knowledge into existing schemas. The framework exists, but new data are entered. This is the most common type of learning.
2. Structuring - the formation of new conceptual structures. The existing schemas will no longer suffice; new schemas must be formed.
3. Tuning - the fine adjustment of knowledge to a task. This occurs when the appropriate schemas exist complete with necessary knowledge, but they are inefficient for the purpose due to being too general or not yet matched to a given task. Tuning changes mere knowledge of a topic into expert performance.

Despite the potential usefulness of Norman & Rumelhart's model in relation to an analysis of nursing CAL, there is however one other cognitive psychologist - Ausubel - who seems to offer a cognitive learning model which is both more elaborated and perhaps more pertinent to this discussion. However before looking at the work of Ausubel in more detail it is worthwhile stressing through one last paragraph the commonalities which exist in all of the foregoing cognitive learning theories.

It can be taken, then, that the foregoing cognitive models can be distinguished more by degree of emphasis than by fundamental difference. Central to each position, for example, is the stress laid on the integration of new information into pre-existing cognitive structures. Each model, moreover, is hierarchial in that different learning tasks demand increasingly complex mental representations - variously known as frameworks, schemas or conceptual categories. What follows from each hierarchial model is that learning should be sequenced by promoting mastery of logically-structured dependent levels from vocabulary memorisation via concept formation toward problem solving. Similarly, the principles of capacity limitations and cognitive economy can be seen to underlie each learning model in that material to be learned should be presented in sufficiently small units or 'chunks'. For at least the purposes of this discussion, therefore, it can taken that Ausubel's learning model will enjoy the support of several other cognitive models. For more coverage on how each cognitive theory might apply to nursing education (see Quinn 1980).

Ausubel and Assimilation Theory The central idea in Ausubel's (1978) cognitive model of learning is Assimilation Theory. The essential idea is one of 'meaningful' learning occurring when new information interacts with the learner's existing cognitive structures. What the learner already knows is held to be so important that Ausubel goes so far as to exclude 'connectionist' or rote learning as a relevant (ie meaningful) mode of learning. To paraphrase Ausubel, ascertain what the learner already knows and teach him accordingly. Meaningful learning will take place when three requirements are satisfied:

1. The learner's 'set' - a meaningful (rather than a rote) learning set must be adopted by the learner. More simply, the learner must be ready and willing to relate new ideas to what is already known.
2. The learning task - it should be logical in that it can be related to the learner's existing cognitive structures in a sensible way, ie the new material should be *potentially meaningful*. Structures are taken as hierarchically organised sets of concepts and ideas relating to a given topic.
3. The learner's existing cognitive structures - it should contain relevant ideas with which the new information can interact. It is important to note that a cognitive structure is actually comprised of *content* and an *organisation* of content.

Clearly there is already much here which is of relevance to nursing CAL. On the positive side, the common practice of scheduling CAL sessions after more traditional teaching fits well with these requirements. Similarly, the logically sequenced arrangements of material in programs, while originally a behaviorist idea, seem well suited to Assimilation Theory's principle of subordinate subsumption. Thus the mental representation of this knowledge which a nurse might hold would have HEART at its topmost level with PUMPING VENTRICLES at the next level - a properly sequenced CAL program would establish that the nurse had mastered this much of the concept and only then display material such as BUNDLE OF HIS.

The most obvious difference thusfar between cognitive and behaviourist theory is the emphasis on the 'mental set' of the learner. Nursing CAL, as previously discussed, is often taken as motivating for learners and as such might be taken as satisfying Ausubel's first requirement. However the support must only be partial since it is by no means always the case that programs deliberately set out to foster a 'meaningful' mental set in the learner by, for example, offering parallels between the new material and concepts which have been previously grasped.

Thusfar nursing CAL programs seem broadly to fit in rather well with Assimilation Theory. However a closer inspection of what meaningful learning is not potentially reveals a problem with, more especially, the types of program with least learner control - Drill & Practice in particular. Ausubel conceptualises meaningful learning as lying on a dimension with rote learning at the other pole. Learning by rote is taken as being both costly in effort and inefficient in terms of retention since word-for-word learning fundamentally opposes the principle of incorporation into existing cognitive structures. A CAL program which required the learner to repeatedly apply formulae in order to work out drug dosages or unit conversions would be an example of rote learning. On closer inspection, however, a feature of the fairly common Drill & Practice programs in nursing can be found which saves them from automatically being dubbed as rote – the provision of *organisers*.

Organisers (or advance organisers) are considered by Ausubel to be the introductory or after-the-fact contexts which are provided for the student in order to facilitate assimilation of new material. Visual illustrations or verbal metaphorical relationships are common organisers. Straightaway it is seen that even the most basic type of Drill & Practice program can become 'dignified' by this insight. Consider an anatomy program which sets out to teach names of parts of a system – the organiser provided would be the diagram gradually being augmented on the screen as the learner correctly names a part. And yet there is an important sense in which this type of program fails to merit the title of a meaningful learning medium. What is missing from the interaction with the learner is any check on whether real understanding is being gained by the learner. Mills (1985) offers as an example of this failing the program which asks the learner to 'fill in the blank' with the name of concept given earlier. In nursing CAL this approach (no more than a memory association test), seems to be fairly common, eg the programs reported by Bratt & Vockell (1986) often feature an 'identify the correct spelling' mode.

Nevertheless, while evidence is generally being accumulated to support the effectiveness of organisers as teaching strategies, there are some important qualifications. For example, studies by Mayer (1978) and Lesh (1976) show that organisers may be especially important for the learning of poorly organised material. However, in Drill & Practice programs the proper sequencing of material has never been a weakness. The paradox is that it is the type of CAL program which superficially seems to follow a cognitive model of learning – Simulations, and Inquiry & Discovery – which are perhaps most at fault when it comes to facilitating meaningful learning. The crucial issue turns out once again to be that of feedback. Hence in programs where the student task is one patient assessment there is often little or no attempt to explicitly relate material to previous learning or experience.

To illustrate the point, in the program of Sweeney et al (1982) the task is to assess a patient suffering cardiac distress. The program, when closely inspected, actually does almost nothing in the way of teaching in that it operates by displaying on the screen no more than the verbal responses of the patient to typed questions such as "Where is your pain ?". This program will be considered in more detail later, but for the moment the discussion must return to Assimilation Theory in order to search for an account of the 'discovery' principle which apparently underlies these 'higher' types of program.

Learning through 'discovery' or through 'reception' is a second unrelated dimension in Ausubel's model. Unlike Bruner, Ausubel can see no benefit in the learner independantly discovering underlying principles on her own. Rather the student should 'receive' the entire content of material in a meaningfully organised, final form. In contrast, exponents of the discovery approach emphasise that learning through discovery is more motivating, less authoritarian, and more suited to the development of critical thinking skills. For nursing, at any rate, it is wrong to dwell on the dichotomy of these two positions. Nursing knowledge is of a reality-based and problem-solving nature, and, since even Ausubel concedes that discovery learning is apt in these circumstances, it is of more benefit to this analysis to accept that much nursing education (including CAL) will be of a discovery nature. What is required, rather, is a cognitive learning model more directly focused on the problem solving process which can then be used to analyse specific Simulation, and Inquiry & Discovery programs. To find such a model, the discussion turns finally to the field of Artificial Intelligence - specifically the ACT* (ACT star) model of Anderson (1983).

The value of the computer as a tool to teach, improve and practise problem solving skills is a popular theme in the nursing CAL literature, eg Mirin (1983), Hoy, R. (1983). The types of program which are said to fall into the problem solving domain are those classified within Rushby's (1979) scheme as "revelatory". Koch & Rankin (1987) elucidate the central idea as being one where the computer acts as mediator between a hidden model of a real life nursing situation and the student, whose task is one of fact collection in order to define and analyse the problem prior to making a decision and observing the effect of that decision. Norman (1983) is not alone in drawing the clear parallel between Revelatory programs and the nursing process *in vivo*. Simulations, and Inquiry & Discovery programs can therefore be seen as offering a safe and cheap means of orientating novice nurses to complex reality (although the reality might be somewhat degraded - D. Hoy 1985 comments that it feels like nursing a patient at the end of a telephone). Nevertheless, given the hype - a word which may or may not be derived from 'hyperbole' - it is all the more extraordinary that a theoretical analysis of the learning processes involved cannot be located in the literature.

It is first of all necessary to determine the pitch of revelatory nursing CAL programs in terms of the level of intellectual skill aimed at in educational terms. Gagne (1977) postulates a hierarchy of learning processes which can be recruited to aid this task. From simple to most complex, the hierarchy is one of stimulus-response learning, chaining (memory association), discrimination, concept formation, rule learning, and problem solving. It cannot, however, be automatically assumed that it is only the topmost skill of problem solving which is of concern to this part of the discussion. As will become apparent, a close inspection of revelatory nursing CAL programs acts to considerably dampen the claims made within the literature. Starting, therefore, at the level of what Gagne refers to as "intellectual skills" ie discrimination learning, the cognitive processes involved in each skill will be outlined prior to analysing what skills, if any, are indeed fostered by revelatory nursing CAL programs.

Discrimination learning basically involves learning the difference between stimuli. In terms of an attribute and its values such as a patient's face and its colour, a nurse would learn to distinguish between the values 'flushed' and 'pale'. While discrimination processes are of interest, nevertheless for the purposes at hand the role discrimination plays with respect to the next category of intellectual skill is of greater importance since *concept formation* seems more directly applicable to revelatory CAL. When a nurse acquires a concept such as 'fever' it is generally considered by cognitive psychologists that the mental representation of this concept will be in the form of a hierarchical list of the features which define a prototypical instance of 'fever' (Mervis & Rosch 1981). However the point here is that it is how a concept is used (*rule learning*) that is the crucial component in problem solving. In addition, a clear distinction must be made in terms of the purpose to which a concept will be put with respect to problem solving.

Before going on to explain why these points are central to this part of the discussion it might be helpful to elaborate on the 'customised' terms which are part of the currency in use within the cognitive psychology of problem solving. The distinction is made between 'declarative' and 'procedural' knowledge. Declarative knowledge is taken to refer to the sort of 'received wisdom' which people hold. For instance a nurse might read or be told how to set up a trolley for a dressing. There is a considerable difference between this and the knowledge gained from actually carrying out this job in real life – procedural knowledge. Using declarative knowledge to carry out a task is often laboriously, slow, and liable to the forgetting of component parts. In skilled and experience-gained procedural knowledge the components become better integrated with the result that the task quickens and is done more efficiently while paradoxically using less mental effort. In other words the whole pattern is responded to, the concept now includes actions along with verbal knowledge. These *condition / action rules* (Newell & Simon 1972) are known as productions which take the form of IF ..*condition*.. THEN .. *action*.

To illustrate these points, consider the concept 'fever' as mentally represented in a qualified, working nurse. Most probably, the concept will contain the defining features in terms of attributes + values which over the years have come to allow the nurse to most economically recognise the signs of fever in a patient, eg 'face flushed and movements restless'. However the purpose of this working nurse with respect to fever is to act on the recognition of fever signs – therefore there will be actions inseparably linked to the signs in the form of ... 'IF the face is flushed and movements restless AND pyrexia is confirmed THEN inform doctor, institute cooling, start chart'.

Proceduralisation of declarative knowledge, to use Anderson's terms, can be further illustrated with respect to an unqualified nurse in that the concept represented might comprise .... 'IF the patient is restless THEN report to Sister'. Finally, the purpose of the learner nurse might lead to procedural knowledge of a very different kind since she might be more concerned with being successfully examined on her declaritive knowledge ... 'IF the examination question is about fever THEN write about thermoregulation'. It is to promote the qualified nurse's procedural knowledge (and to avoid the 'exam' version) that clinical practice and simulated clinical practice are considered to be important consequent to declarative teaching.

The learning theory of Anderson (1983) is at its most fundamental a model of how declarative knowledge becomes gradually more proceduralised until the final stage of more expert performance where the productions become *tuned*. Tuning is achieved through experience-based knowledge of the essential conditions necessary to invoke a particular production. In addition, several production rules can be *composed* more economically into a single rule, eg in the fever example above the separate rules for informing the doctor, alerting junior staff, and writing up the kardex may become composed into a single 'reporting' production comprising all of these actions. Two more processes are important. Firstly, for every time a production rule is used then it becomes *strengthened*, and, secondly, when the actions in several productions are identical then the the rules can become *generalised* in that all the conditions leading to the same action go together to form a new, more economical production. For example, several situations in ward nursing lead to 'reporting'.

Given that strong support for ACT* theory can be found in the success learning of the various computer models which Anderson has implemented, there are several useful implications for revelatory nursing CAL programs which arise from this analysis of problem solving learning. Most importantly, practice in tasks will be beneficial. While this practice should be guided through feedback, there should also be provision of worked-out examples from which the learner can extract general principles. The relationship between conditions and actions should be made explicit in the illustrative examples, as should the relationship between several conditions and the same action. It is insufficient to lay emphasis on conditions (eg collecting data) rather the action links should be stressed. However, the relative importance of individual conditions should be made clear in order that certain rules are strengthened. How, then, do the problem solving nursing CAL programs compare to this ideal?

It first becomes apparent that the major class of problem solving teaching programs seems to meet some but fail on other criteria in the above prescription for a useful tool for teaching the skills involved. While these 'branching' programs, as exemplified by the PLATO approach, lay emphasis on the declarative knowledge necessary to prescribe actions, there is nevertheless no provision of worked examples from which the learner might be expected to adduce general principles. Similarly, the underlying principle of these programs is to place the student in the position of discovering for herself how to determine the correct answer. There is no facility for the 'teacher' to provide a model of a learner's difficulties, just as there is no facility whereby the student can 'look over the shoulder' of an expert modelling the task. The more fundamental problem with these programs is that there is no 'reality practice' where declarative knowledge can be proceduralised and tuned. These points, as well as a fuller analysis of branching programs, will be developed later.

The other, smaller, class of revelatory program seems initially to fit more closely with the principles of cognitive science. These programs, where the learner interacts with a simulated patient in case history form, are notable for two features. Firstly, there is often a facility (eg in the program of Sweeney et al 1982) whereby productions can be tuned. For example, if the learner asks a non-specific question such as "Tell me all about your pain" then a non-committal and unhelpful reply is received. Secondly, this type of Simulation sometimes offers more tailored feedback in the form of a comparison of the learner's care plan priorities with that of how an 'expert' would prioritise care. Nevertheless, there are significant areas of weakness in these programs. Principally these weaknesses again involve the poverty of the student and teaching models contained within the program - although on the credit side there is clearly a commitment to seeing 'learning to solve problems' as an intellectual process which requires refinement. The overall thesis of this discussion is that 'we can do better' - after first offering a conclusion on the theoretical basis of CAL in nursing the discussion can return to how this might be achieved through Intelligent CAL.

Theoretical Constructs underlying ICAL

This section of the discussion will be considerably shorter than the corresponding section which analysed the theoretical basis of CAL in nursing. There are two simple reasons for this. Firstly, there is only one reported implementation of ICAL in nursing (other than the present project) which can be examined. Secondly, the foregoing exposition of cognitive models of learning - especially the latter ones - has served to introduce all of the theoretical principles necessary for a full understanding of ICAL. All that remains is to elaborate on cognitive models such as Anderson's ACT* and on conceptual hierarchies which have been implemented within teaching programs.

O'Shea & Self (1983) point out that a tutorial comprises three interacting components: the subject, the student, and the teacher. This deceptively simple observation can be seen as almost encapsulating this entire discussion in that the review above found fault with one or more of these aspects of computor tutorials for any given nursing CAL program. Fortuitously, these same three components can be utilised to construct a theoretical analysis of ICAL. The analysis will not be completed at this point since a later section will review learning and teaching models embodied in CAL and ICAL, nevertheless an analysis of theory base is a good pump primer.

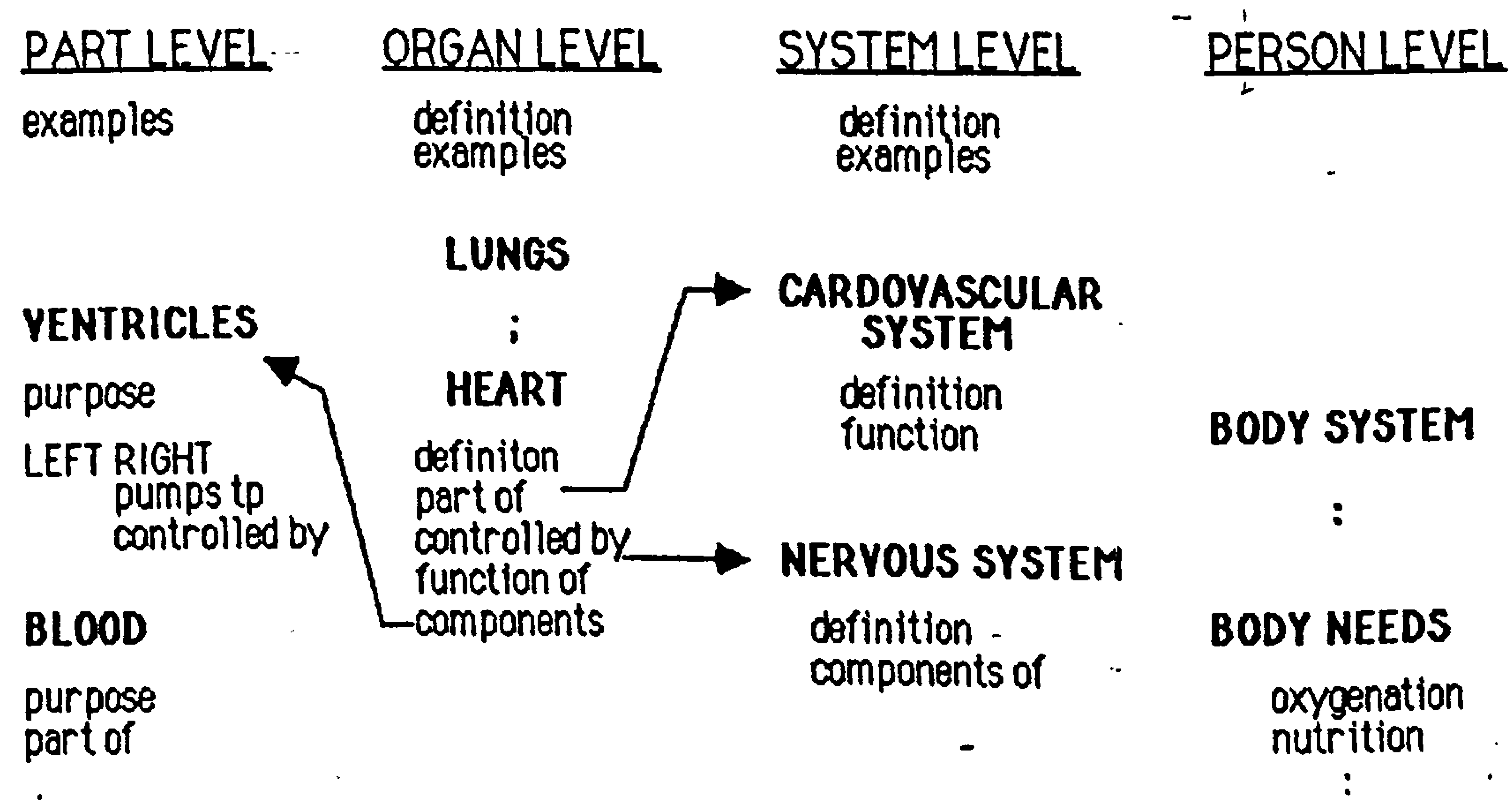
Knowledge representation in ICAL

In the foregoing discussion, criticisms were made both of the representation of knowledge within CAL programs and of the depth of knowledge itself. The factor cited as largely accounting for this criticism is, paradoxically, the aspect of nursing CAL which is currently receiving most praise in the literature - the use of 'authoring' tools. An authoring tool can be thought of as a template of a CAL program into which the non-computer-literate teacher can insert instruction.

However, while Cousins (1986) extols the virtues of authoring tools (eg MICROTEXT), Norman (1983) gives a hint of the drawback by warning that most authoring languages "lock" the user into the tutorial teaching strategy. There is the rub, since this means that the program must inevitably take the shape of an inflexible frame-by-frame sequence. There is no way the program can respond to an 'extra' or unanticipated inquiry. The alternative, admittedly more expensive and specialised method, is to represent the knowledge in a fashion which will allow the learner to explore the material in a more truly individual manner. Two of these alternative, ICAL-style representations - semantic networks and rule-based formalisms - can now be outlined in order to illustrate the point.

Semantic networks, firstly, can be thought of as hierarchical data structures of facts, concepts and procedures with appropriate inter-relationships. Originally described by Collins & Quillian (1969), networks are hierarchical representations of information based on the principle of categories (or nodes) going from broad and general toward more specific and subordinate. The 'semantic' part is denoted to convey that the 'meaning' of a node is contained in its relationships with other nodes. Hence in the example below of a very sketchy anatomy network (Figure 1.2), 'heart' is stored under 'ORGAN LEVEL'. It is part of 'cardiovascular system', has 'ventricles' as components and is controled by 'nervous system'. These relationships (indicated by an arrow) denote both functional and category membership meanings. Relationships with other anatomical systems and parts are denoted by nodes which are one or more steps removed in the hierarchy.

Figure 1.2 Outline of anatomy network (after Carbonell, 1970)



Networks within a ICAL program provide some interesting solutions to some of the difficulties discussed earlier with Tutorial 'branching' systems. The sole reported ICAL system in nursing – COMMES – employs a semantic network representation of its stored knowledge. Using Figure 1.2 as an illustration, it can be seen how such a program would operate when interacting with a student. Although based on SCHOLAR (Carbonnell 1970), retrieval of information can be seen to be a fairly straightforward search along the node links for appropriate answers. So if a student types "What system is the heart in ?" then the program simply enough finds the answer at the other end of the link which goes from 'heart' to the SYSTEM LEVEL. In this fashion the COMMES system responds to queries from learners – Evans (1983) gives an example where a student types that she wants to know about "patient care and emphysema in the context of rehabilitation". The system responds by searching its knowledge base for links between the key words 'emphysema' and 'rehabilitation' before displaying on the screen a summary of central literature references which incorporate these 'nodes'.

This last example offers a clue to real usefulness of this 'richer' form of representation - such a trivial question would probably not be anticipated by someone using an authoring language where every single exactly-worded question must pre-written into the program along with the desired feedback. Even if the teacher did anticipate this question, typically the capacity limitations within microcomputer-based authoring tools will mean that the teacher must decide upon the most important questions likely to be asked. Moreover, a program based on a semantic network can cope with quite general questions in a similar 'inference' fashion by heeding the 'distance' between nodes. So if the learner types "Tell me more about the circulatory system" then the information contained in relevant nodes will be displayed starting with nearest nodes first. It is these 'human-like' teacher qualities which Carbonell considered to be the strengths of the SCHOLAR system - certainly these features seem to fit the prescription offered earlier for a theoretically sound teaching system, eg incorporation of material into existing frameworks through the use of organisers.

One of the chief benefits of semantic networks, then, is the flexibility of responding to learner enquiries. Another advantage, as Rumlehart and Norman (1985) point out, is that it matches many of our intuitions for the representation of a large domain of our knowledge. There are, however, problems with the semantic network type of representation. These problems are summed up by O'Shea & Self (1983) in terms of the difficulties of networks to cope with the teaching of processes. In terms of the discussion above, networks are effective teaching media for declarative knowledge but become less useful with respect to procedural knowledge. As such, they constitute a serious ICAL rival to Drill & Practice and Tutorial programs but for the programs more directly concerned with procedural knowledge - Simulations, and Inquiry & Discovery types - the discussion can now turn to the more suitable rule-based knowledge representation system.

The rule-based system is the second method of representing subject matter which should be considered following from the discussion of cognitive theories of learning. As outlined above, the strength of this approach was in the attention paid to 'proceduralising' knowledge in that the *implications for actions* of the declarative knowledge is made available for the learner.

The most notable implementation of a rule-based approach to teaching is the GUIDON system under development by Clancey (1979) at Stanford University. This program, itself containing some 200 'tutorial' rules, is designed to work alongside the MYCIN program for diagnosing meningitis and diseases of the blood – the learning principle being that students can observe the process of an 'expert' diagnosing and prescribing treatment while also being able to interrupt and ask for explanation of steps which are hard to understand. When GUIDON is used, it is the student who attempts a diagnostic task in case history form (which MYCIN has already 'solved'). In this case, as in Anderson's ACT*, the teaching comprises feedback given to the student which is based on a comparison of the student's performance against that of the 'ideal' solution arrived at by the computer.

The discussion will return to the methods by which these systems model both the learner and the teacher. For the purposes of elaborating the theoretical basis of learning with respect to how the knowledge is represented, however, it should be recalled that the underlying principle of representing knowledge in 'chunks' of information called procedural rules is that this is how human experts represent knowledge. Moreover, the manner in which the rules are processed by the program is held to be analogous to how humans reason and make inferences. These claims, which to an extent underpin the very legitimacy of these programs as tutors, are high claims indeed. In terms of learning theory, this type of ICAL sets itself on a far higher level than traditional CAL. It might therefore be useful if a worked example were provided of how procedural knowledge rules would actually be processed by a program.

EXAMPLE of 'propogating an inference' in ICAL – a system is being asked to decide whether a patient (Mr. Smith) should be 'turned' 2-hourly or 3-hourly. The specific rules used in this example are ....

RULE A	IF THEN	movement -is- reluctant encouragement -is- necessary
RULE B	IF THEN	skintype -is- papery pressure sorerisk -is- high
RULE C	IF THEN	pressure sorerisk -is- high position change -is- frequent
RULE D	IF AND THEN	encouragement -is- necessary position change -is- frequent turning -is- 2 hourly
RULE E	IF THEN	skin -is- shiny and thin skintype -is- papery

When asked the question ....."Is turning 2 hourly (for Mr. Smith)?" ... the system would procede in the following manner - which the reader might more easily understand if a pencil is used to chart the inference procedure as the system consults its knowledge base of 5 rules.

STEP 1 : system finds a rule where the conclusion 'turning -is- 2 hourly' is present. RULE D is found.

STEP 2 : system tries to find out if the two causes in Rule D are present, ie. true. If they present then the system can say 'yes' to the question asked of it. The system takes the first cause ('encouragement -is- necessary') and sees if it is true by searching for another rule which has 'encouragement -is- necessary' as its conclusion. RULE A is found.

STEP 3 : system 'reasons' that it could conclude 'encouragement -is- necessary' if the patient's 'movement -is- reluctant', so a question is asked on the screen ..... "is (Mr. Smith's) movement reluctant ?" ..... to which the user (who knows Mr. Smith) types "YES". Rule A and first cause of Rule D are therefore taken as true.

STEP 4 : system drops to second cause of Rule D ('position change -is- frequent') and again tries to see if this cause is true by searching for a rule which has this attribute + value as its conclusion. RULE C is found. Now Rule C would be true if 'pressure sorerisk -is- high' were true, but at this point the system does not ask the user because a rule is found (RULE B) which would allow the system to 'infer' that sorerisk is high if only it knew that the patient's skintype was papery.

STEP 5 : system finds a rule where 'skintype -is- papery' is the conclusion. RULE E is found so the system asks the user via the screen ... "is (Mr. Smith's) skin shiny and thin ?" ... to which the user might answer "YES". System therefore concludes that pressure sore risk is high and in turn can conclude that Rule C is true.

STEP 6 : system returns to Rule D (at last!) and sees that both causes have been satisfied either by getting answers from the user or by inference from stored knowledge. The system will therefore display on the screen ... "Yes, turning is 2 hourly".

This example will give an idea of how subject matter is processed by the program. To the extent that both the knowledge implemented and the processing strategy used can be taken as psychologically valid (in lieu of much evidence), the learning approach is one of 'looking over the system's shoulder'. There are, however, additional learning facilities typically available in expert system-based tutors - feedback, explanation, and modelling of student-generated hypotheses. To illustrate, consider the following imaginary discourse between learner and ICAL tutor which follows on from the above example ....

The user - perhaps a learner nurse caring for Mr. Smith - might wonder why the system (or the Ward Sister?) came to this conclusion. If she types in WHY?, meaning 'why is turning 2 hourly?', she will receive the answer ... "since position change has to be frequent because pressure sorerisk is high". If she again types WHY? (why is sore risk high?) then the next rule is displayed for her inspection .... "sorerisk is high because skintype is papery". And so on.

It is of course possible to augment this feedback with further descriptive (interactive video?) or biological information made available simply by typing MORE. An additional function is important - when responding to a question like "is skintype shiny and thin ?" the user can type WHAT IS? (and receive further explanation). This type of extra knowledge might well be represented in the form of semantic network. The final facility which could be available relates to the learner receiving feedback on hypotheses that she generates. In the example she might type WHAT IF YES meaning 'what would happen if I typed YES to your question?'. In this latter case the system will 'forward-track' and display the goal or hypothesis that it is trying to work toward.

As the discussion turns further to knowledge representation so it strays into the domain of the next section - models of learning within the program and models of teaching of the systems. After offering a brief conclusion, therefore, that section can therefore begin.

Conclusions

Various theoretical models of learning have been analysed with respect to their ability to account for the learning medium of Computer Assisted Learning in nursing. It was found that no single theory could be applied successfully to all the various types of CAL. Furthermore, the case was made for only a limited explanatory power of specific theories with respect to specific types of CAL. Of the two major contender theories of learning - the behaviourist and the cognitive models - it was found that a 'general' version of the cognitive approaches could be taken as being able to offer the more adequate explanation. Nevertheless, even the more explicitly 'cognitive' types of CAL were seen as having fallen short of a full implementation of the theories to which they owe their allegiance. The tone of the discussion may thusfar be adjudged to be critical in a blanket fashion. This is a 'partial' misrepresentation - an innovation must expect (and even demand) criticism if the commitment to development and improvement is being taken seriously.

The usefulness of a cognitive psychology approach was supported by the brief look at the theoretical basis to ICAL. Hence the semantic network fitted well with a Tutorial type system. A Simulation type of program, which will set out to model human *processing* of knowledge, is the end-goal of the present project. Possibly the ACT* manner in which inferences are processed by using production rules will be the appropriate form of representation of knowledge within that system. One crucial point remains, however. The ICAL systems which claim expertise normally found in human cognitive domian such as diagnosis have not been constructed from first principles. In other words, *there has been no rigourous study of human expert cognition prior to the system construction* Claims regarding 'emulation' of human cognition must necessarily be resisted and replaced by sober statements such as 'operates in the same domain as'.

MODELS OF LEARNING AND TEACHING EMPLOYED BY CAL AND ICAL

The concept of 'student model', as introduced earlier, was defined as the representation of the student's understanding of the subject expertise which is built up by the program. O'Shea & Self (1983) put it more simply as any information which a teaching program has which is specific to the student being taught, and which will help the program to decide on appropriate teaching actions. Glaser (1976) puts it even more simply by coining the phrase 'adaptive education'. It is appropriate, therefore, to undertake a review of the manner in which CAL and ICAL have *interfaced* a model of a user (learner) with a specific teaching model. Most work in this area has focused on the 'learner' part of this equation, consequently the review will offer more detailed treatment of student rather than teaching models.

In order to set the scene and prime the discussion, it would perhaps be worthwhile to look at the particularily apposite research carried out by Pask and Scott (1972) on the matching of learning 'styles' to teaching 'strategies'. The initial distinction made was between 'serialist' learners, who learn by making simple links between items of knowledge, and 'holist' learners, who learn by forming knowledge into a complex whole. The prediction arising from the theory was that each type of learner would have a preferred learning strategy. If serialists and holists were 'mismatched' into teaching styles unsuited to their preference then learning would be less effective.

This hypothesis was tested by the authors assigning students, predetermined as serialists or holists, into each of two teaching groups. The first group was serialist in that information strictly relevant to the the topic was presented in a strictly orderly fashion. The holist teaching group, on the other hand, was based on a strategy which emphasised the 'overall' concepts of the topic. The results showed that learning achieved by the various students depended on the group to which they were assigned. Not only, for example, did serialists learn best when in a serialist group, but also it was shown that an inappropriate teaching strategy (eg holist in a serialist group) resulted in students performing at a level about 50% below that of students in appropriate groups.

The usefulness of the idea of matching learner to tutorial style, then, seems to be empirically supported <u>and</u> feel intuitively correct. The demand, therefore, is for a teaching medium which can flexibly respond to the needs of a particular learner. In order to do this the medium must first build up or 'diagnose' each learner - in other words construct a student model. Student modelling is a specific development linked to Intelligent CAL, it is however unjustifiable to dismiss traditional CAL as having no comparable feature. Even a statement such as "You have scored 7 out of 10 - Well Done" is a kind of student model. As such, therefore, this review must firstly outline the state of the art of student modelling with respect to existing CAL in nursing before moving on to consider what ICAL promises.

Models of Learning and Teaching in Nursing CAL

In linear Drill & Practice programs which are classically Skinnerian a student model is conceptually undesirable in line with the idea that students' responses should be ignored. This format seems to be nowadays restricted to test knowledge acquisition - it was discussed earlier that such programs, eg in the University of Wyoming School of Nursing (described by Norman 1983), are perhaps better classified as Computer Assisted Assessment rather than Learning. The more usual case in current nursing CAL programs, however, is to pre-program 'branching' sequences which take the incorrectly responding learner into some frames which at the very least indicate that a wrong response has been given before continuing.

As an example of branching, the anatomy teaching program described by Richards et al (1986) indicates that the student has incorrectly labelled a heart component and then offers one more attempt. Although nothing in the way of remedial feedback is offered to the learner at the time, there is, however, a printout of scores achieved by students which the nurse tutor can peruse later. To the extent that this latter facility does not help the program decide on teaching strategy, the student model descibed here can be considered to be of the most basic type.

An examination of the Tutorial type of program reveals the next level of student modelling. This class of CAL in nursing, as exemplified by programs authored using the PLATO language, would seem at first sight to offer much more by way of a student model. For example, Bitzer & Boudreaux (1969) report that their obstetric program includes a facility to "make and retain a complete recording of the student's responses". Koch & Rankin (1987) are enthusiastic about this feature, which they see as a "map" of the interactions and decisions made by the learner.

On closer inspection, however, there is in reality little difference between this 'map' and the printout furnished by the anatomy program of Richards et al since the program itself makes no use of the analyses - it is once again a report destined for the nurse tutor. No teaching strategies are determined by the student model. There are in fact no program-driven student modelling features in this seminal program developed by Bitzer & Boudreaux. This assertion is justified by the fact that all branches to frames containing additional information or explanatory comment are effected by the learner herself pressing HELP or DICT (dictionary definition) keys.

The aspect of Tutorial programs which can, however, be interpreted in terms of student modelling (ie student response information being used to determine teaching strategy) is the provision of frames which appear automatically whenever the learner responds erroneously. While these pre-programed frames vary considerably in scope, they have one common feature in that their appearance is triggered by only the previous response typed in by the learner. There is no facility for the program to incrementally build up a picture of student performance over time.

To illustrate, at the most basic level the model of the student is one of "don't try to understand errors, just give the correct answer" - hence in Bratt & Vockell's (1986) program when a learner fails to correctly identify true and false statements concerning respiratory physical examination the program simply responds by displaying the correct answer after each error. The discussion is not implying criticism of such a teaching strategy (at least in this section) which may after all be appropriate for teaching this subject matter. What is being argued, however, is that to claim that this type of program is providing truly individualised learning is to seriously misrepresent the idea of a teacher tailoring teaching to his or her understanding a student.

A variant on the 'last response only' approach to student modelling is for the program author to try to anticipate each and every answer a learner might type in. For example, Mirin (1983) reports the following dialogue:

> *Computer*: One must prevent clot occlusion of the drainage tubes by initiating the action of ____________________ the tubes every hour and prn. Complete this sentence.
>
> *Student*: Irrigate or flush.
>
> *Computer*: Oh no ! you have just given the patient a hydro-thorax! Never irrigate a chest tube ! Please consider the consequences of your action and try again.

Other branching sequences in this program are more extensive. For example, when the student answers that he or she would not expect a patient to develop a pneumothorax after a lobectomy, the computer responds by taking the student interactively through a few frames designed to explain the mysteries of the relationship between intrapleural and atmospheric pressure. The principle, however, remains the same - this type of student modelling is *local* in that it is based on single responses from the student.

An important point can be made about this type of branching student model – it is paradoxically a regressive rather than a progressive feature of CAL. To elaborate, the task of anticipating every response by a student must be considerable – a point acknowledged by Mirin (1983). It follows therefore that anticipation of the *pattern* of student responses must be well-nigh impossible. If Student A gets questions 3, 5, 14 and 19 wrong then what does that mean? And what is this student's understanding of the material compared to Student B who gets 3,5,12 and 17 wrong? A branching format, lauded by the literature as an individualising device, can be seen to constrain student modelling (individualisation) rather than to enhance it.

By staying with the above example from Mirin's program, it is possible to identify a type of teaching strategy which differs from the "give correct answer" variety descibed above. This strategy, characterised by giving 'clues', can be illustrated in the following sequence which comes after the interaction about chest tubes ....

Clue 1: What action breaks up clots in the chest tube?
and if the student still doesn't know ...
Clue 2: Working one's fingers along the tube is called?
and finally ...
Clue 3: Don't give up! What is the act of relieving a cow of milk?

Mirin (1983) enthusiastically quotes an early statement on the benefits of student models in order to support her advocacy of branching programs. It is worth reproducing this quote in the light of the foregoing discussion:

> "(Lenvison) ... storing in the computer all the relevant data related to a given student, such as socioeconomic background, aptitudes, I.Q., interest profiles, vocabulary proficiency, motivation indices, and so on. The computer learning program then selects the appropriate learning sequence that best matches the entering behaviours of the student. Then, as the student works through the sequence, and responds actively to the materials presented, the computer 'learns' more about the student and continuously modifies and improves its feedback." (Bundy 1967)

The first part of this statement, 20 years on, seems as unrealistic as it is Utopian. It is not even clear that such machines would be desirable. The second part of Bundy's quote is prescient and refers more directly to the type of student model under discussion here. This vision has probably been realised in some ICAL programs to be outlined below. Nevertheless, Mirin

(1983) states baldly that the branching program discussed above has also achieved this "capability" - a claim put into considerable doubt by the analysis of theoretical concepts which was undertaken in Part 3 of this chapter.

There exists, however, a nursing CAL alternative to local student models. This type of program has facilities to build a 'longitudinal' model of any given student's interaction with the material. To the extent that an 'overall' (rather than local) model is constucted, it seems at first sight that there might be promise in these CAL programs. However a closer look at the nature of the model constructed brings disappointment in that they are invariably quantitative rather than qualitative. Put more bluntly, these Computer Assisted Assessment programs keep a score of number of questions correct and false.

An example where this approach is utilised with apparently good effect is in the drug dosage calculation program descibed by Timpke & Janney (1981). Students are required to achieve 100% correct answers in each section before being allowed to proceed. Since the questions are deliberately grouped into discrete categories (eg decimal point sums), the computer is able to construct a kind of student model which can then be used as a diagnosis of 'learning pathology' before offering remedial teaching. Freed of jargon, however, the student modelling is less impressive. What happens is that if a student fails to achieve 100% in the decimal sums then the computer directs them to specific pages in a remedial text. There is neither modelling of the specific difficulty of each student nor is there computer-based remedial teaching.

Quantitative assessments therefore are less than ideal as student models in CAL just as in traditional education. What is required would therefore seem to be a qualitative model. To illustrate the difference, consider a heart anatomy program which analysed in a longitudinal, qualitative fashion how each student was performing. Specific and individually-tailored comments could be offered, such as .... "you seem to be confusing lefthand parts with righthand parts - think of the diagrams as mirror images and try these questions". Unfortunately there is no evidence that student models both longitudinal and qualitative are in existence in nursing CAL programs of the Drill & Practice, and Tutorial type. Perhaps, however, it is more reasonable to expect the Simulation, and Inquiry & Discovery type of programs to contain adequate student modelling facilities - a possibility which can now be considered.

The short answer to the question is, alas, no. Moreover, the rationale behind these programs in nursing CAL seems actively to oppose the idea that the computer should build up and act on a student model. To take as an illustration the widely-quoted program 'Mr Malone' (Sweeney et al 1982), the idea here seems to be that the computer does its best to simulate the unfortunate patient Mr. Malone by offering realistic (albeit textual) answers to students' inquiries about, for example, pain being experienced. It would therefore not be fitting for 'Mr Malone' to be seen to be modelling the student. Nevertheless, there is not even a covert attempt to build a student model in this program. This is left to human tutors who read the careplans written by students who have assessed Mr Malone. This function, refered to by Koch & Rankin (1987) as "exteriorisation of learning", may indeed be useful but it is not student modelling by the computer.

It would seem therefore that student modelling is given even less emphasis in Simulation, and Inquiry & Discovery programs than in Drill & Practice, and Tutorial types. Given the implicit assumption in much of the literature that Simulations are somehow 'higher in the evolutionary scale', this point is surprising. One notable exception has, however, arisen from the keyboard of Britain's most prolific nurse CAL programer - Steven Ward. Ward has unfortunately neglected to publish (in the written sense) anything on the conceptual ideas behind his programs (a regretfully common failing in the UK). The discussion must therefore once again adduce from the specific program in question - OBSERVATIONS - what the principles might be.

OBSERVATIONS falls into the Inquiry & Discovery classification. The learner is told that a diabetic patient is about to be admitted, the task being to type in which observations and information one would require in order to plan the care of this patient. The program has one or two 'smart' features, for example if the student types 'urinalysis' the computer makes claer that this is much too vague by responding with 'which test?'. At the end of the task the program reveals that a student model of sorts has been built in that the order in which the student requested information (eg pulse first, respirations second etc) is contrasted with the 'ideal' order based on importance of information. The one rather serious problem with the program, pointed out by the nurse tutor who demonstrated it, is that the 'ideal' order is very different to many other experts' views - an almost inevitable consequence perhaps of 'one-person programing' rather than team effort.

To summarise this section on whether nursing CAL programs can be seen to build and utilise student models, the conclusion must be that anything resembling student models seems only to exist in few programs. Where student models do exist they are characterised by the poverty of their conception of what a model of the student can or should be. Moreover, it is even more rare to find programs which make teaching strategy *consequent* on model of the student. Two problems were identified. Firstly, some models were *local* in that pre-programmed feedback was available only with respect to the most recent response. Secondly, where longitudinal models were constructed their nature was invariably quantitative rather than qualitative.

To the extent that student models are integral to truly individualised teaching - whether traditional or via CAL - the frequently-made claim that nursing CAL offers individual learning cannot be supported by the conclusions of this analysis. Nevertheless, in defence of CAL it could be pointed out that the provision of a student model has never really been a declared aim of the typical programs. It was mentioned earlier that student modelling was, however, a declared aim and specific development of ICAL - the review therefore can now turn to student models and teaching strategies in Intelligent Computer Assisted Learning.

Models of Learning and Teaching in ICAL

If a single principle were isolated which could be said to characterise the difference between CAL and ICAL then that principle would possibly be the *dynamic* nature of the teaching response in ICAL. This principle, which may be called 'dynamic adaptation', itself decomposes into two related components.

1. ICAL tutors take a longitudinal, rather than cross-sectional, perspective. In the terms used by Ohlsson (1986), they focus on the fluctuating cognitive needs of a single learner over time, rather than on stable inter-individual differences.
2. Dynamic adaptation extends beyond performance indicators of learning. In other words, the program can change its teaching strategy and content consequent to the degree of understanding revealed by a student. Hence tutoring can be based on qualitative as well as on quantitative indices of learning - apparently a human-like feature.

These are high claims indeed. To paraphrase Ohlsson (1986) once more, can the computer be programed to generate exactly that question, explanation, example, counter-example, practice problem, illustration, activity, or demonstation which will be most helpful at any given moment for any given learner? The brief review below will seek to analyse the extent to which this promise has been achieved by working ICAL systems. Before embarking on that analysis, however, it is firstly necessary to describe the theoretical and practical means by which these promises might be achieved - in short, the anatomy of an ICAL system.

Commentators on ICAL seem generally to agree that design can be thought of as three interdependent components, the first two of which are of particular interest to the present discussion. The components are:

1. A student model which assesses and records the state of a learner's subject expertise.
2. A tutorial module which selects and delivers the computer's tutorial output.
3. A domain module which contains and deals with the subject expertise.

There are signs, however, that use of the word 'agreement' may be presumptious given the differences in emphasis which are beginning to emerge from the few research sites where such systems are under development (see Yazdani, 1986). It will be useful therefore to look more closely at these differences in order to identify the approaches of most promise to a proposed nursing ICAL system.

The first difference in emphasis relates to the construction of the student model by various systems. It is possible to identify three distinct approaches with respect to how best to represent a student's 'state of learning' - by looking at what the student currently knows, by looking at the errors he has made, or by putting together a 'simulation' of his or her performance so far.

Varieties of student model

a) Models based on current state of learner's knowledge In ICAL this variety of model might seem at first to be akin to the performance models discussed above (eg in the drug dosage calculation program). There is, however, an important if illusory difference in that it is not a *global* measure of *how much* the student knows that is built up; the goal rather is to understand the current state of knowledge *as a subset of* the expert's state of knowledge. Such models are usually termed 'overlay' models in the literature (Carr & Goldstein 1977) in line with the idea that the student is being represented almost as a series of tick-marks which is laid over the representation of the subject matter in order that the parts already known are shown up.

The theoretical basis of overlay models considerably pre-dates ICAL technology since in essence the approach rests upon the idea that subject matter can be broken down into a 'prerequisite hierarchy' (Gagne 1962). Hence in a problem solving task, acquisition of declarative knowledge must precede acquisition of procedural knowledge which in turn must precede acquisition of control knowledge of how to go about solving the problem.

To use the example offered earlier, an overlay model could be used to spot that a learner was consistently wrong in labelling parts of the heart in an anatomy program. These errors would be noted and repeat questions would be offered. In the event of performance continuing to fail then the 'lower' prequisite would be assumed to be not mastered, ie the student does not know about the heart at all. Accordingly, a teaching mode would be selected which dealt more with first principles. In this example an overlay model seems to be appropriately used. Nearly all ICAL systems, however, have sought to teach skills at a more complex level such as problem solving. As can now be discussed, it is possible that overlay representations are less than ideal for teaching complex cognitive skills.

The principal drawback with the overlay model approach, then, is the assumption that learners are like experts only less so. Student performance is assessed relative to the stored expert state, therefore learners, as imperfect experts, are taken as being quantitatively rather than qualitatively different to experts. To take an example of nursing problem solving such as pressure sore risk assessment and preventive care planning, so little is known of the cognitive progression from novice to expert practitioner in this area that it is perhaps unwarranted to contrast and compare performance. Moreover, there is no real attempt at understanding the nature of the student error. The next section will outline a diffrent variety of student model which sets out to overcome these difficulties.

To finally illustrate before moving on, suppose that an experienced ward sister had learnt through experience that a nursing assessment could safely omit 'nutritional state' so long as 'patient's build' was elicited (since one could be inferred from the other). Suppose in an ICAL simulation task a learner likewise omitted to ascertain 'nutritional state' - an overlay model would presumably applaud this, but what if the learner had simply forgotten that state of nutrition has a bearing on risk of development of pressure sores?

b) student models based on error analysis In this variety of student model the emphasis shifts from representation of what the student knows toward an analysis of the student's *erroneous* knowledge. To explain, it is considered that in the more usual case learners will misunderstand knowledge presented to them rather than completely fail to pick up anything at all. A slate will seldom be blank, therefore, but rather will contain erroneous procedures, false principles, and incorrect facts.

The student modelling strategy which is based on learner errors seeks to make a diagnosis of a particular student's errors or combination of errors. The usual method of achieving this diagnosis is through an error library of possible student errors to be contained within the system. The task becomes one of identifying which error best accounts for particular incorrect answers. Subsequently, appropriate remedial teaching best suited to the type of knowledge and type of error displayed can be delivered. The few projects - still largely experimental - to have used such an approach are those reported by Burton (1982), Sleeman (1982), and Anderson et al (1985). Respectively, these ICAL tutors teach arithmetic, algebra and, appropriately, computer programing in LISP.

To return the 'heart parts' example, an error-analysis model could be used to spot that a learner was consistently confusing 'left' parts with 'right' parts in an anatomy program. The diagnosis of learning pathology would therefore be that the learner was failing to compensate for mirror-imaged diagrams and an appropriate remedial teaching strategy could then be chosen. A few questions might be generated in order to test the validity of this hypothesis. In the event of performance continuing to fail then it might be assumed that the student does not know the difference between left and right. Accordingly, some teaching aimed at achieving this skill would be offered. As Fox (1984) notes, however, most of the existing systems have been constructed to teach problem solving. As will be argued below, it is the considerably more complex nature of these cognitive skills which leads both to the promise and to the pitfalls of such systems.

There are basic obstacles to the realisation of such systems. Not the least of these is the task of building up the error library - Anderson's LISP tutor contains 325 rules for planning and coding LISP programs and all of 475 error versions of those rules in its 'bug catalogue' (Anderson & Reiser, 1985). Notice once again the underlying supposition that the subject matter is represented in the 'correct' way within the system. By implication, therefore, considerable research effort is demanded prior to implementation. Not only must designers find out what domain experts know and how they represent their knowledge, but they must also do the same for students in order to assemble a comprehensive list of typical erroneous knowledge representations. Before going on to elaborate on the non-trivial nature of this task, however, the final variety of student model in ICAL systems can be outlined.

c) student models based on SIMULATIONS of learner performance In this variety of student model the goal is to describe the cognitions of an individual learner by constructing a simulation of his performance. Although by far the most ambitious of the varieties, there is a sense in which a simulation which will perform in the same way as the learner is the most logically suited to an expert system-based computer tutor. ICAL tutors are designed to solve problems in a manner akin to experts; it is therefore a rational goal for the system to be able to solve problems in a manner akin to learners. It therefore becomes possible not only to explain the learner's answer to a problem but also to trace the learner's steps toward arriving at that answer.

The idea of simulation models originated with Newell & Simon (1972). By focusing on think-aloud verbal protocols obtained from students completing a problem-solving task, the goal was to encode into rules both the declaritive and procedural knowledge evident in the protocols. A set of rules which contained the learner's encoding of the problem and the cognitive operations used to solve the task could be assembled and, in principle, run on computer as a simulation of that learner. More recently, Anderson and his co-workers (eg Reiser et al 1985) have sought to incorporate this model-tracing pricinple into ICAL using the large number of rules already stored within the program. The system constructs a model of the student based on his step-by-step approach to solving the task.

In some respects the simulation student model can be taken as encorporating the advantages of both the overlay approach and the error diagnosis approach. Hence if the learner solves the problem in the same manner as the 'internal expert' then that learner will be modelled using the set of knowledge units designated as expert. Similarly, the performance of the student (and the diagnosis of his errors) will be gauged through comparison of the student model with that of the expert model. The differences, however, are that the procedure followed by the student is modelled rather than simply putting tick-marks against evidence of knowledge acquired. Nevertheless there remains one major obstacle to the ultimate utility of student models, however adequate, and that is *the use to which they are put.* The only reason, after all, to incorporate a student model into an ICAL tutor is to utilise that model in deciding on choice of teaching strategy. It is, therefore, to a consideration of the teaching strategies used in ICAL systems that the discussion must now turn.

Teaching Strategies in ICAL Systems

By recalling the earlier descriptions of CAL tutors it is possible to list the range of teaching actions which any computer-tutor can perform. Thus the most basic computers are usually able to afford the opportunity for drill on practice problems, some can provide general feedback and others can be more specific in analysing learner performance. At other times, the computer might simply inform of the correct answer. Sometimes an explanation is provided, sometimes a hint is given, sometimes a question is asked. The ranges of actions, arguably, is quantitively rather than qualitively different to the range of actions offered by a human tutor. The crucial difference, however, lies with the expert human tutor's ability to select the teaching action which is appropriate with respect to a triad of factors - appropriate at that time, for that learner, and for that subject matter. As Ohlsson (1986) points out, computer tutors are as yet far removed from this ability.

To take the last-mentioned of this triad, subject matter, it can be seen that ICAL researchers have paid most attention to trying to analyse and represent subject matter in a fashion most appropriate for communicating that subject matter to the student. The collection of papers edited by Brachman & Levesque (1985) on knowledge representation provides ample evidence of this commitment. The reason that there is such a strong link between how the knowledge is represented within the computer and individualised instruction becomes plain when it is realised that there are many different ways in which material related to a single subject can be presented to learners. One student may learn best through being told the correct answer, another through demonstration of the problem being solved, and yet another through explanation of underlying principles. This implies, in turn, that the subject matter must be represented in a form which is deeper than simply a surface presentation format. The problems in achieving an adequate and verdical representation of a human expert's knowledge lies at the heart of this entire thesis and will be returned to in the next two chapters.

The other parts of the triad which have occupied ICAL researchers have been the subject of much of this section - the achievement of a cognitive diagnosis through the construction of a student model. A student model, properly functioning, should permit the program to register that an error is evident in the student's performance at any particular time. The point, however, is that the adequacy of the decision to make a tutorial intervention is considerably (and perhaps fatally) undermined by the failure to choose the appropriate intervention. However, Leinhardt & Greeno (1986) demonstrate through seminal work, there is little specifically known on how good (human) teachers do what they do. As Ohlsson (1986) complains, there is no *Handbook of Pedagogical Methods* to take down from the shelf and read off what the correct teaching strategy is for even basic parts of the curriculum like arithmetic. The high-level ideas that have emerged from educational psychology such as 'meaningful learning' and 'learning hierarchies' are not only lacking in general acceptance but, more importantly, are devoid of the degree of specificity required for computer based implementation of micro-level teaching strategies.

It should be pointed out that the criticisms levelled with respect to teaching strategy can also be levelled at student modelling. Again the charge of low ecological validity applies in that the student models described above are not based on established evidence of how it is that a good human teacher understands a student's mistakes. Fox (1984) succinctly suggests that this is because almost nothing is known about student modelling by teachers. The fortuitous result, however, of the needs and ambitions of ICAL programs is that there are signs of the necessary research being inspired.

Conclusions

The earlier analysis of 'individualised learning' within nursing CAL programs concluded that there was little justification for the claim. Provision of an adequate student model was seen as being central to realisation of truly individualised learning. Two inadequacies in traditional CAL student models were identified. Firstly, some models were *local* in that pre-programmed feedback was available only with respect to the most recent response. Secondly, where longitudinal models were constructed their nature was invariably quantitative rather than qualitative.

The discussion of progress toward achievement of the goal of individualised learning in ICAL programs suggested that several imaginative ideas might usefully be applied to computer-tutors in nursing. The keys to progress emerged as being answers to the criticisms of traditional CAL - flexible, global, and qualitative models of the student. Additional emphasis was also placed on the need for adequate analysis and representation of the subject matter in order that flexible tutoring might be offered.

One issue identified in the review of teaching and learning in traditional CAL was nevertheless left unresolved by the outline of ICAL advances - the selection of appropriate teaching strategy. The situation, for CAL and ICAL, remains one of the system designers relying upon hunches from themselves or from cooperating teachers on how to proceed in a given student-error situation. To the extent that human teachers might not have full access to their own expert knowledge on selection of best teaching strategy, this weak point will deserve more attention in the future.

ROLE OF CAL AND ICAL WITHIN EXISTING NURSE TEACHING SYSTEM

In this section an examination will be offered of the incorporation of computer-based innovations into the existing nurse teaching system. Firstly, the literature will be reviewed with respect to reported reaction of practising Nurse Educators. Secondly, some consideration will be given to the current thinking on the appropriate use of computer-based teaching.

CAL and the Nurse Tutor

A Nursing Times Educational Supplement (Quest, October 30 1985) begins with the banner headline of ...... Computers: A threat and a challenge ? Leaving aside the interesting choice of the preposition 'and' rather than 'or', the point being made by this title is that ambivalence exists within the nursing profession with respect to computer technology. The issue of ambivalence is hardly new to this discussion, indeed every previous section has found different angles from which to view the same coin.

There is, however, an angle to CAL-related ambivalence which has yet to considered - how 'real' teachers of nurses feel toward the innovation. A consideration of this important aspect will logically lead into a more general consideration of the proper role of CAL within the nursing curriculum. This way of ordering the discussion is logical simply because the power to adopt or eschew CAL lies with nurse educators themselves, although, as will be argued, the forces of technological advance have not always been known to respect human opinion.

The pressures to adopt CAL

When contributing to the nursing literature on CAL (or on computers in any aspect of nursing) it would seem to be almost *de rigeur* to write with a carrot in one hand and a stick in the other. Difficult though this may be, the point being made is that the tone of the literature fluctuates between exhortation and warning; between opportunity offered and opportunity lost; between 'gee whiz' and 'you had better'. Starting from the viewpoint of computers as revolutionary agents of society itself, the assumption is made that computers must necessarily revolutionalise nursing education and to ignore them would be to jeopardise the image and advancement of the profession. This thesis has already been rehearsed in this discussion, nevertheless here as well as elsewhere it is necessary to criticise unjustified assertions by means other than by making unjustified assertions - hence the need to look in more detail at these issues.

Since much of the rest of this chapter looks at the 'carrot' (educational benefit) aspect of pressure to adopt CAL, it is sensible here to concentrate on the external pressures which act on nurse educators. At the topmost level this pressure is political. D. Hoy (1985) suggests that the desire of governments to promote industry and jobs in the new technology has led to support for the use of computers in education as a means of ensuring an adequately prepared workforce. With respect to nursing, part of this argument seems undeniable for the following reasons:

1. nurses will increasingly encounter computers in their workplace - the wards. It is necessary therefore to expose nurses training to computers during their basic nursing programme.
2. administrative tasks in nursing education, such as the complicated learner allocation task, can be more efficiently performed by computer.

At one level these reasons cannot be refuted. Computers in the wards, for example, will undoubtedly require nurses who can operate them - even if the origin of this development (North America) can be traced to the desire to make more efficient the 'billing' system. At another level, however, the point has been missed since there is a considerable conceptual and practical difference between teaching nurses <u>about</u> computers and teaching nurses about nursing <u>by</u> computer. As an analogy, there is a considerable difference between teaching a nurse how to operate a sphygmomanometer and teaching her the principles of care of the hypertensive patient. Interestingly, the nursing curricula for teaching 'computer literacy' which have been described in the literature are heavily weighted towards traditional (ie non computer) teaching methods.

The conclusion which must be taken, once again, is that the debate about using computers in nursing education should be decided <u>only</u> on the merits of the medium as an educational tool.

Political and pragmatic pressure, therefore, offers reason for teaching computer literacy. At a level more proximal to nursing, however, the pressure from above has become translated and operationalised by the governing bodies of the nursing profession into terms which much more directly implicate the teaching of nursing by computer rather than just about computers themselves. In a seminal document entitled 'Project 2000 - A New Preparation for Practice' (UKCC 1986) the adoption of the assumption of CAL was made very clear......

> "The use of technological aids that will allow students to experience situations at a distance, and not to infringe unnecessarily upon privacy, will be crucial." (p. 46)

To anticipate the later section on Evaluation, this statement does not easily fit with other positions such as that of Baker (1984).......

positions such as that of Baker (1984).......

> "... neither educational progress nor curriculum improvement is a direct and inevitable consequence of technological advance." (p. 115)

It is concluded, therefore, that of the outside pressures on nursing education - from social, political, pragmatic, and educational directions - only the last of these has any direct relevance on the issue of whether CAL is in fact a medium worth adopting. There is, however, another direction from which pressure can bring to bear on nurse teachers - from within.

CAL as a perceived threat or challenge

There are two types of perceived threat which are at least mentioned in the literature. These may be categorised as 'role threats' and 'redundancy threats'. Before going on to discuss the first of these it is worth mentioning that there are no published papers on CAL as a threat which have been written by a nurse teacher who holds this view - once again the literature is 'one-directional' in that only proponents of the idea write about it. What tends to result is the proponent offering only a brief statement of criticism before going on to comprehensively refute it. This is regrettable. One-directional views are moreover regrettable *per se* and must be seen as having contributed to CAL having got as far as it has without being subject to serious critique. It will be necessary, therefore, to again adduce much about these threats.

Redundancy threats, firstly, is taken here to cover the idea that CAL will so completely usurp the human teacher in terms of skill and cost that they will be reduced to mere machine-minders or even replaced altogether. Initial reaction to such an extraordinary idea is to see it as absurd in the context of current CAL. The idea is absurd given even the most optimistic of informed predictions. The more circumspect reaction in the literature, however, is to take the fear seriously and embark on careful reassurance that there will always be a place for the teacher (R. Hoy 1983) or that some teaching will always require the human touch (Koch & Rankin 1987).

There are, however, problems with taking this threat seriously. It might be attempted, for example, to trace the origins of redundancy fears with respect to CAL to the writings of Skinner (1954) on Programmed Instruction. Skinner's declaration which is possibly most to blame is that as a mere reinforcing mechanism, the teacher is out of date. However, to lift this quote out of context is to misrepresent his view which was that teachers would be freed to concentrate on the uniquely-human aspects of teaching. Nevertheless, D. Hoy (1985) reminds us that the idea is still with us - Huckaby et al (1979) suggest that CAL in nursing will lead to accomodation of increasing numbers of students. The conclusion from many parts of this chapter is that this assertion is both premature and beyond the evidence.

Rather than rehearse these 'serious' arguments over again, then, this discussion is going to rely more on its initial reaction of absurdity. After all, as mentioned earlier the literature reports no proponents of this belief and it is not in this author's experience to meet anyone with a working knowledge of CAL who believes such a thing. Rather, this type of redundancy threat seems more akin to common 'techno-fear' related to, for example, robots replacing workers in car factories - a very different scenario to CAL replacing human teachers. It is of more interest, therefore, to look behind this facile view of perceived threat in order to identify what might be the true source of worry or doubt. This aspect brings into focus the other type of perceived threat identified - 'role threats'.

The literature seems to assume a subtle distinction between the different types of fear related to the nurse teacher and his or her role. Firstly, there is the threat 'to me as a teacher' and secondly, there is the threat to 'my ability to master this new teaching tool'. The end result, which Koch & Rankin (1987) dub a form of Luddism, is the same in that hostility and scepticism (see comments reported by Norman 1983, p.8) prevail while realistic funding is withheld.

It must be stressed once again, however, that no empirical evidence supports the subjective feelings of how the proponents of CAL see the psychology of their detractors. The approach is rather to offer presciptions on how best to overcome the role insecurity of less-enlightened colleagues - prescriptions which invariably include converting scepticism by means of "hands-on" experience of computers and CAL. Yet there remains a sense in which there is an inherent contradiction in the reassurance offered by advocates of CAL. On one hand the benefits of the medium for teachers are extolled while on the other there is the claim (even if unrealistic) that CAL programs can mimic human teaching activities. Seen from this perspective it is perhaps small wonder that teachers remain resistant.

The most influential subjective evaluation of what underlies teachers' 'worries about mastering this new tool' is Rushby's (1980) suggestion that specialist mathematical-type expertise is a prequisite. Perhaps also there is something in Grobe's (1984) suggestion that gender is an important factor since computers seem generally to be very much a male preserve in society. While there may indeed be something in these ideas, what is perhaps more certain is that teacher's perceptions cannot be taken as static over the years. Both technology and gender perceptions are changing. There are,for example, rapid advances being made in the field of human-machine interaction which serve to ensure that each single year sees the introduction of more easily used and 'friendly' computers.

The suggestion which does not, however, appear in the literature is that it might be the CAL proponents who are the 'blinkered' ones. Perhaps CAL sceptics, looking at the typical nursing program available, are sufficiently percipient to recognise impoverished teaching material when they see it. This idea is offered only to illustrate a point. To the extent that it is untested, it is no more or no less valid than some of the foregoing ideas about role insecurity in nursing teachers.

The second way in which CAL is seen as threatening role - 'to me as a teacher' - is slightly more tangible. Here the focus becomes one of altering the teachers' role rather than usurping their role. In Morgan's (1977) terms, the role of the teacher will be modified to include being a resource person and learning manager. In the nursing literature the terms more commonly used are those coined by Ball & Hannah (1984) - teachers must become 'facilitators, moderators, and coordinators'. At worst, apparently, teachers will continue to schedule timetables for their student's computer experience while at best their role will become one of offering specific remedial instruction on topics which the computer has identified as weak points for specific students. Once again - difficult though it may be - this idea deserves to be examined more closely.

The central idea, then, seems to be one of devolution of responsibility. What is not clear, however, is *to whom* the responsibility is being devolved. On the one hand the mainly American CAL literature seems to assume that responsibility is being devolved to the computer. A different perspective, however, would be to view this shift as being devolution to the student for responsibility of learning - an idea certainly consistent with current UK nurse education trends and philosophy. Given this perspective, therefore, it becomes clear that the statements in the nursing literature concerning teacher role insecurity with respect to CAL are once again seeking to oversimplify a much larger issue.

Alongside current major shifts in teaching philosophy, CAL introduction must surely rank as relatively trivial or, perhaps, as a scapegoat rather than a cause of role insecurity. Koch & Rankin (1984) provide an example of the unjustified over-promotion of the effect of CAL by chastising nurse educators in the following strong terms ...

> "(teachers) see the computer as a 'young pretender'..... Sadly, this may be a reflection on how they perceive themselves as teachers, for surely the

> authority that any teacher lays claim to is at best provisional. Teachers, by definition, must be challenged; they should rely on challenge and innovation to maintain dynamism and flexibility in the learning environment. To argue otherwise is to advocate stagnation or, at worst, authoritarianism." (p. 17)

This is overstatement - for two reasons. Firstly, in the context of the earlier outline of the changes in educational philosophy it is clear that CAL is of relatively small consequence. The second reason is of more pertinence to this discussion and is repeated once again - CAL in its current state of development has not yet earned a status which merits a chastisement of such ferocity. Teachers, moreover, are no strangers to 'wonder developments' in educational practice. The onus of demonstrating efficacy rests with the developers of CAL, not on the receivers of CAL. Teachers do, however, have the responsibility of open-mindedeness and of withholding judgement until that point has been reached.

CAL and the Nursing Curriculum

It was suggested earlier that the tone of the nursing CAL literature was, with one or two exceptions, almost grandiose about what CAL was doing (or was about to do) to the nursing curriculum. The impression gained was that there was no curriculum topic which could not be wonderfully delivered via the computer screen. Two points dispel the illusion. Firstly, some more recent literature which is considerably more circumspect in terms of claims about appropriate subject matter for teaching by computer. Secondly, a closer look at exactly which subject domains are currently covered by programs.

The first of these points - the more restrained tone of some recent literature - is an encouraging sign in that it denotes a 'maturing' of the field. Koch & Rankin (1987) support this conclusion by offering some previously unheard of criticism about some aspects of CAL. On the point of appropriate subject matter, the tone now seems to be one of 'what CAL cannot teach' rather than 'CAL can teach everything'. To use Koch & Rankin's (1987) illustration, an interactive program could not deal with the philosophical issues of nursing as well as a small-group discussion could. Nevertheless this example is perhaps over-extreme given the earlier criticisms of many of the so-called 'problem solving' programs which are in existence. Until programs can be developed which are truly individual then the conclusion of this discussion is that traditional CAL is not an appropriate medium for teaching this higher-order skill.

The second line of evidence which acts to temper the over-sell of CAL was mentioned as being a closer look at the range of avaliable programs in a Scottish College of Nursing and Midwifery. At the time of the author's visit, this College was arguably the best-equiped in Scotland due both to an enthusiastic Director and to the two Tutors known nationally through their program creation and writing. Some twenty-five programs falling into four categories were available for student use. The categories comprised Anatomy & Physiology (13 programs), Nursing (6), Ward Management (3), and Drug Dosages (2). Given that the Nursing programs, by virtue of their small number, necessarily representated only a fragment of the topics in the nursing curriculum then the conclusion is that Tutorial-type Anatomy & Physiology is currently the principal application area of CAL. If a reminder were needed, it is clear that CAL in the UK has not yet evolved further than the experimental/developmental stage.

Since there are as yet no available ICAL programs in nurse education, it would be speculative only to discuss the role of such programs in the curriculum. One point is nevertheless relevant here - the siting of computer-based training. All existing CAL programs both in the UK and in the USA are located within Colleges of Nursing. The logic of this, however, has never been articulated. A drug dosage program, for example, might be appropriately sited in a ward environment. ICAL programs which operate in the Expert System mode (ie learners can consult the program about patients they are currently working with) would be most sensibly sited in the wards. Therefore although a discussion of the role of ICAL in the curriculum would be at best hypothetical, it could be suggested that the 'threat' such a system might offer is considerable.

In conclusion, there has been little other than superficial analysis of the role of CAL alongside the existing nursing curriculum and alongside existing nurse tutors. Not only is it unclear what the current position is in this respect, but also it is difficult to accurately predict what the future position of CAL is given that program creation and implementation is as yet embryonic. One conclusion, however, is clear - the need for careful academic research endeavour in order that light rather than heat might be generated.

EVALUATING THE EFFECTIVENESS OF CAL AND ICAL

Evaluation is an all-embracing term of such breadth as to raise doubts about it's usefulness. At the outset it is therefore necessary to establish what is to be taken here as 'evaluation'. More importantly, it is necessary to restrict the working definition to those aspects of evaluation which are pertinent to the overall project.

That several calls for evaluation of nursing CAL have been made seems to confirm that evaluation is a neglected step (eg Koch and Rankin 1987, Ball and Hannah 1986). What is meant by evaluation? Evaluation can be taken, in nursing CAL, to refer to measurement of goal attainment related to the system's objectives. This definition begs several questions - measurement, for example, can refer to the system itself, the personnel involved, the impact and general acceptance of the system, the changes in knowledge and behaviour before and after the system, and so on. Moreover, measurement implies not only reliability and validity but also the existence of a metric against which to measure. The issue is complex and suggests to the reviewer that a narrow focus is required on those studies which take seriously the need to address these issues before offering 'an evaluation'.

Search of the literature reveals that in fact there have been several studies conducted with evaluation in mind. These studies can be categorised into those which measure cost-effectiveness, those which measure educational transfer, and those which measure the actual system itself in terms of variables such as quality of presentation. Taking the criteria of relevance to the present project, it is clear that the first two of these categories is of interest. Using the same criteria, an additional category demands attention when broadening the review to include ICAL - the extent to which the performance of the system mirrors the performance of the nurses on whom the system was based. On various parameters, therefore, a review at this point will begin to explore the ways in which the ICAL being put together in this project can itself be evaluated.

At this point the review can begin by looking at the first (and least complex) category ....

Evaluation of Cost- Effectiveness

Estimates vary rather widely on how long it can take for CAL programs to be developed, written and programed into computer language. Writing of general education CAL, Kochar & McLean (1985) give the following figures (based on Manpower Services Commission data) for time in hours taken to develop one hour of courseware :

Conventional course	- average time	5 hours	Range 2 - 10
CAL course	- average time	150 hours	Range 50-250

In nursing CAL, estimates for development are even higher or much lower. Kirchoff & Holzemer (1979) estimate between 200-400 hours for 1 hour of CAL while Bitzer (cited in Norman 1983 b p.9) claims that between 12 and 80 hours are necessary. The notable factors which must be borne in mind with respect to Bitzer's figures are firstly her 25 years of experience in the task and secondly her use of 'authoring language'. An authoring language can be thought of as a CAL program 'shell' with only specific nursing facts and comments requiring to be added. Interestingly, Norman (1983 b) found that users of this authoring language in a separate site from Bitzer themselves estimated 360 hours to create 1 hour of CAL.

To put development time finally into perspective, Kirchoff & Holzemer (1979) found that "approximately two months were required to translate an *existing* assignment of postoperative nursing care to a computer-assisted instructional program" (present author's italics).

Given these development times and their concomitant values in money terms, there is clear support for the observation made earlier that the onus must be on proponents of CAL to demonstrate the value of the medium. Most easily this value can begin to be demonstrated if evidence could be found to back the claims that significantly lower tutor involvement goes with running students through unattended CAL programs. Similarly, evidence is required that the more students to use a program then the lower the development costs become - Kochar & McLean (1985) estimate that 300 trainees need to take a computer-based program before it becomes more economic than a conventional alternative.

Unfortunately, the literature on CAL in both nursing and general education reveals extremely little in the way of comparative cost analyses of program implementations. Hannah (1983) quotes a single nursing study by Larsen (1983) which found that there was indeed significant cost benefit in using CAL when compared with traditional teaching strategies when the subject material was the teaching calculation of intravenous flow rates.

Nevertheless a lone study is far from convincing in terms of weight of evidence. Even if it were, it could well be that this type of program - typically cheap to develop - is cost effective when compared to traditional alternatives while more complex therefore more expensive programs may not be cost effective. The need for work in this area is as urgent as its omission is remarkable.

ICAL seems to fare no better in respect of apparently large development costs. As O'Shea & Self (1984) point out, the few systems which are in existence have each been the fruits of PhD projects. The ACT* system of Anderson and his co-workers (Anderson 1983), for example, has been developed over a decade of rolling funding of several researchers. The single program which might be classed as a nursing ICAL program - COMMES (Evans 1983) - has recently been released on to the market along with the sales pitch that it has taken 12 years to develop and 4 million dollars. Nevertheless, it becomes less meaningful to think of the costs of single hours of teaching time of such systems given the innovative nature and different nature of these programs. COMMES, for example, can be closely likened to an enclyopaedia of medical and nursing factual knowledge which can be consulted rather than a program which will deliver a set lesson.

To compare ICAL with CAL in terms of cost, then, would only be possible if the original CAL programs of 20 years ago were examined. Moreover, given the ambitious subject domains of ICAL systems, comparative development costs would only be meaningful if an estimate of 'development difficulty' was to be encorporated. Since this factor clearly applies to comparison of CAL programs both with teachers and with other CAL programs, at this point the discussion will profit from moving to the second category of evaluation.

Evaluation of Educational Effectiveness - Qualitative and Quantitative

The question of cost-effectiveness cannot ignore the ultimate metric - the measurement of educational transfer. The expense of an alternative teaching method must always be considered relative to the effectiveness of that method versus traditional alternatives. If, for example, a new and very costly method were shown to be a radical improvement on traditional techniques then the expense can be reconsidered. Clearly also the subject domain becomes important - if improved education led to direct saving of lives or even of expensive resources then once again development cost can be reconsidered. These questions, however, imply that educational transfer can be and has been reliably demonstrated with respect to CAL - a conclusion which this review will find extremely difficult to support.

Educational effectiveness, for the purposes of this review, can be measured either directly or indirectly. By direct measurement it is meant that some form of pre and post index is taken of student's knowledge. Indirect refers more to attitudinal measurement. Cognitive versus affective would be an alternative schema - no study has been located where behavioural consequences have been measured in terms of improved nursing practice consequent on exposure to a CAL program.

Indirect measures

Studies identified which fall into this category seem to rest on two articles of faith - firstly that feelings toward an experience will correlate with educational transfer, and secondly that the experience has been causal in producing that alleged educational transfer. Hence if a student reports that a CAL learning experience has been 'valuable' then the first article of faith states that learning has been facilitated by this positive feeling while the second assumption holds that this supposed facilitation has been brought about by the CAL medium.

These assumptions may be valid; on the other hand they may not. The point is that attitudinal evaluation studies must themselves be evaluated in the light of how adequately they address these two assumptions. In one study, for example, Rankin, Koch, and McGuire (1986) administered questionnaires to nursing students who had been using the CAL facility in a Scottish College of Nursing and Midwivery. Without giving details how it was carried out or the frequencies involved, the authors divided the responses into 'positive' and 'negative' categories. The authors then went on to report examples of comments and to draw conclusions about educational effectiveness. Aside from reliability and validity issues, the point is that neither assumption has been addressed.

A more specific yet similarly flawed 'opinion poll' is reported by Richards, Al-Basri, and Minshull (1986). Here the nurses who had completed a CAL program which taught cardiac anatomy were administered a questionnaire. The questionnaire, although anonymous, required the students to fill in what mark they achieved in the computer-administered test. Of interest here is the written responses to the question "In your view what are the benefits of such a system?" - all the answers to this rather leading question which were reported were favourable. Another question asked students how they felt about whether their mark was above or below what they expected - although no figures were reported apparently those who did well were pleased and those who did badly were grateful. Clearly the review has not yet reached the point where it can take conclusions on indirect measurement of educational effectiveness.

A more comprehensive study is reported by Huckaby et al (1979) where the aim was to design a controlled trial of CAL versus lecture-discussion in the teaching of management of hypertension. Learning measures were measured before and after the learning experience and Likert-type scales of affective response to the teaching methods were administered on post-test. Potentially, then, this study holds much promise for addressing the articles of faith outlined earlier. Indeed this study, by four United States assistant professors, has become widely quoted as evidence of CAL educational effectiveness (eg. Norman 1983, Pleasance 1984). As the authors state, there was a trend toward the CAL group giving a more positive evaluation towards this form of teaching.

However, the predicted difference between groups on the affective measures failed to achieve statistical significance. Nevertheless the authors stated that "the trend was in the predicted direction". It is hardly permissable, even in a well-designed experiment, to draw this conclusion (it was drawn seven times) - especially when future readers may lack the skills to evaluate the quality of the research. This experiment, moreover, had several serious design flaws in the methodology which severely limit such conclusions as could be drawn. Each group actually received equal lecture-discussion teaching *followed by* either exposure to a CAL program or, for the controls, exposure to a "reading assignment method". The difference in terms of novelty value and stimulation between these exposures weighs heavily in favour of the experimental group. Later this study will be revisited when discussing the cognitive measurement of educational effectiveness (significant results are reported). For now, however, it is sufficient to note that the cognitive post-test for each group used multiple-choice questions related to patient case studies - the CAL program comprised two hours of question and feedback related to patient case studies.

One final study which can be examined in relation to attitudinal measures of educational effectiveness is that of Kirchhoff and Holzemer (1979). In this study the design was to obtain 16 measures from 100 nursing students related to their learning styles, attitudes toward the PLATO CAL program, and their experience with this type of learning media or the nursing domain being taught (postoperative care). All of these variables were entered into a regression analysis along with a post-PLATO score on knowledge gained, the goal being to identify which student variables most powerfully predicted the observed learning. Interestingly for the authors, the students' perception of the degree of dullness of learning via PLATO was the highest predictor of learning (the less dull it was perceived then the more they learned).

On first inspection this result would seem to support that educational effectiveness can be attributed to CAL by indirect attitudinal measurement. However the only other variables which had more than minimal predictive power were:

1. the length of time between using PLATO and being tested (the longer the interval the more students forgot),
2. the current ward the student was working in (surgical ward students did best), and
3. previous experience with PLATO (the more previous use the better the learning).

None of these variables is impressive evidence for the effectiveness of CAL rather than any other teaching method. Crucial points are that there was no control group in the experiment, non-volunteers were excluded, and subjects were aware that they were taking part in an evaluation of a method 'believed in' by their tutors. Hence the 'dullness' finding begins to lose appeal.

In closing this section of the review on quests to demonstrate educational effectiveness via attitudinal measurement, the conclusion must be made that this difficult methodological venture has thus far been less than successful in demonstrating reliable effect. The review can now turn to studies which have used measures of knowledge gained.

Direct measures

Direct measures of learning in the context of evaluation refers here to the search for enhancement of knowledge levels which are contingent on exposure to a CAL experience. In fact, the achievement of this aim has been shown to be an exceedingly difficult experimental question, the major problem being the focus on the learning effect being 'contingent' on a particular teaching method. Hence if CAL is that method then it is insufficient to assign one group of students to CAL exposure and a comparison to 'no CAL' – not only must the control group receive a valid substitute method but also that substitute is required to be matched to CAL on all of the many variables which hold the potential for affecting learning enhancement. Quite apart from matched content of the methods, additional variables include factors such as duration of exposure and perceived value of method by learners. This assumes, furthermore, that the validity of the experiment has been established along with reliability of criterion measures. With these requirements at hand, therefore, the key studies in the literature can be inspected.

The studies by Huckaby et al (1979) and Kirchoff and Holzemer (1979) have been introduced with respect to indirect measures employed. It is worthwhile revisiting these frequently-quoted papers in order to evaluate the extent to which methodological problems have been overcome. Huckaby et al, firstly, state that the experimental (CAL exposed) group transferred their learning to clinical practice at a significantly improved level when compared to controls. The groups, however, did not receive comparable interventions since the control subjects were given a 'reading assignment' whereas experimental subjects received two hours of novel and possibly exciting CAL teaching.

The internal validity of this experiment was further jeapordised by the CAL group being sensitised to the criterion measure by receiving identical teaching to the nature of the post-test.

Finally, the conclusion that transfer to 'clinical practice' was achieved (thereby implying external validity) cannot be accepted given that the post-test was a paper-and-pencil case study followed by multiple-choice questions. In fact, the CAL group did not actually learn more than controls - the basis for this paper being so widely taken as evidence of the success of CAL rests in the finding that a paper case study was scored better by one group versus another.

The study by Kirchoff and Holzemer (1979) must also be seriously criticised on the terms set out for acceptable experimental method since no control group was utilised - all subjects were exposed to CAL. The authors almost accept this conclusion since they offer "only a qualified yes" to the question 'did students learn the material on the PLATO program?' since there was no control group of students not being exposed to the program. They remain confident, however, that PLATO is an "effective teaching technique" (p. 28), despite simultaneously accepting that there cannot be an "implication that the computer program ia a *better* technique than the former written assignments" (present author's italics).

An approach by Hoffer et al (1974) seems at first sight to have taken into account the need for a properly established control group and for valid baseline measures. Hence 34 Registered Nurses were randomly assigned to either a CAL experience (cardiopulmonary resuscitation) or to a control group. No differences on scores for a knowledge level pre-test were demonstrated, yet on post-test the CAL group scored significantly better ($p<.05$). There are, however, limitations. Firstly, the pre-test can be taken as a sign to subjects that they are participating in an experiment - having the possible effect that experimental subjects rise to greater endeavour. The opposite effect is possible for controls, especially since all subjects came from the same hospital and therefore contamination between groups could occur. The final (and fatal) flaw in this study's design is that the control group actually received no teaching whatsoever - hence 90 hours of teaching are being compared to zero hours of teaching.

Between group contamination was eliminated by Bratt and Vockell (1986) since the two groups set up passed through the nursing college a year apart. However another problem arises as one is avoided in that there was no random assignment or matching of groups - one class was essentially compared to another. The post-test administered after teaching on respiratory assessment was part of the curriculum and would therefore be less likely to be seen as part of the experiment. Each group spent the same amount of time on the topic; it therefore seems impressive that the CAL group scored significantly better on the post-test. Nevertheless, the fundamental design flaw has not yet been overcome in that a traditional teaching technique is being directly compared to a novel technique. Since the novel technique is in fact extremely novel - the class had no experience on computer before - then the strong possibility remains that the students' 'mental set' cannot be taken to be constant across groups.

By ignoring the subjective element in program evaluation and by concentrating on quantitative measurement, the danger is that the 'direct' approach will fail simply because of the very real difficulties of designing a sufficiently rigorous yet true-to-life experiment. This observation, powerfully argued by Cronbach (1980), is rather ignored by the most simplistic evaluation studies such as that reported by Timke and Janney (1981). Here the strategy was to compare the exam pass rates before and after the implementation of a CAL program designed to teach nurses the arithmetic of drug dosage calculation. Before CAL, 11 students out of 28 failed. After CAL, 32 out of 32 passed. An undeniable demonstration of success? Unfortunately not. The point is that the conclusion cannot be taken that CAL and CAL only can produce this change - leaving aside the fact that several extra hours were spent by students on CAL; perhaps a completely different innovation would produce equally impressive results.

One final study can be examined in this quest for reliable evidence from a direct, quantatative attempt to establish the effectiveness of CAL in nursing. The rigour of the design of the experiment carried out by Valish (1975) seems at first sight best able to withstand criticism. Hence the method employed was a post-test only comparison of randomly-assigned experimental and control groups where all subjects were exposed to CAL. In an apparently elegant variation, there were 3 CAL programs (on shock, parenteral feeding, and leadership) each of which a third of subjects completed. The scores on post-test of, for example, 'shock' knowledge from the subjects who completed the shock program could therefore be compared to the scores of subjects who completed a CAL program, but not the shock program.

Although the impression is that the 'novelty' effect is being adequately controlled in this design, a moment's thought reveals a serious flaw in that the 'shock' group is actually receiving extra-curricular teaching on shock while the two other groups are receiving nothing whatsoever on shock. It is, however, unnecessary to have at hand this criticism - there were no significant differences between any of the groups in terms of how much the nurses knowledge had been augmented by exposure to a CAL program.

Conclusion

Evaluation of the educational effectiveness of CAL in nursing has not been adequately demonstated through use of either direct or indirect measures. This conclusion should not be taken to be wholly a comment on the methodological designs which have been used but rather a comment on the extreme difficulty of the task. House (1980) argues that it is largely for this reason that 'objective' or 'goal-based' evaluation has increasingly been abandoned in favour of more qualitative mehtods such as case-study or naturalistic methods. The focus here becomes one of taking into account the context within which an educational program is implemented in addition to focusing on the users' perceptions of the innovation. Hoy, D. (1985) is as enthusiastic for this approach for nursing as he is caustic about the adoption of the 'hard' approach by evaluators of nursing CAL. Moreover, Billings (1984) has argued from a nursing perspective that CAL evauation should strive for "verification through observation".

The flavour of this review and of these comments is that assurance of quality must begin at the earliest component stages of system construction and follow through until the point has been reached when performance can be measured. It is easier to infer justifiable user-satisfaction in occupiers of houses known to be soundly designed and built of good brick. Whether the house sells, or whether it looks good, is not a serious test of the quality of that house. Although an important paper by Grobe (1983) has begun this change of emphasis in evaluation of nursing CAL, there is clearly a need for a more muti-dimensional scheme for appraisal.

An eminently suitable framework is suggested by the current interest within the NHS on the assessment of quality - with the goal of quality assurance - where approaches closely focus on the classic formula (eg Donabedian 1976) of scrutinising <u>process</u> (analysing action), <u>structure</u> (the environment of action), and/or <u>outcome</u> (the effect of action). Just as quality assurance has become integral to healthcare delivery, so too it becomes analogous to evaluation of an educational model. Hence, not only should the effectiveness (outcome) of a CAL program be evaluated, but also the educational appraoch (process) and the nature of teaching model employed can be incorporated into the evaluation.

The conclusion, then, is that evaluation of the effectiveness of traditional CAL is neither an exact science nor a settled issue. With these points in mind, the discussion should now move to an examination of how evaluation has been attempted in ICAL. Here, however, there arises the immediate problem of no reported instances of educational evaluation of the innovation. It is not quite accurate to take as the reason for this the relative newness of the medium since there are now several up-and-running systems in various educational fields. Rather such evaluation reports that have been published are concerned much more with the <u>validation</u> of the systems in terms of the extent to which the system comes to the correct decision. In the light of the lessons to be learnt from the review of CAL, therefore, an examination of these reports will be undertaken in order that the direction for evaluation of the system which is the product of the present project can be determined.

Evaluation of ICAL and Expert Systems

It is perhaps worthwhile beginning this crucial section by posing a superficially simple question - how should an 'intelligent teaching machine' in clinical nursing education be evaluated? When the nature of a ICAL system is considered, it becomes immediately clear that the approaches which have been adopted for CAL are of only partial applicability. Of much more central importance is the evaluation of the composition and performance of the system if that system is claimed to be a model of how human experts perform within a domain. What is required, then, is a suitable framework within which to all pertinent aspects of an ICAL system can be elaborated. It is suggested that the process, structure, outcome framework is a suitable candidate for such a framework.

Methodological principles of assessing quality can also be seen to easily cross boundaries into the research domain. Thus, for 'process', read identifying the items or units of the research domain and measuring their reliability; for structure, read assessing the validity of the representations employed and within which these items or units relate; and for outcome, read analysing the results of the research effort. It is also clear that an approach to evaluation of a cross-disciplinary product must, as Brown, Tanner and Padrick (1984) argue, seek to adopt perspectives on appraisal from each of these disciplines. Hence nursing, psychology and computing perspectives must combine if a complete evaluation is to be achieved.

It has furthermore been argued from the healthcare practictioner's standpoint that measuring quality by focusing on only one component in the process-structure-outcome formula is an incomplete approach and that combinations (eg process and outcome) should be measured (Bloch 1975, Mates and Sidel 1981). The evaluation studies of expert systems which have been published, reviewed by Gaschnig et al (1983), have primarily focused on 'decision correctness' as the outcome measure of importance - hence Buchanan and Shortliffe (1984) report figures which demonstrate the accuracy of diagnosis of their medical diagnostic system. The contention here, however, is that this conventional approach to the evaluation of expert systems - that of applying outcome tests only ('does it work or not?' or 'how well does it work?') - is an approach which leads to the same problems as witnessed with CAL evaluation.

Outcome-focused strategies is a particularly incomplete evaluation of a multivariate system that has been designed to emulate human information processing. This position is supported by Liebowitz (1986) who, in a substantial review of expert system evaluation, observes that it ".....has centered mainly on the use of blind verification studies and modified Turing tests". Casting the net wider, more general support for this view comes from Ferrari (1986) who records that it is a lamentable feature of computer science practitioners that their systems-performance evaluation is quite emancipated from software considerations and computer architecture and operating system considerations. For computer systems in general, then, he argues in favour of an integrative evaluation solution where evaluation features at all levels of system evolution.

Evaluation of an ICAL system, therefore, should focus on each of the following stages:

1. Knowledge acquisition (process).
2. Implementation of the knowledge within a machine (structure).
3. Product-testing (outcome)

The aim of the present project is to construct an expert system with a dual goal of clinical support and education for ward-based nurses. If each stage of this construction is evaluated as it proceeds then the 'gestalt' of these appraisals will, it is hoped, exceed a unidimensional approach.

This prescription for evaluation will underpin the present project. For the moment, however, it is important to establish the extent to which other work in this field can be said to support such a scheme for action. Richer (1986), for example, makes the point that it is difficult to evaluate any system without consideration of the system's intended use and users. Nevertheless, the thrust of Richer's argument is not to promote a single dimension of evaluation but rather to show that intended users cannot hope to be satisfied with a product which has not taken seriously the need for assurance of quality throughout the construction period.

Gaschnig et al (1983) have criticised the dependence on validation of decision accuracy as a metric for expert system evaluation and identify the following desirable components for an evaluation of expert systems:

1. Quality of the system's decisions and advice
2. Correctness of the reasoning techniques used
3. Quality of the human-computer interaction (both its content and the mechanical issues involved)
4. System's efficiency
5. Cost-effectiveness.

The approach, it becomes clear, is tending more toward process and structure as an adjunct to (rather than an alternative to) the outcome-focused approach. Liebowitz (1986) has similarly offered a set of evaluation criteria for expert system evaluation which incorporates many of these components - namely, how accurate is the methodology; what resources are needed; how sophisticated is it; and can it be easily maintained?

There is a sense, nevertheless, in which evaluation criteria should be individually tailored to the system at hand. It follows that since the present project aims to achieve something rather atypical of expert systems - construct a program which is a model or emulation of the cognitive skills and styles used by expert clinical nurses - then a shift of emphasis will become appropriate whem evaluating. Cost-effectiveness, for example, becomes less important when the research is fundamental than does reliability and validity of the methods used in the initial study of expert nurses' information processing. It will now be useful, therefore, to outline the important criteria which will be used when evaluating the present project.

Following the scheme adopted above, an initial subdivision needs to be made of the Process Evaluation phase:

Process evaluation 1 - the surface knowledge base

Attending to evaluation of the surface knowledge base must begin with reasoned choice of a domain suitable for study and modelling within an ICAL system. It then becomes important to define the nature of the knowledge which expert nurses will hold with respect to the chosen domain. It follows also that considerable attention should be given to the definition and identification of these expert nurses. Only when these preliminary stages have been completed can a methodological package be devised for the purpose of eliciting that expert knowledge from nurses.

The methodological package must seek to maximise the reliability and internal consistency of elicited knowledge. It becomes important to establish from first principles the factors (or attributes) used by nurses when reasoning within the chosen domain. Similarly, it will be necessary to develop a reliable understanding of the way in which nurses sub-classify their patients on each factor. Put simply, if the ICAL system is to have external validity then it must use the language of nurses.

Process evaluation 2 - the deep knowledge base

Fink et al (1986) argue that a clearly neglected area of expert system construction is emulation of the deeper knowledge held by humans which is used to achieve a teaching role which is closer to that of an 'intelligent arguing colleague' than mere decision support. To obtain this knowledge it will be necessary to devise cognitive science techniques which will provide data on the information processing behaviour of nurses. By employing experimental techniques the rationale, once again, is that quality can to a great extent be assured by the application of rigorous method. The implication for evaluation is that a simulation of actual, rather than idealised, knowledge will be evaluated more favourably by end-users since it will posess more potent deep (explanatory) knowledge.

Structure evaluation - information processing styles

Aspinall (1979) provides some evidence to support the intuition that the style of information processing evident in the ICAL system will be an important determinant of learning by students using the program. While this claim warrants further testing, the more immediate concern is to specify as far as possible the nature of the cognitive skills held by expert nurses when reasoning within this domain. When the stage has been reached of the ICAL system being able to sequentially ask for information about a patient then it becomes important to evaluate the 'routes' which the system takes routes taken by other decision models.

The whole area of measuring the quality of information processing in expert systems evaluation is neglected and deserving of better metrics and criteria. In one major article, for example, Ramsey et al (1986) carried out a comparative evaluation of three types of 'rule' - based on frame abduction, IF/THEN-rule induction and Bayes Theorem - yet employed the single criterion of classication accuracy as a measure. An approach with similar limitations has been adopted by Lewis and Hammer (1986) in testing the significance of rule-based models of human problem solving. Methods, therefore, will be required if this important aspect of the multi-dimensional evaluation is to succeed.

The implication for evaluation is that developments in the evaluation of information processing styles and routes will permit the validity of the ICAL model to be more clearly established.

Outcome evaluation

The point being argued here is that quality can be assured if multilevel criteria are set. It goes without saying that classification accuracy (the output) of the system is important. It is also important, however, that comparison of the output of the ICAL expert system is made against what Gaschnig et al (1983) term a 'gold standard'. Hence decision outcome should be measured in comparison to other accepted instruments with respect to a 'test set' of new patients. More importantly for the theme of this discussion, there must clearly be other criteria set for outcome measurement other than 'Turing testing'. One suggestion that can be made is evaluation through experimental testing of any predictions arising from the ICAL model.

Conclusion

The point which has been argued is that there is a greater likliehood of assuring overall quality if the incremental steps used in production are themselves reliable and valid. Billings (1984) poses three further questions which should be answered when determining the usefulness of CAL in nursing. These questions can be seen as bringing together the separate implications from the review of CAL and from the review of ICAL:

1. is the program consistent with nursing values? The ICAL of the present project should be examined to determine if it fits with the current ethos of valuing the cognitive component of the process of nursing.
2. does it meet the needs of the users? With pressure sores an extensive problem, nurses apparently requiring prediction scales, and frequent calls for improved prevention, the signs are hopeful that an evaluation of the present project will adequately meet this criteria.
3. what teaching/learning activities occur during CAL use? This most general of the questions returns the discussion to the difficult issue of evaluation of educational effectiveness. The thrust of the review, however, has been that there are many other dimensions to evaluation which can be more realistically answered - if the direction of these answers leans towards favourablity then the reliance on the single index of measured learning becomes much less important.

CONCLUSION

This review of literature has attempted several goals. One goal has been to look closely at traditional CAL in nursing in order to identify possible weaknesses in the innovation. Such weaknesses which were found might 'act as a warning' to a proposed ICAL system. Although the distinction between various types of CAL made generalisations difficult, nevertheless many programs were found to have an insecure theoretical base, an unclear role, and poorly established evaluation. It was concluded that the ICAL innovation might potentially resolve and avoid these problems.

With ICAL in nursing established at least as a promising idea, the next goal was to look as closely as possible at ICAL using the same parameters - types, theory, role, and evaluation. It was concluded that there was significant potential as an educational tool of an ICAL system which emulated the cognitive skills of expert nurses. In short, a cognitive model of nurses' patient assessment skills.

The greater complexity involved in all areas of ICAL - and cognitive modelling in particular - leads to an important conclusion for the present project with regard to evaluation. If a cognitive model as ICAL system is to maximise validity then each stage of the construction of that model must necesarily be designed and carried with the greatest rigour. The basis for a cognitive model as ICAL system is the knowledge-based component which actually performs the patient assessment; the student and teaching model come later.

Since accomplishing this component will prove to be a large undertaking in itself, it follows that the remainder of this thesis will be concerned with the cognitive model as goal. Using a step-by-step approach, therefore, it is intended to put into practice the conclusions from this review of the literature on the past, present, and future of computers in nursing education.

CHAPTER 2 KNOWLEDGE ACQUISITION

INTRODUCTION AND AIMS

The aim of this chapter is report the first stages of the construction of an ICAL system which is based on a cognitive model of expert nursing decision making. The chapter begins with the chosen domain of expertise; the factors influencing its choice and discussion of the nature of expert knowledge within this domain. The following parts report the implementation of a stepwise methodological approach to the acquisition of knowledge held by nurses of this domain.

At this stage the nature of the knowledge acquired is 'descriptive' rather than 'processing'. Hence, in the third part of the chapter, an experiment was designed which aimed to provide data on the manner in which nurses process descriptive knowledge when assessing simulated patients. Chapter 4 undertakes to explore this data in some depth, however in the final part of this chapter a preliminary analysis was conducted with the purpose of identifying expert performers in the patient assessment experiment.

NATURE OF PRESSURE SORE ASSESSMENT KNOWLEDGE

In this introductory section it is important to outline the reasons for the choice of the domain of pressure sore risk assessment and to discuss the conceptual basis of knowledge relating to the domain.

Advantages of Choosing Pressure Sore Risk Assessment for Study

There are clearly numerous knowledge bases held by clinical nurses which could have been selected for knowledge acquisiton and representation. In considering some possible domains, a set of criteria were developed which took into account both the constraints and needs of the project as well as the constraints and needs of the nursing profession. These criteria can now be used to structure this discussion on why pressure sore risk assessment was finally chosen. Criteria, then, can now be addressed under three broad headings relating to the domain of pressure sore risk assessment - its nature, ubiquity, and its importance.

Nature of pressure sore risk assessment knowledge

If a view, as unjust as it was superficial, were to be taken of nursing with regard to pressure sore prevention then perhaps the impression would be one of nurses carrying out largely manual and routinised tasks. An apparently 'mindless' approach to this aspect of nursing has been reported recently by J. Jones (1986). Hence daily life on the ward sees teams of nurses frequently attending to the 'pressure areas' of their bedfast patients before altering their position. Such a 'mindless' activity, if this view were correct, would seem little suited to the exercise of cognitive modelling on an intelligent tutoring machine.

Viewed more closely, however, at least some of these nurses are processing information in a fashion made less impressive by its seemingly automatic and subconscious mode. Pressure sore development is a process understood to varying degrees by all nurses - a score or so of factors which contribute to risk can be gleaned from the literature (eg Williams 1972). The cognitive processes for combining this information can be expected to be complex. In recognition of this complexity, a considerable research effort has gone into developing aids to risk judgement - certainly a more extensive effort than for any single other aspect of clinical nursing. Hence, Barratt (1987) is able to review eight of these essentially similar scales of risk factors each with a set of defining characteristics for categorising a patient. Typically, these scales involve addition of points for each 'danger sign' exhibited by a patient.

Such an extensive research effort into aids to prediction can be seen as an implicit criticism of nurses skills available. If the simple premise is taken that all sores are preventable and yet there still many sores (see below), it not surprising that Norton et al (1962) wrote of the widely-accepted link between incidence of sores and indifferent nursing. Yet apparently only Gould (1986) has begun to question the view that nursing should continue to strive for the 'ultimate' prediction scale. After some 25 years of scale-invention the point has been reached when the literature is more concerned with advocacy of a particular scale or with doubtful reliability and validity of scales (Goldstone and Goldstone 1982) than with the fundamental issue of achieving a real reduction in prevalence. A rare attempt to put into effect expert skills which were not being systematically utilised has been reported by Osborne (1987).

Understanding of the aetiology of pressure sore formation is currently well-developed (see Barton and Barton 1981) and basic to nursing curricula. Nursing, moreover, is increasingly questioning its practices and striving for quality. Yet, as Gould (1986) points out, there has been little evidence of the effect of knowledge on the essentially static pressure sore prevalence rates. Explanations seem to strongly implicate education issues. Factors which have been put forward recently include failure to implement classroom learning in the ward (Bendall 1975), misunderstanding of biological science (Wilson 1975), and even refusal to accept the existence of the problem (Kerr et al 1980). The conclusion offered by Gould (1986) is that there is an urgent need to link theory and practice in the ward situation - clearly a sentiment with which the present ICAL project would agree.

Florence Nightingale, surely an eminent cognitive psychologist, argued almost 100 years ago that good nursing must contain a strong cognitive component. 'Observation tells us the fact, reflection tells us the meaning of the fact... observation tells us how the patient is, reflection tells us what is to be done'. (p.255). Clearly also this pioneering nurse accepted that teaching the cognitive skills of nursing was a crucial task......'Training and experience are, of course, necessary to teach us how to observe, what to observe, how to think, what to think' (p.254). It seems remarkable that no specific effort has been made since these comments were made to definitively describe the information processing - 'observation and reflection' - of expert nurses with respect to pressure sore risk assessment with the goal of devising a directed teaching package aimed at passing on the cognitive component of this knowledge base.

Importance of pressure sore risk assessment knowledge

It is possibly not an exaggeration to point out that pressure sore prevention is an issue for every patient in General and Geriatric Hospitals. The incidence of sores which have actually developed is also high enough to ensure that nurses will be involved in preventive and tertiary care from their very first days of entering the wards - surveys have variously reported prevalence of 6.5% (David et al 1983) to 8.8% (Barbenel et al 1977) in UK hospitals. Not unsurprisingly in the light of both human suffering and of financial cost - £200m per annum - there are frequently calls for improved knowledge and teaching. (David et al 1983, Grier 1981).

Considering the relatively untried nature of some components of the methodological approach envisaged, it follows from these comments on the ubiquity of the problem that pressure sore risk assessment knowledge will be held to various degrees by a large pool of nurses. The search for expert subjects will therefore be made easier and the quality of data should be preserved despite the essentially exploratory methods to be used. It follows that a research effort in this area, particularly one in the spirit of Nightingale, may have the potential to make a contribution of some value.

Representativeness of pressure sore risk assessment knowledge as a nursing domain

Pressure sore risk judgement and preventive care planning are perhaps unique in that these skills are regarded as effectively the sole responsibility of ward-based nurses. Unlike other aspects of nursing where the nurse is but one component in a multidisiplinary care team, clinical expertise with regard to all aspects of pressure sores - from prevention of sores developing to management of actual sores - is acknowledged to be within the nursing province. Thus while the physiotherapist, nutritionist and doctor may play a consultative role, in terms of the countless NHS working hours devoted to thinking about and dealing with pressure areas and sores their contribution would be both minimal and secondary when compared to that of the nurse.

It follows, therefore, that nursing authority regarding this aspect of patient care is legitimate and complete. Consequently nursing knowledge possessed might almost be unique in that there will be virtually no 'blank spots' in the knowledge base where the nurse would concede only partial expertise "since that's the job of X". It follows also that nurses can be expected to support a research effort into devising a teaching package based on clinical nurses' skills in an area which is perceived to be the quintessence of nursing - an important point when it comes to seeking subjects' cooperation in demanding experimental exercises. Since Barratt (1987) has stressed that predictive aids are no substitute for professional judgement, it follows that the present study of these clinical skills can be distinguished from the research efforts cited above.

A Conceptual Model of Pressure Sore Risk Assessment Knowledge

What is domain expertise and how is it held ? Who are the holders of domain expertise and how can they be identified ? Such are the types of question which must properly be addressed before even beginning to ask how it is that the defined expertise held by the defined experts might best be elicited. One other factor must also be considered in this 'knowledge equation' - the eventual uses to which the elicited expertise will be put. An introduction to models of expert knowledge will firstly be offered prior to looking in more depth at the nature of expert knowledge on pressure sore risk assessment. It will be argued that benefit to the eventual ICAL system will result if careful attention is paid to these issues.

Nature of expert knowledge

It is perhaps making a broad yet defensible statement that the expert system field can be characterised by a rather ad-hoc 'prototyping' approach to system construction. Greater attention is paid to the goal of 'up-and-running' than to basic principles of knowledge acquisition. It is not unsurprising that since defining knowledge and knowledge holders is a stage which precedes even knowledge acquisition then it follows that this exercise will receive scant attention in the literature. Gotts (1984), in a rare example of an attempt to review work on establishing a typology of expert knowledge, found not only relatively few pertinent references but also relatively little coherence between parallel work.

With reference to types of expert medical knowledge (where most work has been carried out), what seems to emerge from the literature are two distinctions which are presrved despite the variety of terms used by different authors. The first distinction will be characterised here as contrasting *descriptive* (also known as factual or declarative) knowledge versus *processing* (also known as procedural or reasoning) knowledge. The distinction is between the 'facts' of knowledge and the 'processes' which are used to reason with these facts, something akin to ingredients and instructions in a recipe.

Buchanan et al (1983) saw "factual" knowledge as relating to objects in the domain while "strategic" knowledge refers more to problem-solving processes. Friedland (1981) coined the "declarative-procedural" distinction while Kolodner (1982) prefers a "domain" versus "reasoning" separation when refering to knowledge of disease states and the manner in which they are diagnosed. As might be expected, implicit in this distinction is the difficulty of eliciting processing knowledge when compared to descriptive. Processing knowledge, furthermore, is seen as something clinicians acquire experientially while descriptive knowledge is more like that found in the textbook.

The second major distinction which can be identified in the medical expert systems literature is between *deep* and *surface* knowledge. Other terms are relatively uncommon. Hence only Hartley (1981b) seems to prefer "systemic-additive" while Hart (1982) along with Szolovits and Long (1982) among others content themselves with "surface(or shallow)-deep". What is being referred to here is the difference between underlying principles of the domain (deep) and the mere empirical associations between phenomena (surface). Johnson-Laird's (1983) analogy seems apposite - turning on a television gives a picture; but the mental model of causality varies from child to repairman to physicist.

In the psychology of problem solving field there is clear support for a conceptualisation of knowledge which acknowledges both the descriptive-processing and surface-deep dimensions. Moreover, as Chi et al (1981) exemplify, there is better developed understanding of the importance of these dimensions in terms of cognition of experts and novices. These authors showed, for example, that a task of solving physics problems led to novices representing the problem in terms of its descriptive superficial features while the expert physicists built a mental model of the problem in terms both of its deep level principles and of the procedural knowledge necessary to solve that problem.

Applying these dimensions to theories of nursing knowledge is also fruitful. There has been, for example, considerable recent work on conceptual (deep level) models of nursing (eg Roper, Logan and Tierney 1985). Some attention, moreover, has recently been given to study of processing of knowledge by expert nurses (eg Gordon 1983). The field is nevertheless of an earlier standard of development than both the medical and general psychological literature. This thesis will argue later that it is of particular concern (given scant empirical support) that prescriptions of how nurses should represent their patients are increasingly stressing surface level features by advocating categorisation by nursing diagnosis, although some important work from Benner (1984) has called for a reversal of this trend.

If an expert system model is planned which aims to classify patients then it has become clear from this sketch of the nature of knowledge that descriptive <u>and</u> processing knowledge should be represented - particularly if an additional goal is education and training. However, if the system is planned also to be a cognitive model of expert nursing patient assessment then the issue of depth of mental representation of knowledge becomes equally important. It follows that these two dimensions of knowledge, applied to both experts and expertise, must therefore be considered in the light of intended use of the system. Moreover, Gammack and Young (1984) point out, the selection and application of the methods of acquiring knowledge from experts should be made with the domain taxonomy firmly in mind. It follows, therefore, that some consideration should now be given to a definition of 'expert' nurses.

Identification of expert practitioners

It might reasonably be expected that given at least some coverage in the literature on types of expert knowledge there would be corresponding attention paid within the expert system literature to choice of expert prior to the knowledge acquisition exercise. This, however, does not seem to be the case. The principles which guide choice of expert (or experts) seem governed more by circumstances and professional politics than by reasoned strategy.

The example of expert system construction in the medical domain elaborates these assertions. Wellbank (1983) advises finding an expert who is interested in the project and articulate about skills he or she possesses. Avoid those who are defensive when probed or those who feel threatened by the project ('avoid threatening' is a recurrent theme in the literature, advice to knowledge engineers is to stress that the system will "work alongside" and "not rival" experts). Clancey (1983) also underlines the need for cooperativeness given the necessary revision stages of system construction and advises 'rapid prototyping' as a means of helping secure this cooperation and interest. Only Szolovits and Long (1982) come near to considering different types of knowledge holder when discussing the advantages and problems of recruiting university and hospital doctors who between them might span the knowledge domain but whose professional politics may not be compatible.

Nearly always the assumption is of the expert being a singular entity, which given the difficulties outlined above is perhaps understandable. Hartley (1981b) addresses the possibility of elicitation from several experts and explains the inconsistencies which seem to result as being partly due to some being "experts" but others being "practitioners". Davis (1982) notes that the accepted wisdom of knowledge acquisition is to use a singular expert – a "knowledge Tsar" – and comments that as yet there are no good ways of dealing with inter-expert disagreement.

Gotts (1984) has suggested that the reliance on a single expert is more a reflection on the difficulties of coping with an uncertain domain in system construction than it is on the lack of ways of dealing with inter-expert inconsistencies. A counter-suggestion would be that it is the foregoing points about cooperation, defensiveness and politics which have brought about the reliance on a single expert while no serious attempt has been made to develop methods of producing 'average' expertise from a pool of experts. There is, however, an argument which alone can overcome all attempts to brush aside the foregoing points - *an ICAL system must in every respect achieve a high degree of validity*. It follows, therefore, that any model based on a single expert cannot potentially achieve the external validity that a model based on the 'collected wisdom' of several experts can potentially achieve.

The idea of an 'expert nurse' seems to have been only recently accepted within the nursing literature. The term, however, has been applied with varying degrees of stringency. For example, Broderick and Ammentorp (1979) simply denote a sample of associate degree nurses as experts while Corcoran (1986) demands that her sample of peer-nominated experts have previous publications. Benner (1984), in an application of the Dreyfus model of skill to nursing, sought to identify examples of expertise rather than examples of expert. These examples of expertise were then classified into one of five levels of competency pre-determined by the Dreyfus model.

Hartley's (1981) comment about 'experts' and 'practitioners' and the manner in which nursing research has defined individuals as expert seems to suggest a third dimension to add to the descriptive-processing and surface-deep distinctions. Clearly it is practitioners who are being sought for the present project, however Osiobe (1985) has made the point that knowledge can be of a formal or informal nature. Clinical nurses may therefore vary in terms of being 'wardwise' or 'bookwise'; have hands-on versus textbook knowledge. This third dimension - theoretical-practical - can therefore now be carried forward for a closer focus on the identification of potential expert holders of pressure sore risk assessment knowledge.

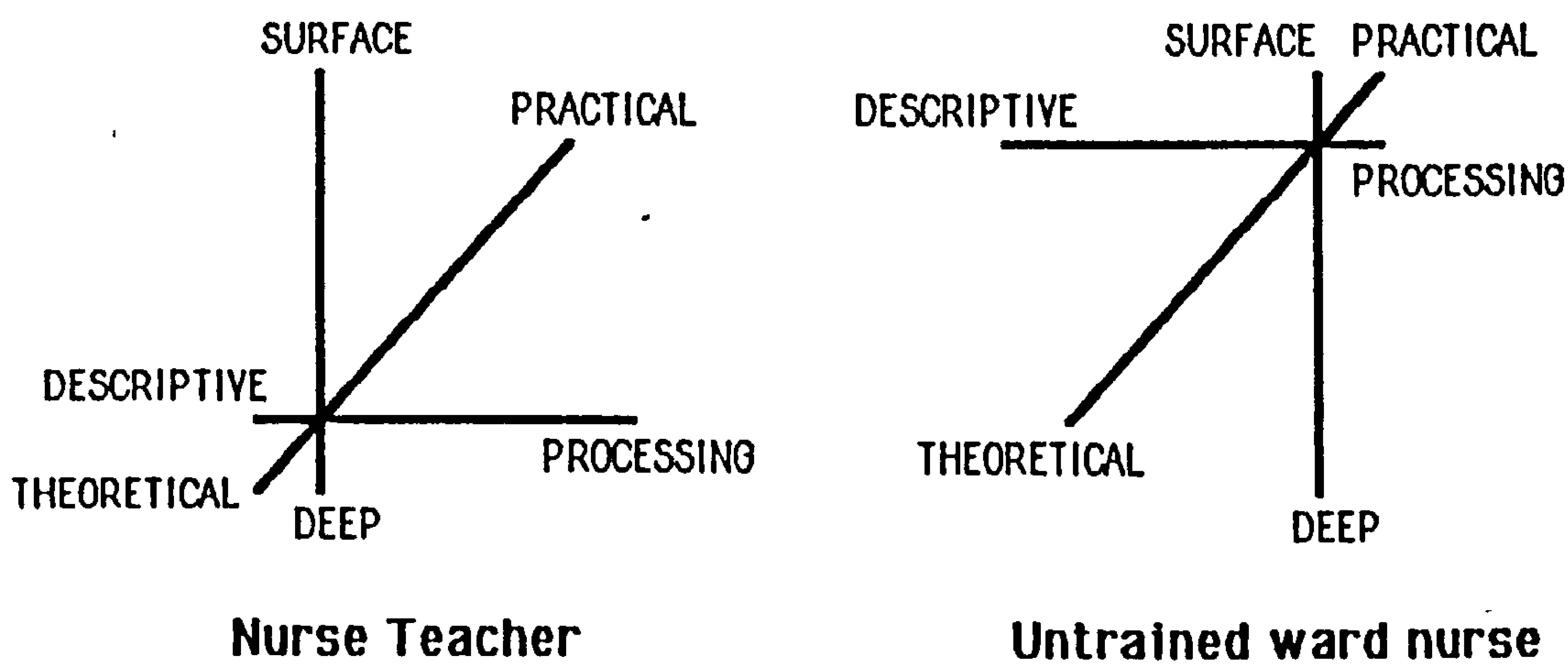
Identification of expert holders of pressure sore risk assessment knowledge

A model which conceptualises different dimensions of expert knowledge seems well suited to the pressure sore risk assessment knowledge. Any given nurse could be placed along three main dimensions of knowledge. These dimensions are:

descriptive - processing
deep - surface
theoretical - practical

The question of "who knows more about pressure sores, the nurse teacher or the untrained nurse of 20 years experience?" becomes rather facile when consideration is given to the differering dimensional profiles possessed by each nurse. As Figure 2.1 overleaf depicts, the nurse teacher can be considered to have deep knowledge of a descriptive theoretical nature while the experienced untrained nursing auxilary has a rather reversed profile.

Figure 2.1 Profiles of dimensions of expertise for two differing nurses



In these figures, for illustrative purposes, each nurse would be located at the intersection of the three dimensions of expertise.

Which dimensional profile should the proposed ICAL system seek to emulate? It is firstly important to define the intended uses to which the knowledge base will to be put and thereby receive guidance as to what constitutes 'proper' knowledge and who might possess such knowledge.

There are three principal considerations to the taken into account when looking to intended implementation of a pressure sore risk assessment ICAL system. Firstly, the embedded knowledge should be consultative. Secondly, the system should be interrogative, and thirdly, the system should be educational. In a general sense, the most important consideration is that the system should be process knowledge, that is it should be capable of emulating an expert nurse actually assessing a patient.

To an extent these goals overlap, hence a nurse consulting the system about a patient's pressure sore risk might be educated through a modelling process. Similarly, interrogation of the system's knowledge base might also be seen as consultative as well as educational. Nevertheless it is hoped that by preserving these distinctions between criteria and by repeatedly holding them up against the three dimensions of expertise there could be some resolution of the circular issues raised earlier about identification of domain expertise.

Considering the descriptive - processing dimension first, there is clearly little compromise with respect to this dimension in the quest for an ICAL system which adequately meets all three criteria above. It becomes clear that each pole of this dimension must be represented. For consultation, for example, expertise must be related strongly to past and present experience of decsribing and assessing patients at risk of pressure sores. The proposed system will not aim simply to arrive at a risk profile of patients (eg using the Norton et al 1962 scale) but will aim to emulate the cognitive processing of an expert nurse assessing a patient. Experience of previous patients, moreover, should ideally reflect not only numbers of patients but also variety of types of patient.

If the theoretical-practical dimension is considered, however, it becomes clear that validity of language descriptors is also important for each of the criteria. It has been argued that the proposed ICAL system should most appropriately be sited within the ward, where Benner (1984) and others have underlined the importance of establishing a consensus descriptive language. This implies that experts should currently be practising nurses.

A superficial view of the 'consensus descriptive language' which is used on the wards would be that it is so much jargon. This view would be to miss the crucial point of this exercise in modelling knowledge. Jargon, viewed charitably, can be a useful shorthand description which can be transmitted economically among understanders of jargon. Leaving aside the well-aired criticisms of jargon, what is taking place between these understanders of jargon is the mutual use of the same mental model. Thus if one nurse tells another that Patient X "is cachexic" then each will understand the many attributes and implications of this statement.

It is clearly desirable that learner nurses acquire mental representations of these models of patients. It therefore follows that the system requirements of interrogation and education will be more easily met if the local shared meaning structures are embedded within the system's knowledge base. For example, ask the nurse above what she means by cachexic and the stuttering reply might indicate a mental model acquired in a rather top-down fashion rather than one built up from first principles and which could be expected to degrade gracefully into explanations based on these first principles. Furthermore, it seems a reasonable hypothesis that the 'best' expertise will be unlikely to be found in neatly-labelled mental models understood by all - the fine discriminations between susceptible skin types held by an experienced ward nurse may be an example of this point.

Mention of 'first principles' serves to introduce the final and possibly the most crucial dimension - surface and deep knowledge. To date, research effort with regard to improving judgement of pressure sore risk has concentrated rather on surface level knowledge. For example, efforts have been made to make nurses more aware of the 'danger signs' which they should notice (Barratt 1987). It seems surprising that study has been neglected until recently of the deeper levels of knowledge which might be held by demonstrably excellent practitioners. Braden and Bergstrom (1987) have suggested one conceptual schema for this deep knowledge, the rationale being that it would be educationally useful to gain an understanding of this knowledge. However, it becomes clear that behind such knowledge there will be an interaction between experience and educational preparation.

The question of whether deep knowledge should be sought from either wardwise or bookwise nurses is to a degree solved by an educational structure which requires that all teachers of nurses have completed at least basic nurse education followed by a period of ward work. It might therefore seem reasonable that nurse teachers who have studied further might most appropriately be identified as holders of expertise. Certainly this contention would seem to be sensible when faced with the opposite extreme - an untrained nursing auxilary of 20 years experience - when the educational requirement of the intended system is considered. Nevertheless, the system must also be consultative, by which is meant that an adequate assessment could be made of any previously unseen patient with any combination of attributes. For the present purpose, then, the definition of expertise should tend more toward deep knowledge derived from experiential learning.

Conclusions

Before summarising the implications of these arguments it is perhaps worthwhile looking in more detail at a factor which has recurred frequently in the foregoing paragraphs - length of experience. The largely North American nursing cognition literature has tended in the direction by defining highly educated nurses as expert (eg Broderick and Ammnetorp 1979). The implicit assumption in the present discussion, however, has rather been to equate longer experience in a nurse working with pressure sores with greater quality of knowledge base. As Benner (1984) points out, however, this may not be the case in that length of experience might more properly be equated with a rather 'mindless' and automatic style of applying fixed ideas to pressure sore prevention which has little regard for current thinking on the subject.

A conclusion on the length of experience issue is, however, far from straightforward. The foregoing has argued for clinical nurses with extensive and varied experience to be regarded as expert. There are nevertheless dangers in this approach.

What is necessary, therefore, is a reasonable set of conclusions which can act as 'points to consider' when designing the methodological approach for eliciting the pressure sore assessment knowledge base:

1. Complete expertise should not be taken as being possessed by any single nurse.
2. Variations in the effect of clinical conditions on pressure sore risk points to no single nursing area as holding expertise in all nursing areas.
3. Depth of knowledge is important yet should not be demanded of knowledge holders possessing exclusively theoretical or practical skills. Ideally, nurses with balanced practical <u>and</u> theoretical skills should be sought.
4. At all times there should be procedures devised and applied which will seek to establish consensus expertise and identify those individuals who deviate from the consensus.

With these points in mind, the discussion can move to a report on the design and implementation of the methodological steps used to elicit the descriptive knowledge base.

KNOWLEDGE ELICITATION EXERCISES

This Part will comprise three sections which between them report and discuss the preliminary exercises carried out to elicit the expert knowledge base necessary for the proposed ICAL system. The maxim which has guided this incremental approach was set out in the final Part of Chapter 1 ...... that quality of a system in an overall sense can be assured through the use of reliable and valid methods during all stages of system construction.

Eliciting Top-level Descriptive Knowledge: the Attributes

Introduction

Following the convention widely adopted (eg Hart 1986), it is convenient to distinguish between 'attributes' and 'values' which an attribute can take on. Hence for any given person the attribute Sex would take the value 'male' or female'. The other term for attribute in the context of pressure sore risk is 'factor' - the task in this section is to specify the range of attributes (or factors) which nurses believe should be assessed with regard to pressure sore risk.

Perhaps because it seems rather obvious which attributes are of interest to researchers, there is not commonly much attention given to eliciting attributes from knowledge holders. Broderick and Ammentorp (1979), for example, give no details of the source of 59 attributes which they used in a simulated patient assessment exercise. Hammond (1966), on the other hand, generated 165 pain cues from first principles using the critical incident technique. The point is that there is a threat to validity within studies of reasoning which have not firstly established the basic components used when reasoning. It is safer to begin by eliciting attributes which are actually used rather than by making assumptions based on what the textbooks say should be used.

It is possible to elicit the broad range of attributes which nurses use through the use of a free-listing task in response to a question such as "what factors would assess when ....". Clearly it would be of benefit if measures could be derived to indicate the degree of confidence held in the lists which would be provided by this task. In fact there are some assumptions which could be made regarding these lists, moreover, these assumptions can be tested. Hence it could be argued from the work of Tversky and Kahneman (1974) that the attributes which appear early in such lists will be of greater significance than those which are recalled in the last positions. Similarly, the frequency with which any given attribute is listed (ie the number of nurses who mention this attribute) might denote importance.

Jaccard and Sheng (1984) provide a suitable index of attribute importance which addresses these points. Hence an index can be computed (see results section below) which assumes not only that frequency of mentions is important but also it is important to take into account the position in a particular list and the number of attributes mentioned in that list. It was therefore determined that the free-listing task is an economical method which is open to testing for reliability of results.

Method

Some attention to the need to establish validity was required prior to putting into operation the free-listing exercise. The most important point to be resolved concerned the nature of the subjects since there are in fact two bases suggested to the compilation of a set of attributes concerning the liklihood of a patient developing pressure sores. The first basis concerns the 'predictive' factors of which a nurse might have theoretical knowledge. The second basis reflects more directly on the 'practical' decision making task itself and would reflect the 'assessment' factors which a nurse actually uses when judging the risk of patient developing pressure sores. It can be seen that the distinction involves the theoretical-practical dimension introduced earlier.

There may of course be no difference in the lists of factors which each set of nurses might list, however the point is that assumptions cannot be made regarding the goal of decison making. Hence a 'theoretical' nurse might refer to an internal list of factors established by research as being important in the aetiology of pressure sores. A 'practical' nurse, on the other hand, might interpret the question about factors affecting liklihood of pressure sore development in terms of the factors which the preventative care which she plans for the patient.

In order to gauge replication of findings and to produce comment on the 'predictive' or 'assessment' issue, it was decided to undertake two free-listing exercises which focused separately on each suggested basis to a list of attributes. A minimal assumption about expertise was made at this stage in that candidate subjects were defined as registered and having had some experience in either medical or surgical (including orthopaedic) wards. The goal, which was achieved, was to recruit approximately equal numbers of subjects who might be characterised as either 'surgical' or as 'medical'.

The details of each sample are:

Sample 1 (more 'theoretical' nurses) An opportunity arose to meet a group of N=32 candidate subjects who were Clinical Nurse Teachers while they attended a study day. Agreement was obtained to 'help with some research into experienced nurses and the nursing process'. Sheets of paper which were blank except for the instruction were distributed (see Appendix 1). A full 15 minutes was given for the task, although all subjects indicated that they had finished before this time.

Sample 2 (more 'practical' nurses) These nurses, by definition, were at work in wards within Glasgow Hospitals when approached by the experimenter with the same request as for Sample 1. To allow subjects to find the opportunity to complete the task, forms were collected from subjects several hours later during that same shift. It was anticipated that this strategy might result in high attrition, therefore some 60 subjects were approached. A high response, however, was achieved with N=52 subjects complying.

Results

'Entries' written by subjects on the forms proved to be unambiguous to categorise, the norm being that exactly the same word was used by differnet subjects (ég MOBILITY). However, since the experimenter carried out the categorisation it was necessary to demonstrate reliability of this classification. A random sample was assembled of 30 entries which potentially were ambiguous to categorise (ie differneces in wording). These entries were given to 2 experienced nurses along with a list of categories into which each could be assigned. Later comparison revealed that one rater agreed with the experimenter on all occasions, while the other agreed on 28 occasions. The benefit of 'clean' data within this domain was predicted in Part 1 - clearly there is significant shared meaning with respect to top-level descriptors.

The 32 subjects in Sample 1 wrote a total of 246 entries which fell into 23 categories of risk factor (mean number of entries per nurse=7.7). The 52 subjects in Sample 2 gave a total of 320 entries which fell into 19 categories (mean entries per nurse=6.15). There is no interest in apparent differences since the experimental situations were rather different. Following the suggestions made in the introduction, quantification of attribute categories was undertaken using the Frequency measure (number of subjects mentioning this category) and the Weight measure (after Jaccard and Sheng 1984).

The Weight (or 'importance') of a category was computed by:

$$\text{Weight (w)} = \left(\Sigma\, O_{ij} / P_i\right) / C_k$$

where O_{ij} is rank order in reverse of an entry j made by an individual i; P_i is the number of entries made by individual i; and C_k is total number of mentions made by all subjects of this category k.

Table 2.1 below displays the 14 factors which appeared in each sample's lists along with the frequency of mention (F) of each category and the cumulative weight (W). The factors which were mentioned by one or more individuals in only one of the samples were DEFORMITIES, PERSONAL HYGIENE, AGILITY, DEHYDRATION, RADIOTHERAPY, INFECTION, SENSORY LOSS, ANAEMIA, PYREXIA and SMOKING. Since none of these factors received more than 3 mentions, they were henceforth considered no more. The point to be made is that it is not whether or not a factor can affect pressure sore risk but rather whether or not that factor is one which been shown to be within nurses' knowledge bases.

Table 2.1 Frequency and Weight values of 15 factors common to each sample

	FREQUENCY		WEIGHT	
factor	S1	S2	S1	S2
INCONTINENCE	45	32	.57	.57
MOBILITY	45	32	.80	.79
NUTRITION	42	28	.63	.67
BUILD	36	30	.67	.66
AGE	19	24	.60	.59
DIAGNOSIS -	17	17	.51	.38
MENTAL STATE	16	16	.30	.35
CIRCULATION	15	10	.48	.51
NURSING STANDARD	20	4	.46	.49
SKINTYPE	8	9	.50	.55
LIFTING & TURNING	13	4	.46	.45
SEX	3	9	.29	.50
DRUG THERAPY	5	4	.25	.23
BLOOD PRESSURE	2	3	.27	.59

Some factors require some explanation, for example NURSING STANDARD is the term given to the entries which directly implicated poor nursing care in pressure sore aetiology. LIFTING & TURNING could also have been termed 'Mobilising - dependency' in that entries here conveyed that the extent to which a patient was dependent for positional relief would affect pressure sore risk (references to shearing force were categorised under NURSING STANDARD).

It is clear that there is close corespondence between these two lists. Statistically this can be demonstrated by Kendall's rānk order coefficient; for Frequency measures K=.79 ($p<.0001$) and for Weights K=.70 ($p<.0001$). Nevertheless, the correspondence is not perfect and contains some interesting anomalies. For example, 8% of the entries given by 'theoretical' subjects referred to poor standard of nursing compared with 3% in similar category for 'practical' subjects. However, it is not intended to speculate here on apparent diferences of emphasis since this will be undertaken within the major analyses which lead to the construction of the cognitive model in Chapter 3. The proper conclusion, for the purposes at hand, is that two measures have agreed that the 14 factors above account for the overwhelming number of those listed by nurses in response to the question.

Further testing - Durbin design

A final measure is required of the degree of confidence which can be placed in a conclusion that nurses concur about factor importance. With this in mind, an additional experiment was designed and administered to the Sample 1 subjects in order to test the hypothesis that the numerical operations carried out on the lists could be taken as valid. If it is the case that there is an underlying hierarchical structure to subjects' mental representations of the factors then there should be demonstrable concordance between subjects. To be more specific, if asked to judge whether Factor A is more important than Factor B then the nurses should agree in terms of direction of importance (eg A is more important than B) at a level beyond that expected by chance.

Five factors were selected for a multiple pairwise comparison test known as a Durbin design (described by Marascuilio and McSweeney 1977). The four principal patient attributes were selected (MOBILITY, INCONTINENCE, BUILD, and NUTRITION) with DIAGNOSIS included as an important factor but nevertheless not patient-state specific. Each subject was asked to consider the 10 possible pairings of these factors individually and to decide which factor of a pair seemed to them to be more important in terms of pressure sore risk assessment (the 10 pairings along with some additional results information is given in Table 2.2 below)

Results of the calculation of the coefficient of concordance showed that the 30 subjects who completed the task had a highly significant level of agreement on direction of importance ($W=.24$, $x^2=29.1$, $p<.001$). The rank order of importance which could be recovered from the procedure demonstrated that MOBILITY was most important, followed by INCONTINENCE, DIAGNOSIS, BUILD, and finally NUTRITION as the least important of these five factors.

Further operations were carried out using the Friedman's 2 way ANOVA procedure described by Marascuilio and McSweeney (1977) in order to test whether the reliability of ranking of individual pairings was significant. Since a significant main effect of Ranking X Factor was suggested by the result (x^2=28.05, p<.001), post-hoc testing of the reliability of each pairing was undertaken. Results for the 10 pairings are given in Table 2.2 below.

Table 2.2 Reliability of post-hoc comparisons of 10 pairings

pairing	p
MOBILITY versus INCONTINENCE	<.05
" v. DIAGNOSIS	<.05
" v. BUILD	<.05
" v. NUTRITION	<.01
INCONTINENCE v. DIAGNOSIS	<.05
" v. BUILD	<.05
" v. NUTRITION	<.01
DIAGNOSIS v. BUILD	ns
" v. NUTRITION	ns
BUILD v. NUTRITION	ns

Given that these 5 factors were 'neighbours' in the lists, it is impressive that subjects concurred sufficiently strongly for there to be reliable differences on 7 out of 10 of the pairings. Moreover, given the close correspondence between the Durbin results of prefered ranking order and the list itself there seems to be grounds for confidence both in the listing procedure and in the subsequent measures derived of factor importance.

Following one final operation, it is therefore proposed to take the factors identified through to the next stage of knowledge elicitation. The factors which are required for the proposed simulated patient assessment exercise are factors which are both nursing-specific and patient-specific, by which is meant that specialist knowledge of medical diagnosis or connative beliefs about standards of nursing should be avoided in order to maximise standardisation and reduce complexity. It is therefore proposed to exclude DIAGNOSIS, DRUG THERAPY, and STANDARD OF NURSING from the list. One further operation will be carried out to INCONTINENCE in that it will be restored to the separate attributes URINARY INCONTINENCE and FAECAL INCONTINENCE which some but not all subjects specified.

Eliciting Micro-level Descriptive Knowledge: the Attribute Values

Introduction

A patient cannot be described as, for example, SKINTYPE. To achieve a full description it is necessary to have sub-classifications or 'values' of attributes such as 'type of skin B', '4.5', or even 'unknown'. The focus of this phase of the knowledge elicitation exercise, then, is to specify the values which each of the 12 target attributes can take on in order that patient descriptions are both meaningful to nurses and discriminable from other patient descriptions. The goal becomes one of specifying the micro-structure of nurses' representations of their patients.

Embarking on a 'scaling' exercise such as this raises several issues such as length of scale, nature of scale anchor points (adjectives, numbers, descriptions, icons?), and the intended purpose of the scale. More equally familiar issues are raised when these questions are addressed. For example, scale length trades off reliability and discriminability - a two-point scale will have strong inter-rater agreement but may not discriminate between patients. This, moreover, relates to the training goal of the project in that an expert nurse might agree that skin is either 'susceptible to breakdown' or 'ok'. The learner nurse, on the other hand, might not possess the deeper knowledge which goes with understanding these descriptions.

Although construction of nursing taxonomies of patients has been receiving increasing attention, it can however be argued that most attention has been given to the issue of 'intended purpose'. For example, the authors of conceptual models for nursing (eg Roper, Logan and Tierney 1985) are perhaps most interested in educating learner nurses to systematically organise their thinking around certain groupings of attributes which fit the concepts within the model (such as activities of living).

Researchers who aim to construct scales of attributes (eg Norton et al 1962) or who set out to build mathematical models of the nursing process (eg Grier 1981) are more concerned with specification of the values which attributes may take on. These researchers, however, are primarily interested in numerical values of attributes - the qualitative descriptions which accompany the numbers are often fairly brief and ambiguous. The need to scale attributes according to statistical properties (eg vpoor, poor, average, good, vgood) rather ignores whether these qualitative descriptions are in any way representative of the symbols used by humans when categorising a patient on their 'internal' attribute scales.

Firstly, the present project is not modelling quantitatively therefore it can aim for qualitative descriptions of attribute values. Secondly, since 'risk classification' by the intended system is less of a priority than educational goals, these qualitative descriptions should be in the consensus natural language which existing senior nurses use to represent their patients. Thirdly, the system is planned as a model of clinical rather than theoretical cognition. It follows, therefore, that sub-classifications of each attribute scale should be constructed solely through study of senior nurses who are currently working with appropriate patients.

One final reason for this concern with establishing valid surface knowledge is that the study of deeper level processing knowledge (the most crucial phase of the project) will be greatly facilitated if the simulated patients which nurses assess are stated in terms which are unambiguous in their shared meaning. Rather than suppose that national descriptive 'norms' exist, this implies that the knowledge base should be localised - a point also made by Ball and Hannah (1984). A methodological approach must therefore be designed which incorporates these various rationale and goals.

Method

The core of the approach to eliciting attribute values was to use interviews focused on patients with a sufficient number of nurses. 'Sufficient' takes on twin meanings. Firstly, along with goal of consensus language, there should be sufficient nurses in order that shared rather than idiosyncratic language could be tapped. Secondly, it is necessary to interview sufficient nurses in order to represent the range of patients which could be encountered. In essence, therefore, a nurse would be asked to give a description of one of her own patients with respect to each of the 12 attributes. This exercise would be repeated with different nurse and different patient until the point was reached when no new attribute values were emerging.

Subjects who were approached with a request to cooperate in this exercise were defined, following the above discussion, in terms of length of experience, localisation, and expertise. Length of experience included both experience in assessing patients at risk of developing sores and experience in communicating patient descriptions to learners; hence the target sample was set at the clinical ward sister. The need for localisation led to targetting of a sample within two central Glasgow teaching hospitals which shared a college of nursing.

Rigorous assessment of expertise is not yet of crucial importance (for eliciting vocabulary); therefore a self-rating task was used as a simple screen - lest a nurse revealed marked lack of confidence in evaluation of her own expertise in the assessment of pressure sore risk. The scale was drawn up to depict 5 categories which followed from the question "If asked to compare your skill in assessing pressure sore risk to that of other qualified nurses, would you estimate that it was ....": considerably below; below; about the same; greater than; much greater than. Only nurses endorsing one of the last three categories were accepted.

Since ward sisters have little available time, it was planned to interview each subject with respect to one patient only. The number of subjects approached was determined by the range of patients targeted. In order to ensure 'spread' of attribute value descriptions, a stratified range of patients was planned from surgical, orthopaedic, and medical wards - some patients with a pressure sore patient, some of a high and some of low risk of developing a sore, and some who were randomly selected. The original intention was to have 10 patients in each of these four categories; this number was reduced during the data collection exercise when it became clear that there was little variation of attribute values which applied to patients in the low risk group.

Procedure adopted was to approach a targeted ward sister with a request for 10 minutes time to help put together a new type of pressure sore risk assessment teaching tool which would be based on the knowledge held by experienced nurses. A category of patient was selected according to the need to complete the categories in Table 2.3 overleaf; the subject was then asked to visualise such as patient if one was present in the ward currently. The interviewer then followed a sequence of asking the question "how would you briefly describe this patient's ....(attribute)....?" and noting down verbatim the answer. Each of the 12 attributes were treated in this way. Finally, the self-rating scale was handed to the subject with a request for the nurse to place herself.

Results

At the end of the data collection exercise there had been 34 patient-focused interviews obtained from 30 subjects (four subjects dealt with two patients). The types of patients and the wards in which they were inpatients are given in Table 2.3 overleaf.

Table 2.3 Numbers of patient-focused interviews by type of patient and ward

	TYPE OF PATIENT				total
	sore present	high risk	low risk	randomly selected	
surgical	4	3	3	4	14
orthopaedic	3	3	1	2	9
medical	3	4	2	2	11
total	10	10	6	8	34

Final analysis of the descriptions which resulted for each attribute was carried out by the experimenter in order to effect maximum standardisation and consensus from the replies, although it was clear that there was a strong consensus of descriptive terms used by the subjects within these two hospitals (each with a common College of Nursing). In addition to other potential weaknessess, the attributes and values which have finally been set down (see Appendix 2) must be seen in the light of this deliberate policy of achieving 'local' validity.

Establishing a database of patients

With top and micro-level descriptive knowledge collected, the task now was to build up a large cohort of patients who are described in terms of the 12 attributes and values and who have been evaluated in terms of risk of developing sores by the nurses who are caring for them. A checklist questionnaire was designed which set out the 12 attributes arranged under activities of living headings (see Appendix 2). Once again, the expertise of the nurses who would complete this questionnaire was not regarded as crucial, and for this reason it was decided to minimise risk of judgemental errors by permitting only three categories of decision - High, Medium, and Low risk of developing sores.

The collection of the database of patients (henceforth refered to as database1) was carried out by the experimenter re-visiting six weeks later the wards of the 30 ward sisters with the checklist forms. If the same nurse was on duty then she was asked to to complete the one form for each of 10 of her patients. To ensure an even spread of patients representing varying pressure sore risk, the nurse was asked to complete a form for every second patient on the ward's sleeping list. 154 properly completed forms were returned (71 Low risk, 44 Medium risk, and 39 High risk).

One of the several uses to which this database is put will be discussed in the next section. At a later point in the project there was a requirement for an additional database of patients - see introduction to Chapter 3 for details.

Eliciting Information Processing Behaviour

Introduction

Descriptive level knowledge used by nurses during the cognitive operation of assessing a patient's risk of developing pressure sores has been assembled. To the extent that this knowledge is valid, it can be taken as corresponding to the symbols which are used by the nurse to represent the patient she is assessing. The aim, however, is to emulate the *active processing* of these symbols by expert nurses. The medium for achieving this cognitive model, it is planned, is the computer. It can be seen, therefore, that the ambition of this project has acted to ensure that there is both descriptive and processing knowledge of sufficient detail for it to be encoded into a computer.

This point about the chosen medium is of crucial importance. If that medium had been the blackboard then boxes could have been drawn to represent the resultant cognitive model. The computer, however, requires precisely specified instruction code. Although this in turn helps add a certain measure of rigour to the theoretical basis of the model, the practical consequence for the moment is that the methodology used must be adequate to the task of preparing this code. It follows that this imperative and the imperatives of reliability and validity for educational goals must at all times be considered when choosing from the literature a suitable method for achieving the more difficult goal of eliciting the processing knowledge. Four principal approaches or methods to analysing processing knowledge can be considered - phenomenological, statistical modelling, verbal protocol analysis, and process tracing.

The phenomenological perspective has, as Tanner (1988) points out, multiple perspectives but nevetheless some common assumptions. With regard to study of information processing, however, it quickly becomes clear that the the present project does not share these assumptions. One point, put strongly by Benner (1984), is the belief that formal specification of clinical judgement cannot be achieved if removed from the context in which action takes place. A more rationalist perspective, which the present project adopts, would reply that it does not follow that these same decision makers cannot make decisions on reduced information. The point, however, is that the observational and retrospective interviewing methods which are used by the phenomenologists are ill-suited for the present purpose. Notwithstanding the issue of reliability (Nisbett and Wilson 1977), the data which results is of insufficient specificity.

Statistical modelling, secondly, has a strong tradition in medical and, more recently, in nursing research on decision making. This work will be reviewed in both Chapters 3 and 4; for now however it can be seen that some aspects of methods used to model decision making might be of interest. Hence Hammond (1980) shows that the typical paradigm is to present a series of cases to subjects for rating on, for example, degree of risk. The aim of the research might be foreign to the present project (ie finding a mathematical formula which fits what the humans seem to be doing), moreover the validity of the descriptive knowledge is open to question (Elstein et al 1983). Nevertheless the principle that reliability can be strengthened through repeated measurement is important for the present search for suitable methodology.

Verbal protocol analysis, thirdly, is possibly the largest category of methodological approach to have been employed both in medical and nursing decision making research. Hence the seminal medical work by Elstein et al (1978) and some recent nursing studies (eg Tanner 1983, Corcoran 1986) have provided some evidence that expert practitioners use the hypothetico-deductive method (see Chapters 3 and 4 for further discussion). Analysis of transcripts taken from subjects who are instructed to 'think aloud' can provide data sufficiently rich to construct computer-based cognitive models (Ericsson and Simon 1983). Moreover, as Elstein et al (1983) point out, the richness of such data is educationally attractive.

Aside from the well-aired dispute about the validity of cognition which is verbalised (see Ericsson and Simon 1983 for overview), there are three problems connected with protocol analysis methodology. Firstly, Lichtenstein (1982) has made the point that as experts' cognition becomes more automatic then the verbalisation from experts may reflect little more than the way these subjects as novices would have gone about solving the problem. Secondly, a point made by Patel and Groen (1986) is that this methodology becomes less applicable in verbally complex situations which depend on a rich knowledge base (in contrast to the 'toy' problems successfully studied using protocol analysis). Thirdly, reliability and validity are jeopordised since the huge volume of data produced by the method acts to ensure that very few subjects and possibly a single patient are analysed.

Notwithstanding the strength of these criticisms of protocol analysis, a further crucial point which can be made is that there have been no medical or nursing research reported which has utilised this method to furnish data of sufficient detail to construct an operational cognitive model. Although the method has been used to this end in other fields (eg Anderson 1983), it is not helpful that there has been no precedent, particularly from nursing.

Process tracing methodology, lastly, seems to avoid the main criticisms made of the other methods above. Moreover, it has been used to effect in studies of nursing cognition by Gordon

(1980). The paradigm, which has been developed principally by Payne (1976), acknowledges the role of subjects' concurrent verbalisations while solving a task but goes considerably further in measurement of the processing of information in predecisional behaviour. This is achieved mainly through a procedure which ensures that monitoring of information use by the subject can be carried out reliably. Hence, in a typical experiment reported by Payne (1976), an 'information board' was set up which displayed envelopes labelled with attribute names. The subject's task was to 'search' through the information as they wished by opening envelopes in order to read the attribute value contained within. Interestingly, subjects were also asked to 'think out loud' while performing this task.

An even more extreme emphasis on process rather than product of decision making is the 'eye fixation' work of Russo (1978) where a record of the sequence of a subject's eye fixations as they examine attributes which are displayed. Nevertheless, there are understandable technical problems associated with this technique such as the limitation of a maximum of 10 attributes which can be displayed (Russo 1978). The point, however, is that these investigators feel it worthwhile to collect data which is 'behavioural' and trustworthy. Although Jacoby (1977) argues strongly that such data is clearly preferable to verbal protocol data, more recent evidence supplied by Ericsson and Simon (1983) shows that certain types of verbal reports are capable of providing a picture of working memory which is both reliable and illuminating. The verbal protocol method, on the other hand, limits the amount of subjects and/or problem situations which can be studied.

The position taken by Payne, Braunstein and Carroll (1978) is that there are clear benefits of using a concurrent multimethod approach which incorporates both information acquisition and verbal report data. Hence these workers found that ambiguities arising from one source could often be made more clear when the concurrent data from the other source was inspected. In the light of the criticisms of verbal comments given retrospectively (Nisbett and Wilson 1977) or given in response to specific questions (Ericsson and Simon 1983), Payne (1976) has shown that very useful verbal protocol data can be provided by focusing more on the information search task while asking subjects (without specific training) to 'think aloud' regarding anything which comes to them.

It is suggested, therefore, that the present project could set up a simulated patient assessment exercise during which data is collected from subjects as they searched the available attribute values. Simultaneously, a record could be made of 'simple' verbalisations. It is hoped that the present experimental design will improve on some of the previous research using this paradigm. Specifically, it is hoped that using a computer both to present patients to subjects and to record covertly their responses will obviate possible effects of having an experimenter present. Secondly, it is intended to run the experiment with a greater number of both subjects and patients to be assessed. Thirdly, there will not be a complete reliance on sequential information search in that patients will on occasion be presented with all information simultaneously available.

Objectives of the simulated patient assessment experiment

1. To provide data corresponding to processing knowledge and hence complete the knowledge elicitation phase of the project.
2. To develop and apply methods which will identify subjects whose performance of the pressure sore risk assessment task can be taken as expert.
3. To carry forward this expert knowledge to a more rigorous analysis from which a cognitive model of human expertise can be constructed.

Method

Design of this experiment involved same subjects carrying out pressure sore risk assessments of simulated patients presented in two experimental conditions. All patients appeared in each condition, therefore each patient was assessed twice by the subjects. These conditions were:

SELECT trials - only attributes of a patient were displayed; subjects were required to elicit values of attributes of their choosing prior to making a risk decision.

ALLUP trials - all attributes were displayed along with their values; subjects were required only to make a decision.

Patients were presented in blocks of 3 in each condition. To achieve between-subject comparison all subjects were presented with patients for assessment in fixed order, although these patients were counterbalanced for degree of risk of pressure sores (see Stimuli below for more details).

Subjects recruited to the experiment were nurses working or teaching in the clinical area of Glasgow hospitals. Since a preliminary aim of the experiment was to identify 'expert' subjects from 'potential experts', selection of subjects was deliberately stratified in order to represent different specialities, lengths of post-registration experience, qualifications, and roles. The 14 subjects who completed the experiment are listed in Table 2.4 overleaf.

Table 2.4 Descriptive Data on Subject Nurses.

nurse	qualifications.	work area	grade	experience
1	RGN RMN	acute med	Sister	12 years
2	RGN ONC Dip CT	ortho, surg	Clin. Teacher	16
3	RGN ONC RNT	ortho	Tutor	30
4	RGN DN	med	Sister	11
5	RGN DN	geriatrics	Nursing Officer	13
6	RGN ONC	ortho	Sister	15
7	RGN DN	surg	Sister	19
8	RGN Oncol Cert	med	Sister	6
9	BA RGN	med	Staff Nurse	4
10	BA RGN	surg	Staff Nurse	4
11	RGN	med	Staff Nurse	7
12	RGN ITU Cert	surg	Sister	5
13	RGN	med	Staff Nurse	3
14	RGN	med	Staff Nurse	3
			range	27 (3-30)
			median	9 years
			mean	10.57
			sd	7.7

Stimuli comprised 18 patients described in terms of the 12 attributes (eg SKINTYPE) and values (eg 'rather dry and thin'). These 18 patients were selected from database1 to be representative, as far as possible, of the whole sample and of the judgements of the nurses who had cared for them. Hence Patients 1 to 6 were High risk, Patients 7 to 12 were Medium risk, and Patients 13 to 18 were Low risk. In addition, an effort was made to ensure selection of patients representing each value of the 'important' attributes such as MOBILITY and MENTAL STATE. The display of attributes to subjects was in the form of 3 rows of 4 attributes. This display will be discussed and depicted under Apparatus below; at this point it should be noted that the position of attributes within the display was varied from trial to trial.

Counterbalancing of patients by risk was carried out with two principles in mind. Firstly, patients should not appear in clusters of, for example, High risk types. Secondly, since it was planned to present the same patients in each condition, each presentation should be well-separated within the overall sequence of 36 trials. The order of presentation of Patients 1 to 18 is given Table 2.5 overleaf.

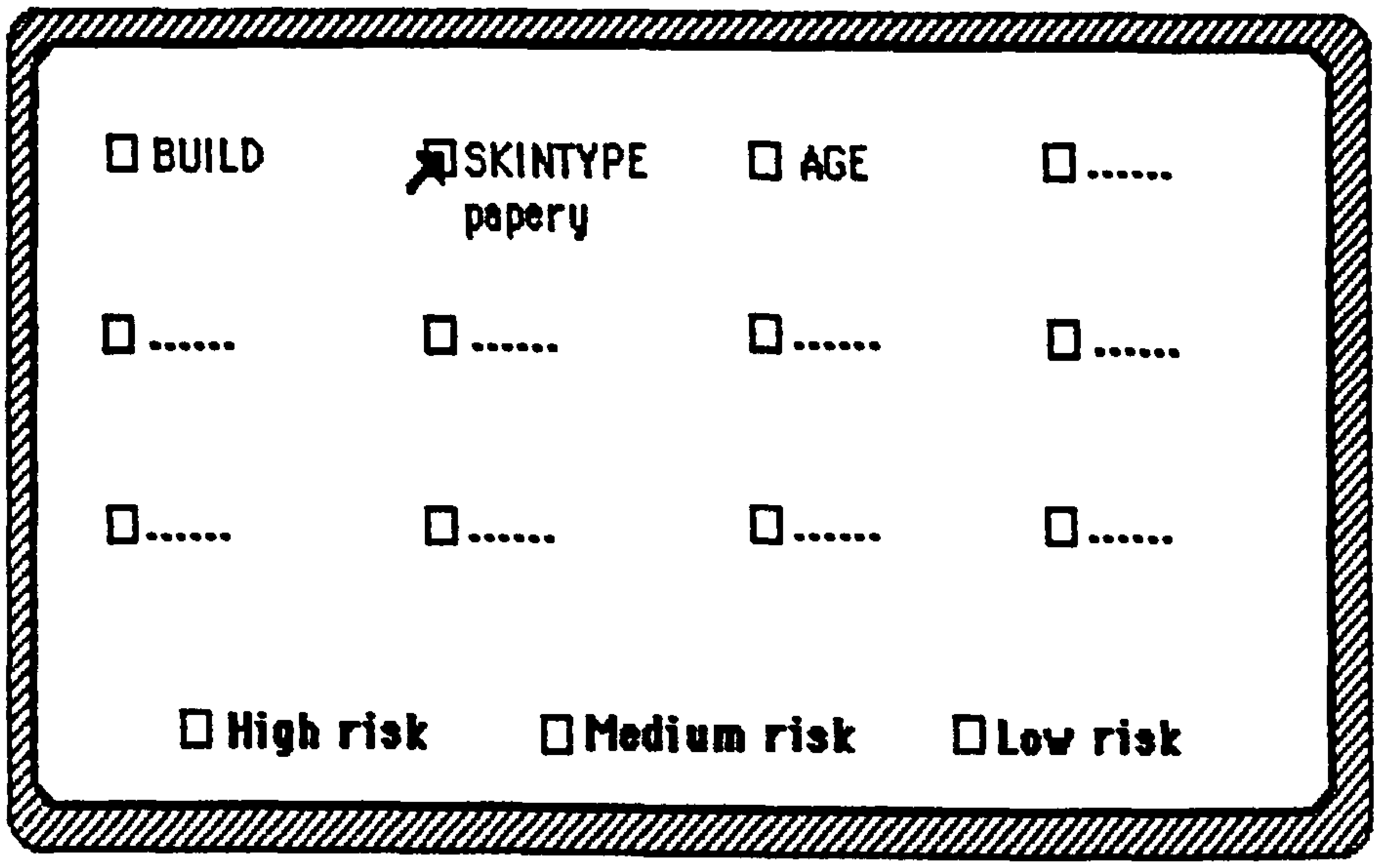
Table 2.5 Presentation order of Patients 1 to 18 within 36 trials

trial	condition	Patient	trial	condition	Patient
1	Select	13	19	S	16
2	S	7	20	S	10
3	S	1	21	S	4
4	Allup	16	22	A	15
5	A	10	23	A	9
6	A	11	24	A	14
7	S	14	25	S	5
8	S	15	26	S	11
9	S	8	27	S	12
10	A	17	28	A	3
11	A	4	29	A	13
12	A	18	30	A	8
13	S	9	31	S	6
14	S	2	32	S	17
15	S	3	33	S	18
16	A	12	34	A	2
17	A	5	35	A	7
18	A	6	36	A	1

Apparatus used both to present patient descriptions to subjects and to record data was an Apple Macintosh 512K microcomputer running a program specifically prepared for the purpose. This program was designed to run the experiment without presence of the experimenter; subjects had only to be trained in the operation of the 'mouse' in order to run themselves through instructions, 4 practice trials, and each trial proper.

A SELECT screen is depicted in Figure 2.2 overleaf. Adjacent to each button (□) the name of an attribute is displayed, although in this example only 3 attribute names are shown. The subject requires to know the values of some or all of these attributes before a decision can be made. To find out a value, the subject manoeuvers the mouse to the appropriate button and 'clicks', whereupon the value appears below the attribute name. In the example the mouse arrow has been clicked on SKINTYPE, hence revealing that this particular patient has 'papery' skin. Three more 'decision' buttons were placed at the foot of the screen - High, Medium, and Low risk. When the nurse had elicited sufficient information to make a 'risk' decision, she ended that patient's assessment by clicking one of these three buttons.

Figure 2.2 Depiction of screen faced by experimental subjects



For SELECT trials, the computer recorded the order in which each attribute was searched (if it was searched) and the eventual decison arrived at by the subject. For ALLUP trials only the decision could be recorded.

Procedure followed was identical for each subject. An explanation was given that they would be presented with patients which they were to assess with regard to pressure sore risk assessment. Each subject ran the program in a room alone after some initial instruction on how to use the mouse. A parting request from the experimenter was to ask subjects to 'think out loud' such thoughts as occur to them while they were assessing the patients. At this point the experimenter switched on a cassette tape recorder left the room.

The experiment, as mentioned earlier, was entirely self-paced. Subjects began by going through a sequence of screens designed to familiarisde them with the mouse and to give them more explicit instructions regarding the task. These instructions are reproduced in Appendix 3. There followed 4 practice trials (2 SELECT and 2 ALLUP) which familiarised the subject with the task and the type of information afforded by each attribute. Subjects were not told how many trials there would be, only that the exercise would take around 45 minutes (a realistic figure arrived at from timing 3 pilot subjects). Midway through the trials the computer advised the subject that she had earned a well-deserved break and she could contact the experimenter for some refreshment.

The analysis of the extensive data provided by the experiment will occupy the next Part of this Chapter (when expert performers will be identified) and a large part of Chapter 3 (when the cognitive model is gradually constructed from the data).

PRELIMINARY ANALYSIS TO IDENTIFY EXPERT PERFORMANCE

Introduction

The first objective for analysis from the process tracing experiment is to identify subjects whose performance could for present purposes be taken as expert. The information processing behaviour of these nurses can then, in Chapter 3, be subjected to both more rigorous and more qualitative analysis. In order to achieve this goal it is necessary to ask the following principal questions of the quantitative data:

1. What measures can be inspected in order to identify good performers?

No assumptions are made at this stage. The following pool of measures are available and could be analysed for possible reliable variation across subjects:

a) decision concordance - the extent to which subjects arrive at the predetermined risk judgement of the 18 patients.
b) number of attributes selected - in the SELECT trials, the number of items of information elicited for each patient.
c) consistency - given the possibility of order effects across the 36 trials, is there evidence of performance decrement within the subjects.
d) condition differences - using measures such as a) and c), was variation evident between SELECT and ALLUP trials.
e) risk decision differences - the extent ot which variation was a function of whether patients were in the high, medium, or low risk classification.
f) experience - the years of nursing experience of subjects.

2. Which grouping of subjects can reasonably be taken as representing expertise on the task?

If identification of such a grouping could be achieved then the individual and collective performance could be analysed at a closer and more qualitative level - the goal being to aim toward constructing an emulation of the cognitive expertise identified.

Beginning with the first question about candidate measures for inspection, the strategy which is adopted is not, where possible, to consider each factor in isolation. Decision concordance, for example, might be a function of the degree of pressure sore risk and/or of whether the trial is a

SELECT or an ALLUP one. The goal, therefore, will be to search for possible interactions between the measures as a sensible precaution against obtaining missleading results. The various analyses below repeatedly employ decision concordance and number of attributes selected as dependent measures; the other factors are analysed in the form of independent variable groupings.

Analysis of Decision Concordance

It is possible to maximally get 18 patients out of 18 correct in each condition. The scores by subject are set out in Table 2.6 below.

Table 2.6 Number of 'Correct' Trials by Subject and by Condition

subject	n correct SELECT condition	n correct ALLUP condition	total
1	12	11	23
2	15	13	28
3	14	12	26
4	15	16	31
5	14	13	27
6	14	12	26
7	13	15	28
8	16	12	28
9	13	15	28
10	13	16	29
11	11	13	24
12	13	14	27
13	10	7	17
14	10	8	18
median	13	13	27
mean	13.07	12.64	25.7
sd	1.82	2.68	4.01

Taking the data at its most 'coarse', as in Table 2.6, there are no apparent differences between the conditions using this dependent variable. A similar picture emerges when a superficial view is taken of the same measure as it applies to other independent variables such as years of experience and risk classification of patient....

Experience of nurses (split above and below the median years)
- more experienced, mean trials 'correct'= 27.14, sd = 2.48, median = 28
- less experienced, " " " = 24.29, " = 4.89, " = 27

(maximum correct = 36, no significant differences between groups)

Risk of patients (high, medium, and low)

- high risk, mean trials correctly judged = 8.14, sd = 3.57, median = 8.5
- medium risk, " " " " = 6.43, " = 2.21, " = 6.5
- low risk, " " " " = 11.07, " = 1.21, " = 11.5

(maximum correct = 12, high vs low t=3.65 $p<0.05$
med vs low t=5.94 $p<0.001$)

The superficial level of analysis, therefore, becomes less than rewarding with regard to achieving the goal of identifying an 'expert' sub-group. A more rational approach to analysis of this data would be to consider that performance accuracy is a function not only of the condition (SELECT vs. ALLUP) but also both of the years of experience of each nurse and the risk grouping of particular patients.

A test of this thesis becomes possible by treating scores (number of correct decisions per nurse) in a repeated measures analysis of variance (ANOVA) using:

1 between subjects factor - Years of experience; 2 levels, above and below the median for the whole sample.

2 within subject factors: A. - Condition; 2 levels, SELECT and ALLUP.
B. - Risk; 3 levels, High, Medium, and Low.

Results, presented in Table 2.7 overleaf, suggest conclusions beyond those possible from the superficial approach of considering the variables in isolation. Of the main effects only Risk is significant ($F=17.54$, $p<0.001$). Post-hoc testing (Tukey's HSD) reveals that in terms of accuracy the low risk patients are most accurately judged followed by high risk and then medium risk. The differences are all significant at the 1% level with the exception of low vs medium which is significant at the 5% level. Clearly there is sense in considering Risk as a separate factor under each Condition.

Table 2.7 ANOVA of Number of Correct Decisions by Experience, Condition and Risk

Source of Variation	df	SS	MS	F	p
EXPERIENCE	1	4.76	4.76	1.90	.19
error	12	30.05	2.50		
CONDITION	1	.43	4.3		
EXP X CONDITION	1	.05	.05	.054	.82
error	12	10.52	.88		
RISK	2	76.78	38.39	17.54	<.001
EXP X RISK	2	30.02	15.01	6.86	.004
error	24	52.52	2.19		
COND X RISK	2	4.07	2.04	2.47	.10
EXP X COND X RISK	2	.17	.08	.10	.90
error	24	19.76	.82		

There is in addition an interesting interaction between Risk and Experience ($F=6.86$, $p<0.005$). Hence although Experience as a main effect fails to achieve significance, the interaction suggests subtle but important conclusions. As Figure 2.3 below illustrates, there seems not to be great differences in performance of the two groups of nurses when they are judging either low or medium risk patients. However, a gap is evident with respect to high risk patients. Analysis of Simple Effects seems to confirm the conclusion that Experience has a significant effect on performance only when judging the risk of patients who were predetermined as being of a high risk of developing pressure sores (see Table 2.8 overleaf).

Figure 2.3 Accuracy of decisions by More Experienced and Less Experienced Nurses when judging high, medium, and low risk patients.

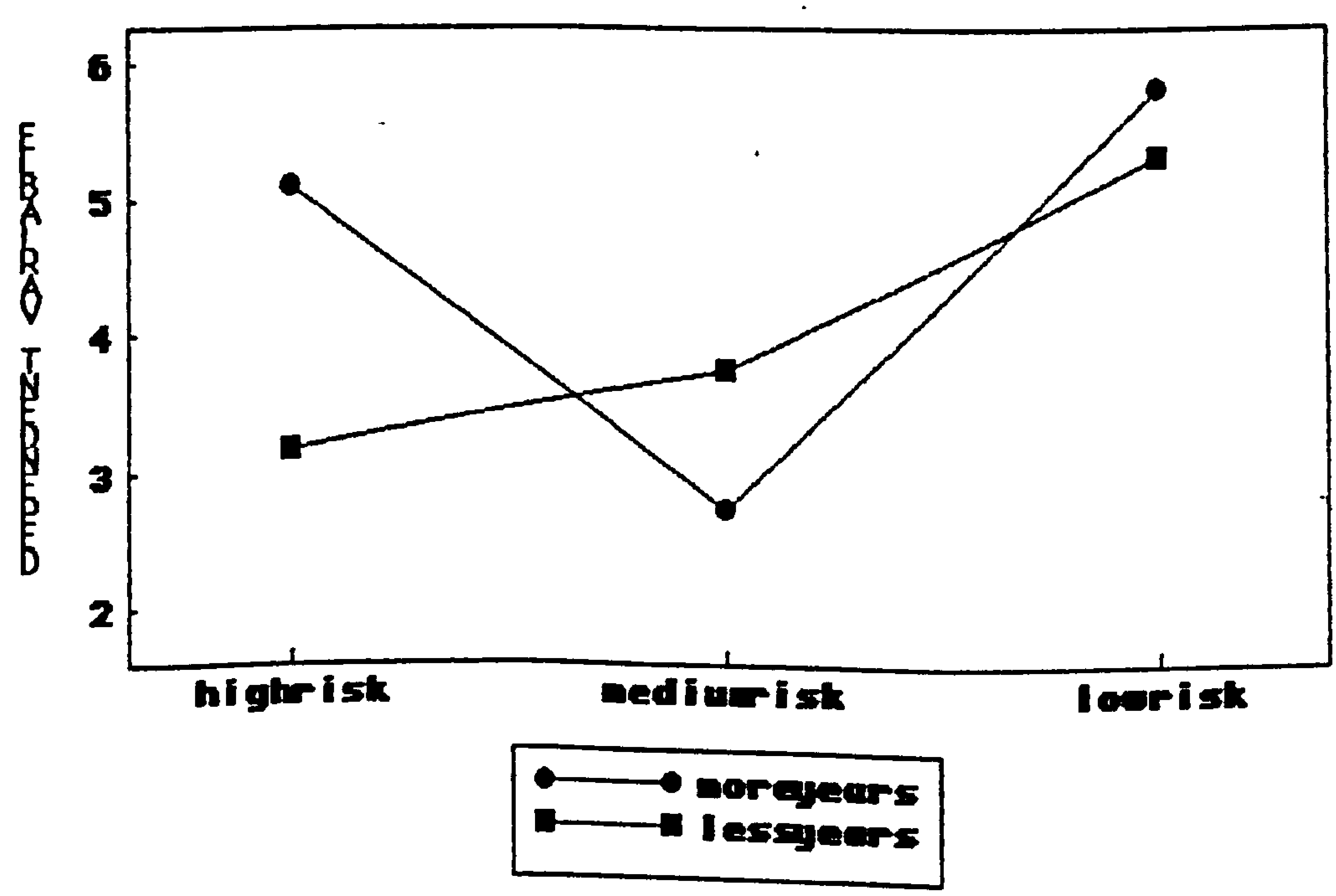


Table 2.8 Simple Effects analysis based on results depicted in Figure 2.3

Effect	MSn	DFn	DFe	MSe	F	p
Experience at high risk	26.04	1	36	2.29	11.35	.002
" " medium risk	7.00	1	36	2.29	3.05	.089
" " low risk	1.75	1	36	2.29	.76	.388
Risk at more Experience	36.17	2	24	2.19	16.52	<.001
Risk at less Experience	17.24	2	24	2.19	7.87	.002

The interim position with regard to the questions posed earlier is that apparently only the variable Risk can be taken as having the clear ability to influence the accuracy of judgements. Years of experience has a subtle but important effect. One possible criticism is that division of the subjects into only two groups (above and below median years) acts to increase the probability of a Type 2 error - the null hypothesis that there are no differences between groups will be hard to reject given that the edges of the two groups 'touch'.

The test of this criticism is to divide the nurses into 3 groups - most, mid, and least years of experience - and to test for possible trend effects across the three groups. ANOVA, as Table 2.7 but with 3 levels of Experience, was performed. Once again no significant main effect for Experience was found (disbarring post-hoc trend analysis), also the significant Risk effect and Experience X Risk interaction was preserved. One final test of the Experience effect was to repeat the analysis by establishing 2 groups which excluded the middle-experienced nurses - testing the most experienced 5 subjects versus the least experienced 5. A repeat of the same ANOVA design again failed to provide evidence for a reliable main effect for experience, demonstrating that for this dependent variable it can be concluded that no simple rule exists along the lines of 'the more experienced the nurse the better the decision accuracy'.

Testing for Order Effects using Decision Concordance

A second analysis using counts of decision concordance as a dependent variable will test for performance decrement over the course of the experiment. If assessing 36 patients was seen as fatiguing or boring then this may lead to reduced accuracy with respect to the patients appearing relatively late in the experiment. Alternatively, there might be a practice effect which could lead to improved performance. Aside from testing for evidence of an order effect *per se*, the hypothesis of interest is that some nurses will be more resistant to performance variation - hence candidates for being taken as more 'expert'.

At a superficial level of analysis, it is possible to correlate the order of appearance (position) of each patient with the number of concording decisions recorded for that patient .....

SELECT condition (n = 18) Kendall's W = 0.014
ALLUP condition (n = 18) " = 0.1
total (SELECT+ALLUP) (n = 36) " = 0.06

There is no apparent support for an order effect suggested by these correlations. However, it is possible to increase the sensitivity of this analysis by blocking the 36 patients into 3 groups of 12 - those presented during the first third of the experiment, those in the middle third, and those in the final phase of the task. If such an effect exists then it could be contingent on whether Experience of the nurse and/or on the Condition of presentation (SELECT vs ALLUP). In order to test these hypotheses a repeated measures ANOVA can be run with:

1 between subjects factor - Years of experience; 2 levels, above and below the median for the whole sample.
2 within subject factors: A. -Order; 3 levels; 1st, 2nd, and 3rd phase.
B. - Condition; 2 levels, SELECT and ALLUP.

The null hypothesis is that there are no differences in decision concordance attributable to the phase of the experiment.

Results, set out in Table 2.9 below, indicate no significant main effect for Order. There was no basis, therefore, for embarking upon trend analysis.

Table 2.9 ANOVA of Number of Correct Decisions by Experience, Order and Condition

Source of Variation	df	SS	MS	F	p
EXPERIENCE	1	2.68	2.68	.90	.36
error	12	35.71	2.97		
ORDER	2	2.95	1.47	1.79	.19
ORDER X CONDITION	2	2.00	1.00	1.22	.3
error	24	19.71	.82		
CONDITION	1	.96	.96	1.56	.23
EXP X CONDITION	1	.11	.11	.17	.68
error	12	7.43	.62		
ORDER X CONDITION	2	7.71	3.86	3.27	.055
EXP X ORD X COND	2	2.00	1.00	.85	.44
error	24	28.28	1.18		

In terms of Question 1, then, the position at this point is that Performance Decrement has not been supported as a variable which might be used to identify a subgroup of 'expert' nurses. In contrast, the variable Risk has acquired high status as a variable which, in conjunction with Experience, might be useful for achieving this goal. In particular it is the assessment of high risk patients which distinguishes the better performers. The status of the variable Condition shows little promise.

Analysis of Number of Attributes Selected

Continuing the analysis of factors which may vary as a function of expertise leads to consideration of the number of attributes searched. The hypothesis here becomes one-tailed in that 'experts' might be expected to have developed more efficacious processing strategies which result in less information being necessary before coming to decision point. Perforce, the analysis refers only to SELECT trials.

The nurses, on first inspection, seemed to vary considerably with respect to the amount of information selected, as Table 2.10 shows. Nevertheless, large variability within each nurse should not be taken as evidence of uncertain expertise - on the contrary, it could be argued that the differing patient characteristics with respect to 'problem attribute values' argues for expertise being about possession of the ability to tailor the amount of information required to the patient at hand.

Table 2.10 Mean values selected by nurses 1 - 14

NURSE	mean values selected	median	sd
1	3.94	4	1
2	3.89	3.5	1.6
3	4.94	5	1.47
4	5.11	4	3.39
5	4.33	4	1.5
6	4.16	3.5	1.72
7	6.66	6	2.11
8	5.72	6	1.13
9	7.22	8	1.52
10	5.0	5.5	1.41
11	8.83	8.5	2.33
12	4.0	4	0.91
13	5.83	5.5	4.38
14	7.61	8	2.2

Similarly, there seemed to be variation between each Risk group - high risk (mean=5.36), medium risk (mean=6.16), and low risk (mean=4.84). It might also be predicted that this dependent variable would be affected by order effects, with less attributes being selected as the experiment proceeds.

An analysis was designed to investigate these hypotheses. Scores (number of correct decisions per nurse) in a repeated measures ANOVA using:

1 between subjects factor - Years of experience; 2 levels, above and below the median for the whole sample.

2 within subject factors: A. - Order; 3 levels- 1st, 2nd, and 3rd phase.
B. - Risk; 3 levels, High, Medium, and Low.

Results, as presented in Table 2.11below, indicated significant main effects for both factors - how many items of information are selected is determined both by the risk of the patient and by the experience of the nurse. The interaction between these two factors was not however significant.

Table 2.11 ANOVA of Number of Values Selected by Experience, Order and Risk

Source of Variation	df	SS	MS	F	p
EXPERIENCE	1	75.14	75.14	6.03	.03
error	12	149.60	12.47		
ORDER	2	3.65	1.83	.61	.55
EXP X ORDER	2	12.29	6.15	2.06	.15
error	24	71.50	2.98		
RISK	2	37.29	18.64	10.44	.0005
EXP X RISK	2	.75	.37	.21	.81
error	24	42.86	1.78		
ORDER X RISK	.4	31.48	7.87	3.59	.01
EXP X ORDER X RISK	4	2.81	.70	.32	.86
error	48	105.21	2.19		

Notable among the main effects is the significant F ratio for Experience (F=6.03, p=.03). The corresponding mean number of values selected for each were: more experienced - 4.68, less experienced - 6.23. This result, significant at the 5% level, lends clear support to the hypothesis that experienced nurses will have developed 'more efficacious' information processing. Later in the analysis it will be shown that the nature of this reduction in necessary information reflected these nurses' ability to infer missing attribute values and to have a clear representation of which information would most effectively reduce uncertainty. For now, however, it suffices to note that the difference exists.

The Order effect could not, once again, be supported either as a main effect or as an effect interacting with Experience, indicating that there was no reliable trend for either group of nurses to alter the amount of information search solely as a function of the phase of the experiment. This result is impressive given the amount of patients to assess.

The Risk effect, shown to determine to a degree the accuracy of decision, once again demonstrates its power to affect information processing ($F=10.44$, $p=.0005$).

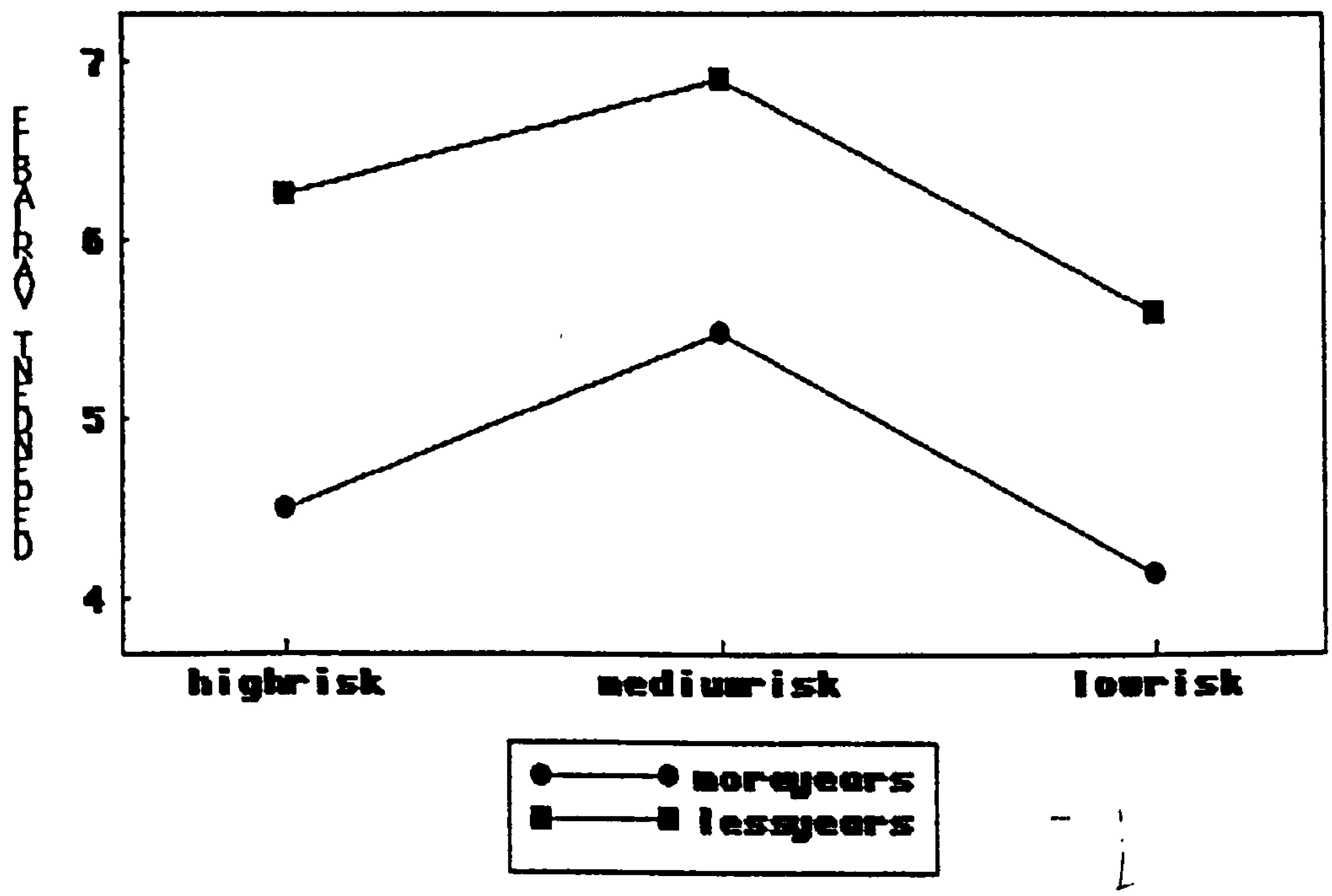
The corresponding mean number of values selected for each risk group of patient were:
low risk - 4.84
medium - 6.17
high risk - 5.36.

Post-hoc testing (Tukey's HSD) showed:
high risk vs. medium risk - $p<.05$
high risk vs. low risk - not significant
medium risk vs. low risk - $p<.01$.

That most information was selected when a medium risk patient was being assessed seems rational - these patients typically had a mixture of 'problem' and 'no-problem' attibute values. Low risk patients have important attribute values in the no-problem category, hence the early end to these assessments.

The graphical display of number of values selected (depicted in Figure 2.4 overleaf) illustrates both the differences across patient groups. Also illustrated is the constant difference between the more experienced nurses when compared to the less experienced group.

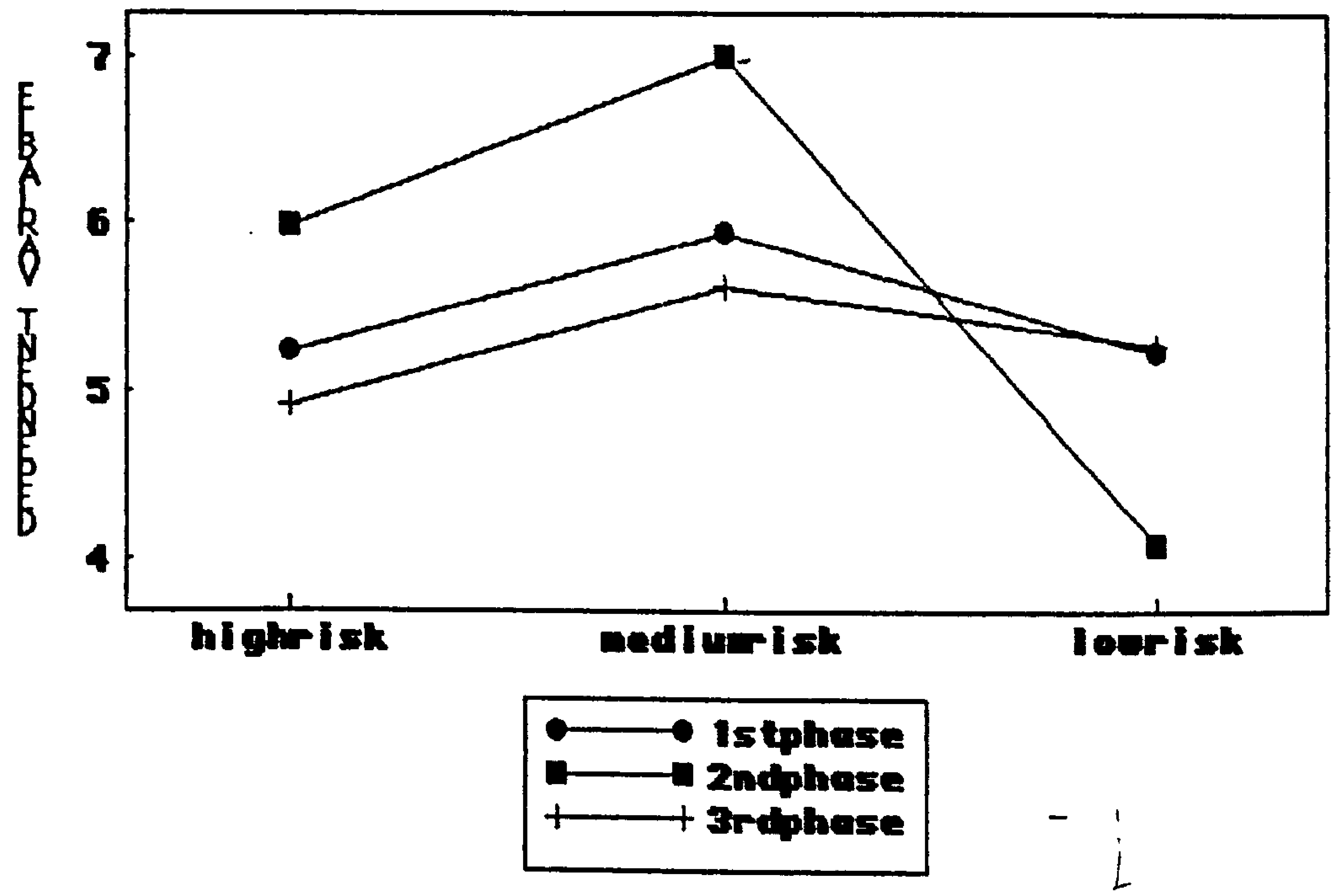
Figure 2.4 Number of attribute values selected by Risk group and Experience group.



The remaining effect which was significant is the interaction between Order and Risk. Potentially this interaction, illustrated in Figure 2.5 overleaf, could pose problems for the conclusion thusfar that Order (and therefore Performance Decrement) has not been a reliable feature of the subjects' processing. However, as the Figure 2.5 shows, there is firstly no trend evident that number of values selected progressively decreases from 1st phase through to 3rd phase - in fact the (nonsignificant) trend is 2nd to 1st to 3rd.

To find the source of the significant interaction it is necessary to calculate the Simple Effects from this interaction. The effect of Order was found not to be significant for any of the risk groups. The effect of Risk, however, was found to be highly significant ($F=17.93$, $DFn=2$, $DFe=24$, $p<.001$) only with respect to the 2nd phase of the experiment - as the graph illustrates particularly at the low risk point. What this signifies is that the low risk patients assessed during the 2nd phase of the experiment had relatively few values selected. This result cannot be taken to be of any meaningful significance.

Figure 2.5 Number of values selected during each phase by Risk group.



Analysis to Group Subjects by Expertise

To the extent that the first part of Objective 1 has been achieved - the development of indices of expert performance - one final analysis remains if the second part of the Objective is to be achieved - the screening out of subjects who do no not perform to an expert standard. In terms of the questions set out earlier which could be asked of the data, which grouping of subjects can reasonably be taken as representing expertise on the task? More specifically, the requirement is to specify relatively homogeneous groupings of subjects who can be henceforward taken as 'expert' and 'proficient' nurses. Further analysis of task performance of a more qualitative nature can then be performed by comparing each group. It should be stressed, nevertheless, that the grouping of subjects represents a definition of expertise which is largely local to this exploratory experimental work. At some later point, it follows, the requirement will become to test the external validity of any conclusions about the nature of nursing cognitive expertise which are offered by this project.

The procedure adopted to establish the homegeneous groupings is to firstly take a 'profile' of each subject and secondly employ an appropriate statistical routine which is capable of discriminating groupings of subjects on the basis of the individual profiles. The profiles comprise individual values on each of the factors which have been found to vary significantly within the whole group. Hence mean number of attributes selected by classification of Risk, performance accuracy, and years of experience values will make up the profiles. The values for the N=14 subjects on these factors are set out in Table 2.12 below.

Table 2.12 Values for all subjects on number of attributes selected, performance accuracy and years of experience

	n attributes selected			Accuracy		Years experience
subject	H	M	L	Select	Allup	
1	4.5	4	3.3	15	16	12
2	4.5	4.7	2.5	14	12	16
3	8.2	7.7	7	15	16	30
4	3.7	6.2	5.5	14	12	11
5	3.5	5.7	3.8	13	15	13
6	4.3	4.7	3.5	16	12	15
7	6.2	7.5	6.3	13	15	19
8	5.8	5.5	3.5	13	16	6
9	5.7	6.2	5.3	15	13	4
10	7.3	8	6.3	14	13	4
11	6	5.6	3.3	13	14	8
12	8.2	9.8	8.5	11	13	5
13	4.2	4.3	3.5	10	7	3
14	7	6.5	4	10	8	3

A cluster analysis procedure was adopted for forming homogeneous groups of subjects based on these profiles. It was firstly necessary to compute an index of the 'distance' between each subject. A matrix of coefficients of distance between all possible pairings of subjects was obtained by inputing the profiles to the SPSSx PROXIMITIES procedure using the measure of squared Euclidean distance. Prior to computing the coefficients of distance and since the factors employed different metrics, each value was transformed to a standardised score on the respective factor.

The requirement from these matrices was a procedure for identifying subgroups of subjects (or clusters) based on the proximity coefficients. To achieve this, each matrix was input to the SPSSx CLUSTER procedure using the hierarchical agglomeration algorithm based on average linkage between groups. The algorithm operates by initially considering each subject as an individual cluster. From these 14 clusters, at step 1 the two 'closest' subjects are combined into a single cluster - hence forming 13 clusters. At each subsequent step an additional cluster is formed either by joining a subject to an existing cluster, two separate subjects into a single cluster, or two multi-subject clusters until all 14 subjects are merged into a single cluster.

Cluster analysis - results

Results are presented for the Experienced and the Beginner groups in the form of Agglomeration Schedules and Dendrograms. The Agglomeration Schedule contains the number of subjects or clusters being combined at each step. In Table 2.13 below, for example, the first line indicates under 'Clusters Combined' that subjects 8 and 11 were joined at this stage. The squared Euclidean distance between these two clusters (subjects at this point) is given in the column 'Coefficient'. The final column - "subjects merged" - indicates which subjects or groups of subjects were being combined at each step.

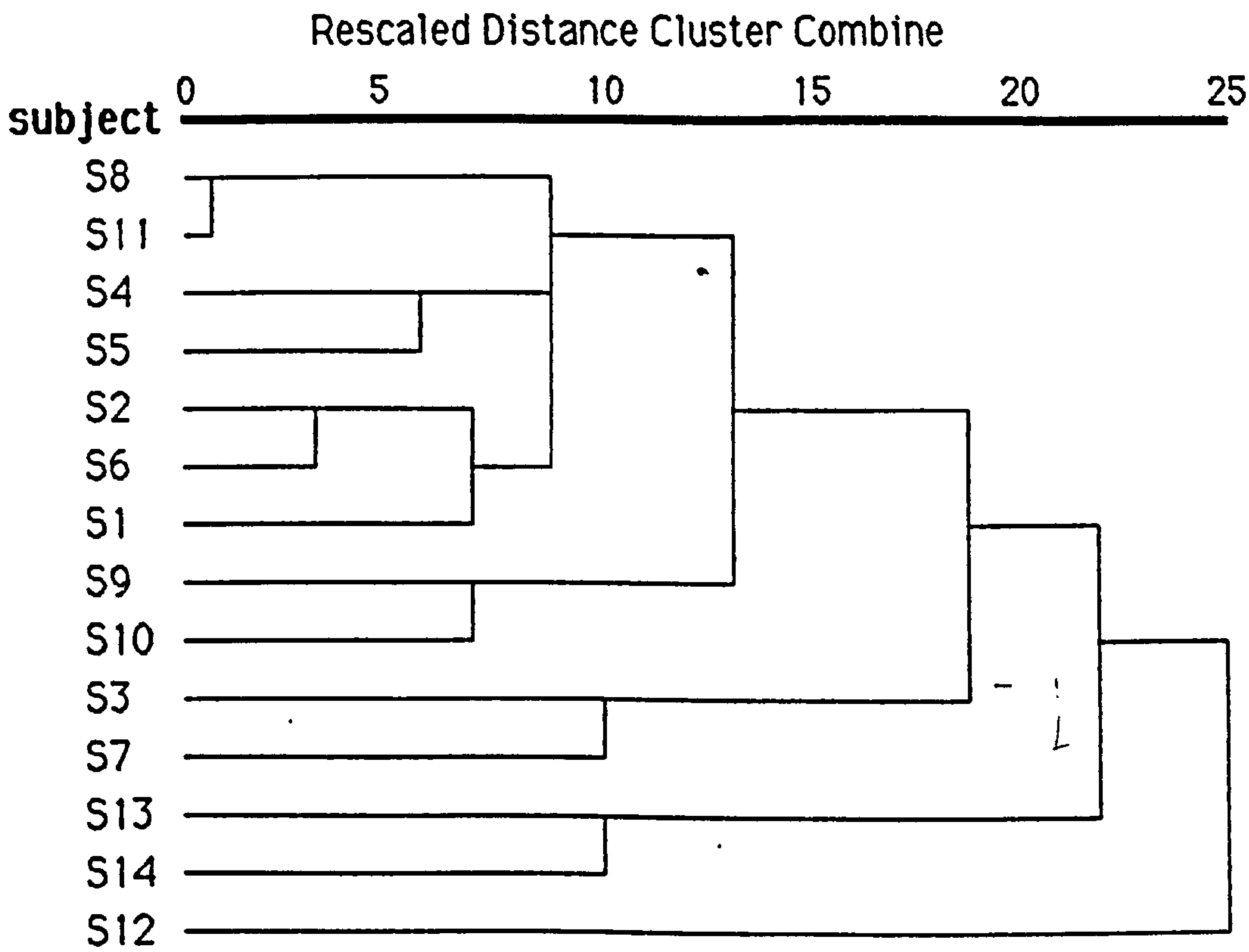
Table 2.13 Agglomeration Schedule for cluster analysis on profiles of 14 subjects

	Clusters Combined			
Step	cluster 1	cluster 2	Coefficient	subjects merged =
1	8	11	.79	**8+11**
2	2	6	1.23	**2+6**
3	4	5	1.60	**4+5**
4	9	10	1.69	**9+1**
5	1	2	1.70	2&6 **+1**
6	1	4	2.00	2&6&1 **+4+5**
7	1	8	2.10	2&6&1&4&5 **+8+11**
8	3	7	2.26	**3+7**
9	13	14	2.27	**13+14**
10	1	9	2.70	2&6&1&4&5&8&11 **+9+10**
11	1	3	3.53	2&6&1&4&5&8&11&9&10 **+3+7**
12	1	13	4.05	2&6&1&4&5&8&11&9&10&3&7 **+13+14**
13	1	12	4.50	all other subs **+12**

The Dendrogram visually represents the steps in the hierarchical clustering solution. The clusters as they are combined are shown along with the values of the coefficients at each step. Produced by the SPSSx CLUSTER procedure, the dendrogram does not plot the actual proximity coefficients of each agglomeration step, rather the coefficients are rescaled to numbers between 0 and 25. The ratio of the distances between steps is, however, preserved.

Figure 2.6 overleaf displays the results of the cluster analysis in the form of a Dendrogram.

Figure 2.6 Dendrogram for cluster analysis on profiles of 14 subjects



In order to interprete the results of the cluster analysis it is necessary to take account of each subject's values on the factors which comprised the profile. The goal is to identify not only a cluster of subjects but also subjects who were seen to be 'strong performers' on the task as well being experienced. With these criteria in mind, the subjects upon whom attention should be focused are broadly those in the top half of listing of profiles. Immediately it becomes clear that steps 2, 3, 5, and 6 are of interest. These early steps represent the merging of 'potentially expert' subjects (1, 2, 4, 5, and 6) from the top half of the profile table. Subjects 8 and 11, merged at step 1, are much less experienced than this group and do not join with this group until step 7 - indicating that while these two subjects cluster together strongly, the subjects in the potentially expert group have greater commonality. Similarly, although subjects 9 and 10 are joined fairly early (at step 4), they do not merge with potentially expert cluster until step 10.

It is therefore beginning to look reasonable that the twin criteria of cluster homogeneity and strong performance should suggest that subjects 1, 2, 4, 5, and 6 can be taken as the 'expert' group. One difficulty is that two subjects had promising profiles yet are not in this group. These subjects - S3 and 7 - joined together at step 8 yet did not merge with the main cluster until a point where the distance coefficient was relatively large (step 11). The reason for relatively great distance from the other nurses who had above average length of experience seems to be the rather large number of attributes selected, yet because of respectable length of experience they do not seem to fit easily with the other nurses who selected similar numbers of attributes when assessing a patient. It is therefore proposed to exclude subjects 3 and 7 from either group and, by making the minimum of further assumptions, classify all remaining subjects as either 'expert' or 'proficient':

expert group - Ss 1, 2, 4, 5, and 6 (n=5)

proficient group - Ss 8, 9, 10, 11, 12, 13, and 14 (n=7).

Conclusions

1. Various indices have been explored in order to achieve the objective of identifying a sub-group of 'expert' subjects whose performance could be more closely analysed. In terms of the two dependent variables used - decision accuracy and number of attributes selected - the indices which reliably varied were principally length of experience and patient's pressure sore risk. More experienced nurses judge risk more accurately and moreover do so on the basis of less information than required by their less experienced counterparts. Experience seems to 'tell' particularily when it comes to high risk patients - presumably because of expert knowledge of the key factors affecting pressure sore risk.

2. The indices which varied reliably were used to construct profiles of each subject. A cluster analysis procedure was then employed to identify a relatively homogeneous groups of subjects who could henceforth be deemed as 'expert' or as 'proficient'. Further analysis will test the validity of the groupings of N=5 Experts and N=7 Proficients which have been established. It is to these more qualitative analyses that the discussion can now turn.

CHAPTER 3 ANALYSIS AND MODELLING OF EXPERT KNOWLEDGE

INTRODUCTION

Explanations of Expert Nursing Cognition - Towards a Cognitive Model

The exploration of subjects' performance on the 'simulated patient assessment' exercise shifts, in the following sections, towards a closer and more qualitative analysis of the observed processing of information. Hence, 'number of attributes selected' will be replaced by analysis of which attributes were selected and in what position. The shift will therefore be from a general through to a fairly specific level of description of the data. Paralleling this progression will be a series of arguments which, it is hoped, will take the level of explanation of nursing cognition from framework, via theory, to up-and-running cognitive model. In short, the aim is to describe the cognitive architecture of expert nursing information processing.

This goal - to provide an explanation of cognition which comprises a unified description of the component mental representations, memory structures and processing mechanisms - is ambitious and must necessarily achieve only partial fulfillment in this exploratory project. However, in order to assess the adequacy of the explanation, the chapter which follows will seek to provide a wide-ranging evaluation of that cognitive model and the predictions which arise from it. It is important to establish that the model aims to emulate expert nursing cognition as measured in the SELECT condition of the experiment - a model based on the ALLUP condition would require experimentation designed more specifically to provide data amenable to protocol analysis.

The overall strategy for providing this explanation is to look at the power of four main 'contender explanations' of the information processing. These explanations are derived from appropriate previous literature and might be expected to feature in this task. 'Attribute importance', for example, may feature in that subjects will search the patient attributes which are most likely to reduce uncertainty and facilitate decision making. The description and analysis of subjects' information processing, therefore, will be set out in four main Parts corresponding to these 'contender explanations'. The format of each Part will be as follows:

a) brief literature review to establish the explanation of the data,
b) broad exploration of the present data in order to establish the presence absence of the explanation,
c) specific test of goodness of fit to the data of a version of the explanation tailored to the task,
d) discussion of implications of evaluated explanation for the construction of the cognitive model.

The four 'contender explanations', and therefore the four Parts, are:

PART

1. Attribute importance. Since factors vary in terms of influence on risk of development of pressure sores, order in which information is selected might be expected to conform to a scheme which reflects this order of influence.

2. The use of higher cognitive processes The degree to which subjects go beyond what is known of a patient in order to infer other information.

3. The point for decision making. It has been established that on all but exceedingly few occasions the subjects left unsearched a varying number of a patient's attributes. It becomes necessary, therefore, to explore the conditions which describe the point at which the information gained is taken as sufficient.

4. The process of decision making The process whereby subjects arrive at one of the three decision alternatives.

Before beginning the exploration a final introduction can be offered to the data and to the method adopted for testing goodness of fit of the various explanations to the measures of information processing derived from the SELECT condition. The form of this data does not easily suggest a suitable testing scheme. Hence, prior to making a decision, subject *s* would search the attributes of patient *p*. Of the 12 attributes, some would be searched and others left unselected. The order of search of the selected attributes is recorded for all 18 patients. As an example of this data, the following table represents the order of attribute selection of one subject when assessing patients 1 and 9:

		patient																	
	attribute	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	MOBILITY	1								1									
2	URINARY	2								3									
3	SKINTYPE	5								4									
4	BUILD	0								0									
5	NUTRITION	4								5									
6	AGE	0								0									
7	MENTAL state	3								2									
8	CIRCULATION	0								6									
9	LIFT & TURN	0								0									
10	SEX	0								0									
11	BLOOD PRESS.	0								0									
12	FAECAL	0								0									
		decision = H								decision = M									

For patient 1, therefore, this subject began by selecting MOBILITY. On the value of this attribute being revealed (bedfast but can move freely in bed), the subject then selected URINARY CONTINENCE and was informed that the patient had a urodome fitted. After searching 5 attributes, the subject decided that the patient was a high risk of developing pressure sores. All unselected attributes are represented by zeros.

The difficulties of analysing this data are considerable, particularly when the aim is to construct a cognitive model based on a group of subjects. The complexity of between subject comparisons is considerable. To illustrate, the entire data comprises the process traces of 12 subjects assessing 18 patients. There are therefore 18x(11+10+9+8+7+6+5+4+3+2+1) = 1188 opportunities for comparing one subject with another. Only in three of these comparisons was an identical match located between two subjects' process traces. At first sight this result might be seen as seriously compromising the goal of basing a cognitive model on 'averaged' expertise.

An attempt to establish how many matches would be expected by chance provides a different perspective, however. There were 12 attributes available for search. The process trace is not only a record of whether a given attribute was searched or not but also a record of the order in which information was gathered from the searched attributes. Hence, the record for one patient of each of the 12 attributes can show a value from 0 to 12. Up to 11 zeros are possible, but for a whole number to be present there must be a) only one of this whole number, and b) no deviation from an incremental series beginning with one (ie if there is a 5 then must elsewhere be a 1,2,3, and 4). The increase in number of possible permutations of data is exponential up to the point when a value (including a zero) cannot repeat, ie 11 or 12 attributes selected. In this case there are 12! possible permutations of the ordinal positions in which attributes were selected in a single process trace. The permutations for each number of attributes selected are as follows:

n attributes selected	n permutations
1	12
2	132
3	1320
4	11880
5	95040
6	665280
7	3991680
8	19958400
9	79833600
10	239500000
11	479000000
12	479000000

These permutations apply to each subject - hence in one of the 1188 comparisons of traces it could be that subject A selected 5 attributes and subject B selected 6. The probability of an exact match is therefore 1/95040 X 1/665280. The possible permutations of data, therefore, are considerable. That there were small numbers of attributes selected on the three mathching traces is not surprising - the first match found (within patient 6) was of 2 subjects each selecting only 1 attribute prior to decision, the matching subjects within patient 13 selected 3 attributes, while for the third match (patient 2) there were only 2 attributes selected.

Despite these difficulties, however, the requirement is for a reliable and stringent index of the power of each of the contender explanations above to explain each subject's process trace. A method would be to consider each subject's process trace as a matrix comprising 216 cells (12 attributes X 18 patients). It would be possible to draw up a matrix for each of the explanations, for example the attributes ranked in order of importance as in explanation 1. The procedure could be to simply match the two matrices and count the number of occasions when exactly the same value co-occured in a given row and column cell.

This index - of perfect agreement - would fulfil the criteria of stringency in that even if matrix 1 had 4 against NUTRITION and 5 against BUILD for patient 1, and matrix 2 had 5 against NUTRITION and 4 against BUILD for the same patient then this seemingly trivial difference would be ignored by the index. The 'perfect index', however, does not provide a measure of reliability and hence can only be taken as a descriptive index, albeit of a high level. It is intended, therefore, to employ this method throughout the next sections in order to test goodness of fit of each explanation to the data.

Finally, reference should be made at this point to a second large database of n=159 patients which has been assembled since some of the analyses in this chapter will utilise the database information. This collection of patients, henceforth referred to as 'database2', can be regarded as more reliable than database1 of n=154 patients (see Chapter 2) due to a design improvement in the questionnaire. As identified earlier, a criticism of the questionnaire which database1 nurses completed was the potential ambiguity of the question ...."Indicate your judgement of this patient's risk of developing pressure sores". As was pointed out to the experimenter subsequently, some nurses who had confidence in their provision of pressure sore preventive care may have answered 'low risk' for a patient who might have been of a higher risk of developing sores by nature of their problems. For this reason, therefore, a second database of patients was established using a form which went to some lengths to ask for risk judgements independent of the adequacy of preventive care (see Appendix 4). Greater confidence, therefore, has been placed in this database.

ATTRIBUTE IMPORTANCE

Selected Literature Review on Attribute Importance

The concept of attribute importance or significance is central to each of the principal theoretical models of decision making. Hammond (1980), in a review of the similarities between six of these models, makes this point with respect to 'weights of cues' which are processed whether by mathematical formula or by human decision maker. Hence decision theory, behavioural decision theory, social judgement theory, information integration theory, and attribution theory each make use of the concept of the importance of a piece of information to individuals' judgements.

Each of the theories reviewed by Hammond (1980) are concerned with mathematical models of decision processes, and as such much of the research endeavour has been concerned with the establishment of numerical weights which are optimal for the differential specification of the various attributes in the decision formula (McClelland 1978). However, as Fox (1980) argues, it might be more appropriate if this type of linear prescriptive approach concerned itself with demonstrating correspondence with human psychological processes. One approach which has taken this route and largely abandoned preoccupation with formal mathematical models of decision making is that of Tversky and Kahneman (1974). The focus of this 'heuristic' approach has been to demonstrate that humans often employ rules of thumb based on factors such as how easily a fact can be recalled when making judgements. For the purposes of the present discussion, however, it can be noted that the concept of attribute importance (or salience) is even more acutely stressed by this more cognitive model.

The central idea, then, is one reduction of uncertainty through information gain. Heuristic search, according to Newell and Simon (1972) is the process whereby humans reduce the problem space by selectively gathering that information most likely to produce a solution. The more interesting question, however, is not that information has differential importance but *what* information will reduce *which* uncertainty. In other words, the scheme of attribute importance will depend to a large extent on the nature of the mental representation which the problem solver is constructing of the problem. This 'representation issue', central to cognitive psychology, will become increasingly central to this study of nursing cognitive expertise.

The recent concern of the nursing literature is more and more on the importance of cues and information to the task of patient assessment and care planning. Hammond and co-workers conducted the seminal studies in the area from a mathematical model perspective. It was demonstrated that nurses collect far more information than necessary and that much of the information is irrelevant (Hammond et al 1966). Moreover, the order of information acquisition by the nurses agreed only weakly with the 'significance' of the cues. These conclusions, however, are from the point of view of a prescriptive numerical model which assumes that a nurse's representation of the importance of cues is (or ought to be) the same as utility values within the mathematical formula. Moreover, the more serious assumption that is being made is that the nurse is attempting to arrive at the same mental representation as the prescriptive model - a diagnostic decision. As Fox (1987) argues, the demonstration that humans do not (at least under laboratory conditions) weigh up pros and cons of decision alternatives as well as they 'should' may not mean very much if they are not trying to weigh up pros and cons in the first place.

More recent nursing literature has continued in this theme. On the one hand demonstrations of failure by nurses to process 'relevant' information and on the other prescriptions for improvement. Virtually all of these findings or theorisations have the common assumption that the nursing assessment task is one of fitting incoming patient details to a stored diagnosis. Hence Gordon (1973) confirmed the earlier findings about nurses' collection of data which was irrelevant for the testing of hypotheses about the nature of the patient they were dealing with. Elaborate teaching texts have been written by Carnevali (1983) and Gordon (1987) containing several references to the need for nurses to search for 'relevant' or 'diagnostic' cues which will confirm a stored pattern called down from memory. Thiele et al (1986) have reported apparent success of a CAL program designed to teach novice nurses the importance of paying attention to diagnostic cues.

The cornerstone to these approaches to attribute importance is, therefore, the hypothetico-deductive model of clinical reasoning. This model, slightly adapted from medicine (where it is beginning to be challenged), assumes that the goal of nursing patient assessment is to match surface descriptors of pateints to nationally-agreed categories such as 'skin integrity, impairment of: potential' (ie at risk of developing pressure sores). Discussion, criticism, and testing of this assumption will figure large in the remainder of this project. For now, however, it is sufficient to note that a fairly superficial representation rather than one based on deeper knowledge is being suggested.

If superficial representations were optimal then it might be expected that evidence for their existence would be found in expert nurses. Leaving aside the surprising lack of evidence of any kind, it seems that this prediction has not been fulfilled. Broderick and Ammentorp (1979), for example, found no differences in the priority given to patient information by experts or by novice nurses. Stainton (1988) cites many other demonstrations that expert nurses do not appear to be using a hypothesis testing style - findings which have usually led to calls for improved teaching (eg Padrick 1988).

However, if a nurse's scheme of information importance does not appear to conform to the requirements of this surface level representation it does not follow that her performance is suboptimal. It could well be that the nature of her representation of the patient's problems is at a deeper level which implies a different scheme of attribute importance. Since every nursing author agrees that there are very good nurses around, the point might well have been missed that it is faulty emphasis within the theoretical frameworks rather than faulty cognition within the nurse. Would it not be better, as Stainton (1988) suggests, to base our models on an understanding of our expert nurses' cognition?

This introduction, then, has established that a useful starting point in the analysis of the observed information processing of the experimental subjects would be to look for evidence of attribute importance. The more complex issues raised will begin, it is hoped, to be clarified and further explored in subsequent sections.

<u>Exploration of Attribute Importance in the Data</u>

The literature review above strongly suggests that it might be fruitful to search for evidence that nurses in the present experiment are using some sort of scheme of attribute importance. There is, perhaps, no more appropriate place to begin this exploration than at the first (or 'header') attribute selected by the subjects:

<u>Header attribute choice</u>

A useful starting point in this qualitative analysis is to look at the subjects' choice of first attribute in the SELECT trials - as if the subject is saying "What will I find out about first?". It becomes immediately clear that the attribute MOBILITY was the overwhelming first choice of the 12 subjects who comprise the Expert and Proficient groups. Hence out of a total of 216 trials (12 subjects x 18 patients) the attribute MOBILITY was 'clicked' first on 138 occasions (63.8%). The next-most popular attribute for header selection was AGE - a poor second place at 38 occasions (17.5%) of trials.

Differences were evident in group patterns. The Expert group (N = 5) selected MOBILITY on 75 out of 90 occasions (83.3%), while the Proficient group (N = 7) opted for MOBILITY on only half of the trials (63 out of 126 occasions). This difference was found to be highly significant (x^2= 23.86, df1, p< .001).

Before drawing conclusions from the apparent primacy of MOBILITY it is worth looking at the occasions when other attributes were selected in the header position. It has been mentioned that AGE was selected first on 38 occasions. This occured in 10 out of 90 trials for Experts group (11%) and on 28 out of 126 trials for Proficient group - representing a significant difference (x^2=5.26, df1, p<.05). Of the 5 other attributes selected first on some occasions, no one attribute achieved prominence.

A very strong associate fo AGE appears to be SEX - on a total of 33 trials out of 216 these attributes were selected as initial pair with either AGE first then SEX (most commonly) or vice versa. Interestingly, the Expert group accounted for only 8 of these 33 pairings in comparison with the Proficient group (X^2 = 5.74, df1, p<.05). Taken together, MOBILITY and the AGE/SEX pairings account for 171 (79%) of all trials.

The pattern of individual subjects header node preference is interesting. The relative 'loyalty' of subjects can be summarised thus:

	Experts	Proficients
- loyal to one attribute on every trial	4	1
- 'fairly' loyal (single attribute selected on at least 12 out of 18 trials)	1	2
- little discernable loyalty pattern	0	4
	5	7

The picture which emerges, then, is that the Expert group were more 'sure' in their approach to eliciting information on each trial. Another way to describe this is that Expert information processing shows signs of greater automatisation. Following from this it becomes interesting to speculate on what may have accounted for the deviations observed in the 'fairly loyal' subjects - why should a different header be selected on few trials only? One possible explanation is that this 'shift' pattern is an artefact of the experiment; perhaps boredom or fatigue lead to clicking the nearest or random button.

This explanation would be weakened if it could be establshed that the attributes to which subjects shifted were 'important' attributes in terms of predicting pressure sore risk. By the end of this section the relative importance of the 12 attributes will be established - by anticipating this result, however, it is possible to state that 4 out of the 5 attributes which were selected on the 12 trials in question were indeed 'important' in that they came from the 'top six'. That is, they were seen by nurses to be attributes which are important for assessing pressure sore risk.

An alternative explanation is that the trial immediately prior to the trial when the deviation occured contained a 'vivid' attribute value which had strong influence as the risk decision. This explanation would hold that the salience of this attribute would serve to lead to its selection as header when the next patient was presented. As the literature review established earlier, salience of cues has been established as having an important influence on cognition. This explanation holds good for the single Expert subject - on each of 3 occasions when she deviated the prior patient had informative attribute values which exactly corresponded to the attribute which she choose to begin the next trial. Two of these three values were rare and therefore even more salient, for example on one occasion she selected MENTAL STATE as a first attribute - the previous patient she assessed had been unconscious.

However before suggesting tentative theory for further investigation - along the lines of learning mediated by salient exemplars - it becomes clear that the effect is less apparent for the two Proficient subjects. Thus each of these subjects deviated on four occasions, but of the 8 prior patients to these occasions only 3 had informative values on the attributes which headed the trials on which they deviated. This finding, along with the generally low number of trials in question, makes any conclusion about shifted header choice other than the blanket term 'error' seem to be untenable.

Header attribute choice, in conclusion, is reliably MOBILITY for the Expert group - implying a desire to be informed of the key attribute in pressure sore development. As such this represents heuristic search - searching for information most likely to lead to goal attainment. What the nature of this goal is (and the nature of the representation being consructed) will be further explored below. Support for the general assertion, however, comes from some examples of verbalised statements given by subjects to explain header attribute choice......

> "mobility's obviously the most important thing" (Expert(E) subject 2)
> "really it's got to be mobility" (E 1)

"again I'll start off with mobility because of the problem of unrelieved pressure" (E 5)

On the occasions when factors other than MOBILITY was selected first, AGE was predominant - particularly in the Proficient group. When AGE was selected there was a strong association with SEX as a 'paired' attribute - almost exclusively a Proficient group style of beginning an assessment. Knowledge of the ward situation suggests an explanation in that the routine form of formal representation of patients is in the form... "Mrs Smith (48), admitted for investigations....". It seems clear, that the Proficient nurses are routinely beginning their assessments in this fashion, even though a patient's gender has clearly no bearing on their risk of developing pressure sores. As such, the representation being constructed is superficial. Expert nurses, on the other hand, seem to have eschewed the routinised for the informative.

Following from the literature review on attribute importance, the requirement now will be to establish the importance value of the attribute MOBILITY. If it can be shown that header attribute choice conforms to a reliable scheme of attribute importance then a suggestion can be made about the theoretical import of this finding. This suggestion would concern the form of the mental representation which the nurses are constructing of the patient they are assessing. The difference between Expert nurses and Proficient nurses seems to be that the informativeness of the attribute which heads this representation is crucial - a finding which carries clear analogies to 'reduction of uncertainty'.

Subsequent attribute choice

Having explored the data with respect to header attribute choice, the next logical step would be to proceed to identification of patterns of choice throughout the remaining information gathering process. The explanation being tested currently - that attributes vary in terms of importance - requires that subjects' process traces are tested against some sort of scheme or schemes which reflect this importance. Prior to this more specific testing, however, it is useful to continue to explore the usefulness and applicability of attribute importance as an explanation. For subsequent attribute choice, therefore, a continued exploration of the data should be undertaken in order to look for signs of nurses using some sort of knowledge based scheme of attribute importance.

At this point it would be helpful to look to the verbalisation data for evidence of a scheme underlying subjects' move from attribute to attribute. The verbalisation data, however, was not the primary focus of this experiment - the requirement rather was for process traces of several subjects assessing a comparitively large number of patients. The repetitive and lengthy nature of the task, as well as the focus on quantitative data, precluded the rigorous protocol analysis methodology. The nurses who were subjects were at work at the time and the experiment took on average 1 hour. Nevertheless, it was decided to ask nurses to 'think out loud' as much as they could.

The transcripts which were collected, however, were often extremely elucidating, particularly when subjects seemed to be going beyond the data (see below). Nevertheless, there were several occasions when a silence came before or after an attribute selection. Alternatively, a subject would simply state the name of the next attribute to be selected. It becomes interesting to consider these 'quiet events' in terms of their implications for the analysis, since it is possible that they may have explanations other than simply 'nothing to say'. Perhaps, for example, 'quiet events' were instances of cognitive processing switching from one 'active' mode to another more 'automatic' mode. From higher-level to low-level cognition. Schneider & Shiffrin (1977) distinguish two general modes of cognitive processing....

Mode 1 a type of processing involving parallel processing which is automatic, less capacity-limited and invoked directly by stimulus input (bottom-up processing)

Mode 2 a more serial type of processing which requires conscious control, has severe capacity limitations, and is invoked in response to internal goals (top-down processing).

Based on the rationale that Mode 1 processing will not lend itself to illuminating verbalisation to the degree that Mode 2 processing will, a working hypothesis can be set out....

> silences or occasions when a subject simply states the name of the next attribute to be selected are instances of cognitive processing which involves 'default to next-most important attribute'.

In short, a knowledge based scheme of attribute importance may be an example of Mode 1 processing. An example transcript (Expert #4) will serve to clarify the foregoing. Line numbers in the transcript (see overleaf) are designated by L (number) and "click" refers to a button being pushed and, therefore, an attribute being selected.

L 1		mobility first....
L 2 -	click....	bed chairfast with assisted walks .... mental state to see whether they are confused
L 3	click....	alert and orientated .... nutritional state
L 4	click....	nutritional state seems adequate. Check out whether they are continent or not
L 5	click....	continent if supplied with a commode, I suppose I will check faecal incontinence at the same time
L 6	click....	full bowel control .... go for circulation
L 7	click....	it's poor .... what age is this patient?
L 8	click....	50-69 .... build
L 9	click....	slightly overweight
L10	click....	female patient

This segment of a transcript demonstrates both the variable nature of the transcript data as well as the points at which 'quiet events' occur. Some sort of Mode 2 processing seems to be underway at least at L2 and L5 when the nurse seems to be actively setting out to search a particular attribute - more on this apparently 'higher' cognition later. Using similarly strict criteria for identifying 'quiet events', these were taken as occuring at Ls 1, 3,6,8, and 9. The question becomes one of testing to see if it could be possible that this subject was refering in a comparatively automatic fashion to an internal 'list' of next-most important attribute for reducing her uncertainty of this patient's risk of developing pressure sores.

Heuristic search, as cognitive expertise, might be stated as the selective search of the information which examines those parts of the problem space that are most likely to produce a solution. - The task now becomes one of more specific testing of the data in order to establish firstly the existence and secondly the form of this hypothesised 'list' of attributes set out in rank order of importance.

Specific Testing of Attribute Importance as an Explanation of the Data

The preliminary analysis established that header attribute choice varied significantly between nurses identified as 'expert' and those identified as 'proficient'. In the exploration of subsequent attribute choice the suggestion was made that nurses 'default' to a list of attributes to assess when no specific attribute is suggested as being the next one to search. Furthermore it was suggested that these attributes are ranked in terms of importance or 'most likely to furnish solution to the problem'. The processing control implications, it follows, are that nurses will default to the first attribute on this ranked list which is not yet *directly or indirectly* known. 'Directly' refers to an attribute previously searched for a particular patient. 'Indirectly' refers to an attribute for which the value has been assumed. Two analyses can now be undertaken to test these suggestions.

Goodness of fit of ranked attributes to the data

To test these suggestions against the data it is necessary to firstly set out a list or lists of attributes ranked in the order of usefulness to the assessment task. Deciding on the ranked order is not, however, straightforward since several schemes are suggested. Two of these schemes refer to the initial data collection exercise in this project - the Knowledge Elicitation phase. Hence, the first approach to establishing lists would be to look at the attribute listing tasks. Here, however, it was found that some differences emerged between the lists compiled by the two groups of subjects. The suggestion made is that the 'more theoretical' nurses could have been listing attributes in terms of power to predict pressure sores while the 'more practical' nurses might have been responding to the question by listing attributes they would assess when wishing to plan pressure sore preventive care. This subtle point will be developed more fully in later sections.

A third list which would be a contender is ranked order of attribute importance within a mathematical model which has been constructed. This model, fully explained and reported in Chapter 4 (Part 1), was constructed using the a Discriminant Function Analysis (DFA) procedure. A fourth list, finally, would be suggested by the desire to have some kind of baseline or 'chance' list from which comparisons might be drawn.

Four ranking schemes, or lists, can be considered in this specific testing of evidence for attribute importance within the data. These lists are:

1. 'Chance' list - it will assumed that attributes are ranked in an order which has been derived from chance. This list was constructed by sampling with non-replacement from a collection of cards labelled with each attribute. First card picked out became the 'most important' attribute, second card became rank=2 most important and so on.

2. 'DFA' list - the attributes in this list were ranked in the order in which they were selected during the stepwise analysis performed on the patients comprising database2. As explained in Chapter 4 (p.209), the selection algorithm used was based on minimisation of Wilk's Lambda. In the terms set out in the introduction, this list would conform to a 'mathematical' prescriptive scheme of 'ranked order of cue utility'.

3. 'Predictive' list – the rank order in this case was decided by the frequency with which an attribute was listed by nurses set the task "List the factors which are important in predicting whether a patient will develop pressure sores". In the terms set out in the introduction, this list would conform to a 'human' prescriptive scheme of 'rank order of cue utility'.

4. 'Assessment' list – as for list 3 except that the frequency of listing was derived from the 'more practical' (ward based) nurses. In the terms set out in the introduction, this list would conform to a deeper level representation where nurses were 'looking ahead' in their assessment to the ultimate goal of planning care.

The rank orders corresponding to each list are as given in Table 3.1

Table 3.1 Four different prescriptions of rank order of attribute importance

	Ranking Scheme			
attribute	Chance	DFA	Predictive	Assessment
MOBILITY	8	4	2	1
URINARY CONTINENCE	12	6	1	2
SKINTYPE	9	8	4	8
BUILD	11	7	3	9
NUTRITION	1	2	5	4
AGE	6	10	6	5
MENTAL STATE	3	5	7	3
CIRCULATION	10	3	9	6
LIFT & TURN	2	1	8	10
SEX	7	9	10	11
BLOOD PRESSURE	5	12	11	12
FAECAL CONTINENCE	4	11	12	7

With contender schemes of attribute importance now established it is possible to calculate the point-by-point agreement indices for each subject with each list. The procedure, handled by a short ad-hoc computer program prepared by the author, was to compare the cells in the 18 x 12 matrices formed for each subject with the corresponding cells in a matrix formed from each list. In this way it is being assumed that, for example, the DFA will always search the MOBILITY attribute in the 4th position for each patient. Over all 18 patients, a count is made of the frequency of a '4' being against MOBILITY for a particular subject. This count is incremented each time an exact match is discovered over all 18 patients and 12 attributes – giving a maximum possible 'predicted search position' index of 216.

Results, given in Table 3.2 overleaf, display the number of exact matches for each subject and each list (in columns labelled 'n') In addition, the number of attributes explained is expressed as a proportion of the maximum agreement (in columns labelled P).

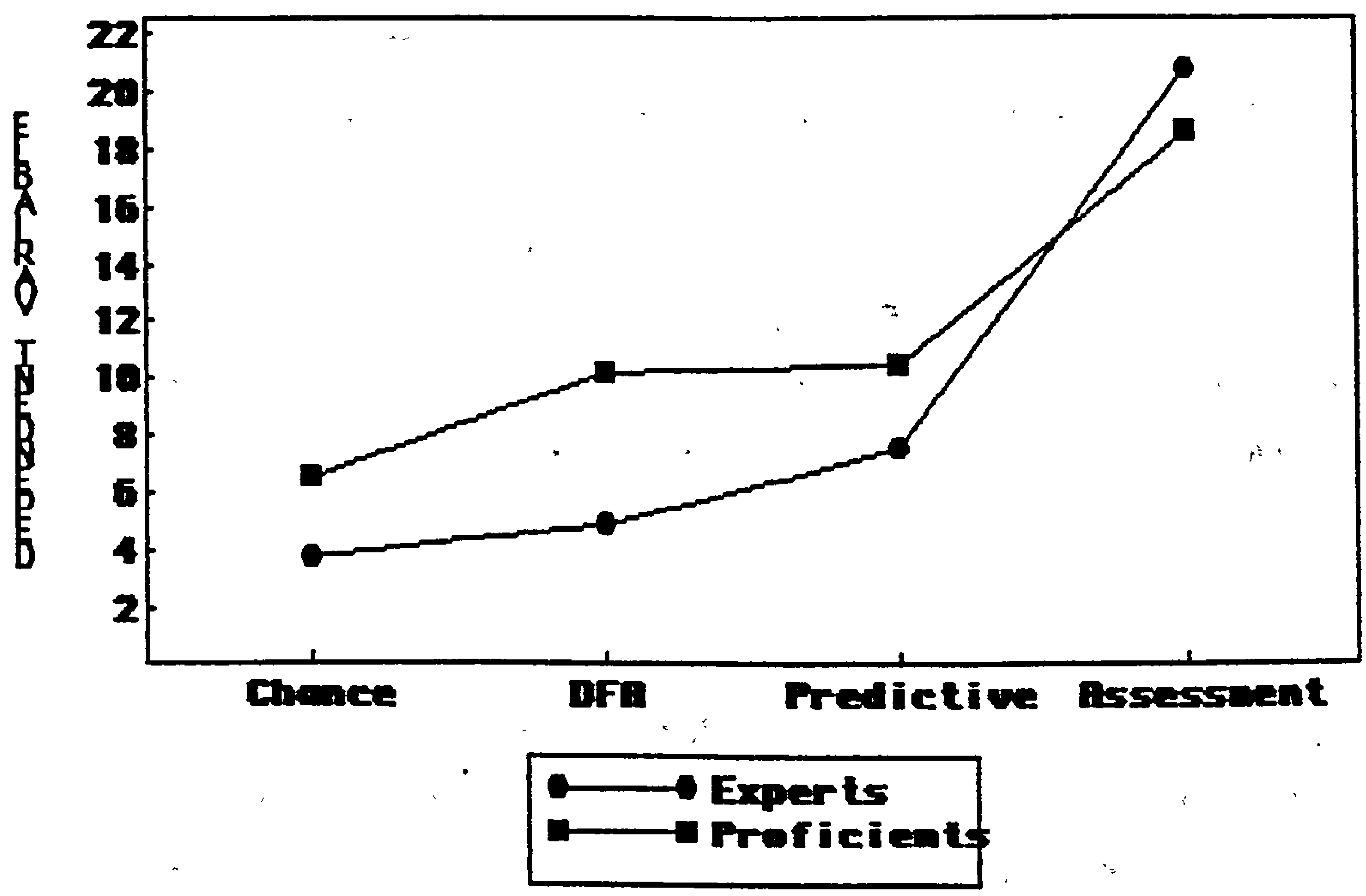
Table 3.2 Point-by-point indices of agreement between each subject and each of 4 ranking schemes of attribute importance

subject	total attribs selected	Chance		DFA		Predictive		Assessment	
		n	P	n	P	n	P	n	P
E1	71	2	.009	5	.02	10	.05	7	.03
E2	68	5	.02	2	.009	5	.02	27	.12
E3	92	2	.009	8	.04	12	.05	28	.13
E4	78	5	.02	3	.01	2	.009	34	.16
E5	74	4	.02	6	.03	8	.04	7	.03
P1	89	1	.004	6	.03	3	.01	21	.10
P2	103	1	.004	16	.07	10	.05	9	.04
P3	131	4	.02	3	.01	15	.07	15	.07
P4	90	9	.04	10	.05	10	.05	15	.07
P5	167	18	.08	20	.09	14	.06	19	.09
P6	72	2	.009	7	.03	10	.05	17	.08
P7	105	10	.05	8	.04	10	.05	33	.15

(Ranking Scheme: Chance, DFA, Predictive, Assessment)

The index is informative only as it varies within each subject, since it takes no account of the total number of attributes selected by each nurse. Every subject had varying amounts of cells in their matrices with a zero to denote unselected attribute while the ranking lists had an integer in every cell. It would be expected, therefore, that the nurses who selected the most attributes would have a higher probability of a matching cell. For descriptive purposes only, therefore, the average number of cells explained for each group are displayed in Figure 3.1 overleaf.

Figure 3.1 Average number of matrix cells explained by 4 Contenders



The problem of varying numbers of attributes selected is avoided by a within subjects analysis, with the dependent measure as number of cells matching (range 0 to 216). Separate ANOVA were performed on the Expert subjects and on the Proficient subjects scores. The design was wholly within subjects, with 4 levels of the within subjects variable - each corresponding to one of the schemes of attribute importance. Source tables of results are given in Tables 3.3 and 3.4

Table 3.3 ANOVA of proportions matched by each of 4 attribute ranking schemes for Expert subjects

Source of Variation	df	SS	MS	F	p
subjects	4	133.3	33.3		
ranking scheme	3	919.4	306.5	6.06	<.01
error	12	607.1	50.6		

Table 3.4 ANOVA of proportions matched by each of 4 attribute ranking schemes for Proficient subjects

Source of Variation	df	SS	MS	F	p
subjects	6	338.7	56.4		
ranking scheme	3	540.8	180.3	6.05	<.01
error	18	536.1	29.8		

Further testing - Expert subjects:

Mean values: Chance= 3.6 DFA= 4.8 Predictive= 7.4 Assessment= 20.6

Since the main effect for ranking scheme was significant ($p > .01$), post-hoc comparison of means was undertaken using Tukey's HSD. Results showed that the only significant differences between means were Assessment vs. Chance ($p < .05$) and Assessment vs. DFA ($p < .05$).

Further testing - Proficients subjects:

Mean values: Chance= 6.4 DFA= 10.0 Predictive= 10.3 Assessment= 18.4

Since main effect for ranking scheme was significant ($p > .01$), post-hoc Tukey's HSD suggested significant differences between means; Assessment vs. Chance ($p < .01$), Assessment vs. Predictive ($p < .05$).

The main conclusion from these analyses is that the Assessment scheme of attribute importance seems to explain the information processing of subjects beyond a level expected by chance. This conclusion is strengthened by the finding that the Assessment list concurs with the earlier analysis of header attibute choice - MOBILITY is the attribute first chosen by both subjects and this ranking scheme. A final specific analysis can now be undertaken to ascertain if 'quiet events' support these conclusions.

Analysis of 'quiet events'

The suggestion made earlier was that quiet events found in the transcripts might be indications of Mode 2 processing when subjects are defaulting to a list of attribute importance. To test this suggestion, an analysis will be performed only for Expert subjects since on this occasion it is the description and explanation of expert cognition which is important rather than the comparative exercise. Since the Expert group's performance was seen to most closely correspond to the Assessment list, this will be the scheme used for the analysis, although the 'second best' list (Predictive) will be retained as a comparison 'default list'.

The procedure for analysis involved the following steps:

1. identify in the transcripts the 'quiet events' ie. the silences / simple statements of next attribute to be selected. Do for n=5 Experts each over n=18 Select trials.
2. for each instance, establish which attribute would be selected next if a ranked list was being automatically referred to. Do for Assessment and Predictive lists.
3. compare the attribute predicted by each list to the actual attribute next selected. Do for each 'quiet event'.

Steps 1 to 3 above were carried out with results as set out in Table 3.5. The number of quiet events identified for each Expert subject is listed against the number and percentage of these events which lead to agreement on next attribute selected by either Predictive or Assessment ranking scheme.

Table 3.5 Numbers of successfully predicted attribute selections following quiet events

Expert	n 'quiet events'	PREDICTIVE LIST n successful	%	ASSESSMENT LIST n successful	%
1	47	22	47%	31	66%
2	25	19	76	20	80
3	26	19	73	20	88
4	32	22	68	24.	75
5	36	23	64	26	72
total	166	105	63.2%	124	74.7%

Each list was found to have predictive power which was impressive when it is considered that the range of possible attributes which could have been selected on these occasions was vast - with the exception of first selected attributes, if the average situation is taken as being a 'quiet event' occuring at about the third attribute point then the range of possible selections is 9. Therefore the probability of the same attribute being chosen by the subject as is chosen by the list is 1/9 or .11. The probability of successful prediction by chance on 124 instances becomes becomes $.11^{124}$ Clearly the level of predictive accuracy - especially of the Assessment List - is considerably beyond that expected by chance.

No further testing of this result has been undertaken, partly due to the problem of non-independence of data points. The important point to be taken, however, is that the analysis of quiet events has produced findings which are concordant with the general conclusion - that attribute importance can explain a degree of the observed information processing, and that the Assessment list seems to offer the most powerful explanation.

Implications of the Findings on Attribute Importance for the Cognitive Model

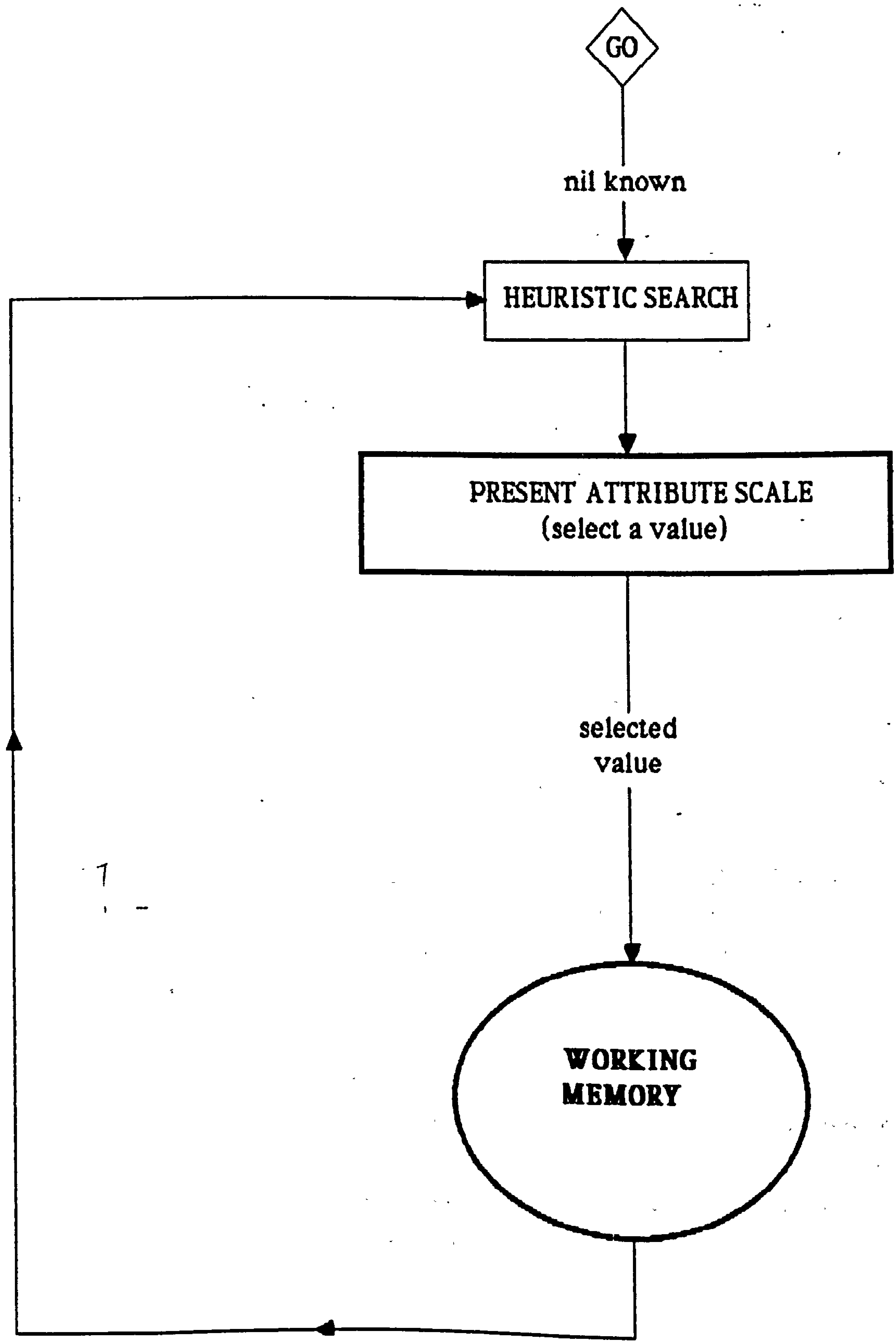
Firstly, it is useful to sum up the lines of evidence which support what might be termed the 'Principle of Heuristic Search' - cognitive expertise contains a component of selective search of that information which examines those parts of the problem space that are most likely to produce a solution. The interesting finding is that the solution cannot be simply taken as an answer to the problem of uncertainty of risk. 'Care planning' styles seem evident, even though the nurses were not asked to plan care. These lines of evidence which support this conclusion are:

- header attribute choice conforms reliably both to attribute most likely to produce a solution to the risk assessment problem and to attribute most implicated by care planning. Pressure is the principal cause of pressure sores; the most potent tool in the nurses' armoury for preventing pressure sores is to relieve pressure.
- non-conscious processing is more likely to result in 'quiet' verbalisations than conscious goal-directed processing.
- 'quiet events' in the verbal transcripts are most adequately explained by a mode of processing which calls up the attribute value next-most important in planning the preventive nursing care of the at-risk patient.
- the scheme of attribute importance which explains the data most powerfully is one which is based on the numbers of nurses mentioning an attribute in a free-listing task of patient assessment priorities.

The first implication for the cognitive model is that an underlying scheme of attribute importance should be adopted and that this rank order list should be that provided by the Assessment List. The second main implication following from this relates to the flow of control within the cognitive model. Put another way, the manner in which the Expert nurse (or to be more precise, the emulation of expert nursing cogntion) utilises her mental representations of attribute importance. Header attribute choice, firstly, shows that when nil is known then the most important attribute should be searched. The only secure assumption which can be made is that the information gained from searching the attributes will be held within working memory.

A model constructed on the basis of the explanation thusfar would be as illustrated in Figure 3.2 overleaf. The rectangular box in the centre of this flow diagram represents the point at which the attribute values are presented on the screen in order that the one most appropriate for the particular patient can be chosen. This loosely corresponds to the expert responding to a request to assess a patient's pressure sore risk by asking for descriptive information of that patient.

Figure 3.2 Flow diagram of cognitive model based on heuristic search



Another major implication arising from viewing this model and the results of analysis is that the search for explanations of the data cannot yet stop – the highest level of prediction achieved by the Assessment list only matched 34 out of the 216 cells of Expert 4 (16%). Although it might be pointed out that this Expert subject actually selected only 75 attributes and that the match represents 34% of this figure, this strategy is self-defeating and takes no account of the importance of all cells in the matrices. If an attribute is left unsearched then a complete explanation must account for this.

As will be more fully explored in the next section, however, information processing after this point becomes highly contingent on the particular attribute values elicited. Nevertheless, the second suggestion for the cognitive model, which arises from the exploration of 'quiet events', is that the stored scheme of attribute importance will only be refered to subsequently if there is no directed search. What is being tentatively suggested, then, is that nurses alternate between Mode 1 and Mode 2 processing – when no higher cognition is taking place and the nurse does not feel directed to the next attribute to be searched (as evidenced by a quiet event) then that nurse might be falling back on to a default list of attribute importance.

More specific testing of these suggestions will be undertaken in subsequent sections, however before closing the discussion on attribute importance it is useful to offer some tentative conclusions on the relative performance of the three ranking schemes. The most clear finding was that nurses' search of the problem space does not seem to conform closely to the prescription offered by a mathematical linear model – the DFA list. It cannot, however, be concluded that nurses are therefore ignorant of the predictive power of attributes. The question rather becomes one of searching for an understanding of what was the basis to the order of importance which nurses gave to the attributes. To accomplish this, it is initially more useful to look at the Assessment ranking scheme which clearly outperformed the DFA list and to compare this list to the Predictive scheme.

At first sight, that the Predictive list explained less of the data than the Assessment is counterintuitive. This list, after all, was derived from nurses asked to note down factors which affect pressure sore risk and the experimental task was similarly one of prediction of pressure sore risk. The Assessment list, on the other hand, held apparently less potential for explaining the data since it was derived from a question about planning pressure sore preventive care. A closer look at the form of the two lists becomes illuminating.

The Predictive list tends to overvalue (by comparison) SKINTYPE and BUILD, while the Assessment list places more importance on NUTRITION and MENTAL STATE. This seems to powerfully illustrate the apparent 'care orientation' of the Expert subjects. MENTAL STATE, to illustrate, is an attribute which may not alone be of unambiguous power for the purpose of predicting pressure sore risk. It is, as the earlier review established, an attribute which nevertheless strongly implicates the plan of preventive care most appropriate for a particular patient. The nurses in the process tracing experiment, although ostensibly 'only' assessing risk, clearly felt it crucial to elicit at an early point the extent of the patient's *capacity tc relieve pressure themselves*, or, more simply, their dependence and self-care potential.

As a factor, MENTAL STATE may not possess the predictive power of SKINTYPE or BUILD; but for accurate care planning it becomes essential. The discussion must at this point be reminded of the rationale for pressure sore risk prediction - so that pressure sores may be prevented. Since pressure sores are prevented through the planning and application of optimal preventive care, it therefore becomes understandable that asking nurses to predict pressure sore risk results in information processing which owes more to care planning than to risk assessment. In the day-to-day situation nurses assess patients in order to plan care. The nurses in this experiment might simply be employing their day-to-day styles of searching the problem space.

The explanation which is suggested for the superior performance of the Assessment list, therefore, is that this list most closely corresponds to the care planning orientation of the Expert subjects. What is being suggested is that the emphasis on diagnostic cues might be misguided. This may even be a valid idea in medical studies - perhaps Doctors, like nurses, collect information with an eye to treatment. A headache may not be an powerful cue for diagnosing heart trouble, but it would be a poor clinician who ignored it. To the extent that diagnosis and treatment cannot be assumed to be separate entities, it is hard to understand why the seminal studies (eg Elstein et al 1978) have concentrated only on asking clinicians to diagnose. This suggestion, henceforth to be known as 'the careplanning hypothesis', will be more fully explored and evaluated in many of the subsequent sections.

USE OF HIGHER COGNITION

The finding which clearly emerged from the earlier analysis was that routes to decision varied both across patients and across subjects.. However, the nurses, both Expert and Proficient, were apparently not collecting information in a random fashion since the evidence showed some correspondence with a systematic approach based on attribute importance. Nevertheless this correspondence was not complete. Therefore in an attempt to more fully explain the cognition of the nurses it might be useful to focus now on information processing which proceeds in the opposite direction to the incoming (or bottom-up) data, ie the use of high level knowledge in the generation of expectations or hypotheses relating to the interpretation of incoming patient information.

Selected Literature Review on Higher Cognition in Expert Decision Making

Interest in the use of higher cognition by nurses can be traced back almost 100 years to Nightingale. It is worth restating her observation which captures the essence of the distinction between top-down and bottom-up processing .....

> "Observation tells us the fact, reflection tells us the meaning of the fact ... observation tells us how the patient is, reflection tells us what is to be done" (p.255)

Reflection, for Nightingale, seems to denote cognitive activity which conscious and deliberate rather than automatic. Clearly also this pioneering nurse saw a strong need for teaching of cognitive nursing skills ....

> "Training and experience are, of course, necessary to teach us how to observe, what to observe, how to think, what to think"(p.254).

Nightingale's aim in these examples was to somehow shake nurses out of traditionalised and automated modes of working. Interestingly, J. Jones (1986) recently studied junior and senior nurses' styles of assessing pressure sore risk and concluded that cognitive activity was minimal and highly routinised. This is seen as highly undesirable, consequently the increasing concern of nursing authors has been to develop nursing models which increasingly emphasise higher cognitive systematic approach.

Many of these authors (eg Roper, Logan and Tierney 1985) suggest a conceptual model of human functioning which interrelates nursing to achievement of optimal functioning. As such these models, often developed with beginner nurses in mind, can be seen as helpfully providing deep level knowledge which the nurse can use to mentally represent her patients. More worrying trends, however, are the increasing advocacy not only of more superficial representations of patients (as protypical diagnosic types) but also of 'even higher' cognitive processes. The validity of these trends, it will be argued, is open to question.

The focus on the use of 'deliberate' higher cognition when assessing a patient can be seen as a 'super-systematic' response to potential information chaos. Hence Gordon (1987) and Carnevali (1983) prescribe specific models of nursing cognition which go to extreme lengths to avoid the danger of nurses becoming overwhelmed by volumes of unsystematically collected patient information. The volume of information which is the proper concern of the nurse has increased with the advent of theoretical models which individualise patients in terms not only of task characteristics. Gordon (1987) goes so far as to write of the need to consciously 'chunk' cues and to deliberately organise information. The move toward what Risner (1986) terms 'deliberation' has also affected advocacy of styles of cognition. In Hammond's (1966) terms, logical and inductive inference (largely top-down) are held to be ideal while intuitive inference or the making of assumptions (largely bottom-up) is to be avoided (Lane et al 1983). Although, as discussed below, there is a growing movement to re-emphasise the role of intuition in nursing expertise (eg Benner 1984).

Hypothesis testing, introduced in the section on attribute importance, can be seen as the key to this trend. Early incoming information which is diagnostic is to be actively noticed by the nurse who then consciously activates hypotheses and goes on to employ maxims or rules which serve to guide subsequent search of the data in order to decide between these hypotheses. Although Gordon (1987) acknowledges that the 'deliberate' component to this process will be increasingly replaced by automatic processing as the nurse becomes more expert, there are nevertheless empirical and theoretical reasons for doubting the usefulness of this prescription for learners. Empirical reasons, reviewed previously and in Chapter 4, are based on the medical and nursing evidence for alternative modes of processing. Theoretical reasons relate to the unworkable demand on human cognitive capacity, as can now be discussed.

Certainly it is possible to 'switch' into a purely top-down processing mode, but there are significant costs in terms of demand on working memory of using this strategy (Schneider and Shiffrin (1977). Following Carnevali's (1983) prescription, the nurse holds activated in working memory an apparently staggering volume of patient cues, conceptual knowledge such as difficulties in daily living and functional health status, and a potentially large number of lists of features known as diagnosic hypotheses. This seems an extraordinary prescription for avoiding information overload in working memory, a point not lost on Corcoran (1986). Moreover, the cognitive processes of chunking and organisation of data are elsewhere accepted as automatic rather than deliberate (eg Chi et al (1981).

Leaving aside working memory limitations, the top-down prescription for running an assessment can be seen to rest on the assumptions of largely superficial representations of knowledge in large packets and predominantly backward search when evaluating multiple hypotheses. As such this denotes in clinical reasoning what Pople (1973) terms abductive inference. Hammond (1966) suggested that this mode of inference represents the 'ideal' in that multiple hypotheses can be entertained simultaneously while the data field is searched for cues which discriminate between competitors. Barrows and Tamblyn (1976) provide some evidence that experienced physicians hold up three to five hypotheses simultaneously and Pople (1977) reports some success (and many problems) in building an expert system which uses abductive logic.

Nevertheless, the applicability of abductive logic to nursing is far from conclusive. Medical knowledge, for example, of diagnostic pathology is considerably more highly specified than the relatively-infant nursing diagnostic concepts. Until there is some resolution of the problems which Kritek (1988) has outlined in relation to these categorisations there can be no serious models of expert nursing based on this technique - particularly in the light of respected work (eg Benner 1984) which argues for alternative modes of inference which see a place for mechanisms whereby missing information can be assumed on the basis of experience. Induction, for example, is the process of using knowledge structures compiled from particular cases to the general case. Interestingly, Hammond (1966) actually recommends this more sober course where nurses use inductive inference.

In terms of the earlier discussion, a recognition that nurses assume unknown information immediately becomes attractive in that the idea fits with the principle of cognitive economy - assumptions save working memory capacity. It is significant that when the medical expert system MYCIN was reconfigured for pedagogical purposes into GUIDON (Clancy 1983) there were efforts made to incorporate implicit knowledge in the rule structure - a key point given the claim that the system was said to be a psychological model of diagnostic behaviour. A further point can be made regarding this system in that the reasoning is not solely of the backward-search hypothesis testing variety. Incoming information acts to trigger smaller units of knowledge than 'diagnostic hypotheses'. The function of these rules is to direct information search. This 'forward reasoning' strategy stands in sharp contrast to the predominant nursing prescriptions but nevertheless finds support from some recent psychological studies of medical diagnosis (Patel and Groen 1986).

The picture which is emerging, therefore, is that the the ability to 'go beyond the data' is a principle of nursing cognition which deserves to explored in relation to the current project. Two main forms of inference have been suggested. Firstly, that information 'suggests' a hypothesis or perhaps more simply an item of information worth eliciting. Secondly, that information can act to permit the nurse to 'assume' knowledge which is implicit in known information. One further observation which can be made is that the representation issue introduced in the section on attribute importance can be seen as central to inference in that the suggestions made above each make separate predictions about whether superficial or deeper level representations underlie internal models of patients.

Exploration of Higher Cognition as an Explanation of the Data

In this section the analysis will shift beyond header attributes and concentrate on the tracing of information processing through the data which ultimately results in the nurse making an assessment decision. The focus, therefore, is on the process of assessing rather than assessment. Put more simply, attention is being paid to the points at which the nurse might be saying ..."where do I go from here if I am to achieve the goal of judging this patient's risk of developing pressure sores?". The position which has been established thusfar is that a single scheme of attribute importance can explain the data only to a point. The foregoing review, moreover, suggests that there may be, in Schneider and Shiffrin (1977) terms, 'mode 2' explanations which can add to the achievement of the 'mode 1' explanation.

Broadly, the picture which emerges, especially from the transcripts, is one of interim hypotheses being generated - usually after the header attribute value was elicited. These hypotheses were then used to guide subsequent data collection. This 'inference' process will be analysed in more detail below; for the moment however the exploration remains at the more superficial level of looking at the pattern of which attributes follow from which attribute values. Three issues regarding higher cognition can be used as a framework to undertake this exploration. These issues relate to the goal-directed nature of cognition, the assumption of unknown information, and the use of hypothesis testing. Each issue can now be explored in turn.

Contingency and goal direction as a feature of nursing cognition

There are, apparently, signs of contingency and goal direction within the subject's process traces, by which is meant that the preceding attribute values collected can be seen on occasions to strongly influence the selection of subsequent attributes. A clear example of this contingent and directed processing can be seen in the attribute next selected following reading that a patient was ....

either MOBILITY - bedfast and virtually inert in bed (pts 2,4,6,11)

or MOBILITY - bedfast but can move freely in bed (pts 1,12).

In the Expert group, for 'bedfast and virtually immobile patients', on 17 out of a possible 25 occasions MOBILITY was selected first. Using x^2 one-sample test, the favouring of this attribute over any other is reliable (x^2=4.84, df1, p<.05). For the 'bedfast but can move freely in bed' patients, the header attribute of choice was exclusively MOBILITY (10 selections out of possible 10). Of interest, however, is the next attribute selected - as represented in the following table:

	attribute selected next	frequency
'bedfast and virtually immobile' patients	MENTAL STATE	13
	URINARY CONTINENCE	4
		17
'bedfast but can move freely in bed' patients	URINARY CONTINENCE	6
	MENTAL STATE	2
	NUTRITIONAL STATE	2
		10

Hence for 'immobile' patients,13 out of the 17 (76.5%) cases the attribute selected next was MENTAL STATE with URINARY CONTINENCE being preferred in the remaining cases. With 'bedfast and freely moving patients', however, on the 10 occasions when MOBILITY was selected first then the next attribute was URINARY CONTINENCE in 6 cases with NUTRITIONAL STATE and MENTAL STATE accounting for 2 each of the other cases.

Bear in mind that for the entire 7 patients there were 11 attributes remaining unselected for these 5 nurses - 385 possible selections were reduced to a range of 4 of which MENTAL STATE and URINARY CONTINENCE predominated. More interestingly, in the case of 'bedfast and inert' patients the consensus was to find out about MENTAL STATE while in the case of 'bedfast and freely moving patients' then URINARY CONTINENCE was preferred. Exploration, rather than specific testing is the focus of this section. Nevertheless it is interesting to test the reliability by looking at the frequency of choice of the attribute MENTAL STATE:

	MENTAL STATE selected	other attributes selected
'immobile' pts.	13	4
'move freely' pts.	2	8

The proportions observed differ from those expected by chance (x^2=8.13, df2, p<.05), the conclusion that the patients are being treated differently is therefore sound for the Expert subjects. Clearly the Expert group seem to practise contingent and directed processing in that the particular attribute values of these two types of patient have brought about significant differences. Later the nature of this processing can be more fully explored, however for the moment it will be illuminating to turn to the process traces of the Proficient subjects for a repeat of this analysis.

<u>In the Proficient group</u>, then, any pattern was less clearly discernible. Firstly, the attribute MOBILITY was not favoured so overwhelmingly as the header attribute on the 7x5=35 occasions when an 'immobile' patient was assessed - it was favoured on 12 occasions. However, for the 'moving freely' patients the number of possible header choices was 7x2=14 and MOBILITY was chosen x 12. Of interest, however, is the next attribute to be selected on these occasions:

	attribute selected next	frequency
'bedfast and virtually immobile' patients	SKINTYPE	5
	URINARY CONTINENCE	2
	NUTRITIONAL STATE	2
	FAECAL CONTINENCE	2
	CIRCULATORY STATE	1
		12
'bedfast but can move freely in bed' patients	SKINTYPE	5
	NUTRITIONAL STATE	3
	URINARY CONTINENCE	2
	BUILD	1
	LIFTING & TURNING	1
		12

Not only are more attributes generally selected by Proficient nurses, but more importantly the distinction between these types of patients does not lead to clear differences in what is next selected - SKINTYPE is the most popular selection in each case. Moreover, while for the Expert group a significant difference could be demonstrated between the types of patient in the popularity of the attribute MENTAL STATE, no such diferences exist within the Proficient group when looking at the attribute SKINTYPE.

The question of why the Experts reliably selected MENTAL STATE in contrast to the more varied information processing of Proficient group is of course crucial and will be discussed in more detail in the specification of expert higher cognition below. It can be noted at this point that the key difference seems to lie in the quality and degree of inference used by more expert nurses. For now, however, the suggestion which can made at this early point of the exploration of this first question arising from the literature review is that there does seem to be evidence that expert nursing information processing is strongly directed on a contingent basis from previously acquired information.

Assumption of unknown information as a feature of nursing cognition

Until now the focus of attention has been largely on the attributes which were selected rather than on the attributes which were left unselected. To say that these attributes were 'ignored' might, however, be to considerably misrepresent the cognitions of the subjects. The literature review earlier also provided frequent suggestions that information is not considered in isolation and that 'reasoning' will be a feature of expert information processing. The implications of the suggestions now being considered are that nurses are able to draw assumptions from certain attribute values about attributes as yet unseen. It is toward an exploration of these suggestions that the discussion now turns.

The implications of this type of higher cognition are that strong assumptions will act to permit nurses to economically leave unselected attributes when the values of these attributes can be reliably inferred. The key words here are 'strong' and 'reliably'. For now, however, two examples can be given which suggest some support for the existence of this mode of cognition. Some analysis can also be offered of possible differences in the use of this mode of higher cognition between the Expert and Proficient groups.

In the first example, one patient (patient 11) in the database had the attribute+value combination MENTAL STATE-unconscious. One reliable inference, guaranteed to be a sound conclusion, is that this patient's URINARY CONTINENCE will not be a problem since the patient (if unconscious for any length of time) will be catheterised. Four out of the five Expert nurses elicited that this patient was unconscious - none of these four subsequently searched the URINARY CONTINENCE attribute.

If the importance of URINARY CONTINENCE as an attribute is considered then this apparent omission cannot easily be explained other than in terms of an assumption being made about the highly probable attribute value. Additional support for this explanation comes from the verbalisations given by the Expert nurses where it was made clear that no moisture was going to be present. When the analysis turns to the more specific level it will be useful to include this kind of additional information, firstly the co-occurence of attribute values in 'real' patients and, secondly, the extent to which this mode 1 cognition is evident in the verbal transcripts.

The second example is more subtle and was actively searched for in the data in consequence to the earlier finding from the analysis of patient database1 that NUTRITIONAL STATE and BUILD were among the highest correlated of the attributes (contingency coefficient = .66). Experience with real-world co-occurence of these attributes should result in awareness of this correlation, ie experienced nurses would have this information mentally represented in some form. Since patients who are, for example, malnourished will also tend to have a thin build, the prediction can be made that these attributes could afford the opportunity for experienced nurses to make assumptions.

For the Expert group, there were 42 occasions when at least one of these 2 attributes were selected. On 5 of these occasions (12%), the subject at a subsequent point selected the second of the attributes. For the Proficient group the proportion was higher - out of 124 occasions both attributes were selected 31 times (25%). These relative proportions reflected a significant difference (x^2=3.98, df1, $p<0.05$ in two tailed test). The prediction that Expert nurses will make use of an opportunity for relatively secure inference seems supported, that is, Expert nurses were less likely to select both of these attributes than were Proficients.

Closer analysis reveals more interesting patterns. By focusing on the trials during which each of the BUILD and NUTRITIONAL STATE attributes are selected, it is possible to examine the conditions which governed the move to select the second attribute of this pair. A matrix of six patient values existed within the 18 patients, as illustrated below (figures represent frequency of co-occurrence of each value in the patients, note that the correlation is not perfect here as in real world).

	Build signif. underweight	Build slight	Build average	Build overweight
NUTRITION -ok	0	6	3	5
NUTRITION -poor	2	2	1	1

It is possible to use this matrix to analyse information search. It becomes evident that the strength of association between these attributes varies as a function of their particular values. For instance, if a patient's NUTRITIONAL STATE is 'evidence of protein/vitamin deficiency' then a stronger assumption can be made to infer that BUILD will have an 'underweight' value than if in the case when NUTRITIONAL STATE is 'adequate'. Taking a model of information as 'reduction of uncertainty' then it might be predicted that firstly eliciting an informative value from the fringes of the matrix would permit the strongest assumption to be made about the co-occuring attribute. Hence BUILD - 'overweight' will strongly predict that NUTRITIONAL STATE will not be a problem.

Finding out that NUTRITIONAL STATE was 'adequate' is not, however, strongly informative. Interestingly, all of the occasions when an Expert selected both attributes within the same patient fell into this category. That is, these nurses are seeking to assess (as evident from the transcripts) the 'boniness' and the susceptibility of the patients to develop sores due to nutritional factors - one attribute (NUTRITIONAL STATE) can provide both items of information. If it does not then these nurses proceed to BUILD.

Put in terms of condition-action pairs (or 'rules'), a nurse's experience with patients has resulted in attribute values which strongly correlate becoming linked into IF.....THEN antecedents and consequents. The suggestion from this analysis, therefore, is that expert nursing cognition will feature the assumption of consequents given strong evidence of antecedents. Expert nurses will make use of an inference when appropriate with the result that useful information does not have to be explicitly searched. The picture which has emerged from this initial exploration of the data is that there is sufficient support for the existence of assumptions to warrant a later more specific testing of this hypothesis within the data.

Hypothesis generation as a feature of nursing cognition

The final question arising from the literature review which can be explored relates to the existence or otherwise of evidence that nurses generate hypotheses in the course of patient assessment. 'Hypothesis generation' in the nursing diagnosis use of the term cannot be a feature of the present experiment since the nurses were not being asked to assess the patient in order to decide on a diagnosis. Moreover, such an experiment would be open to criticism that demand characteristics would affect results. The 'hypotheses' which could, however, be a feature of these nurses' cognition is of a more specified and interim nature. Hence it is possible that the attribute values elicited up to that point may 'cue' a direction for subsequent search, as in .... 'this looks like x, I'd better just check on x'.

Reference was made in the discussion on inference to the case when a nurse elicits an attribute which may possess the potential for making an assumption about subsequent data. It was pointed out that on some occasions this potential is not realised due to an uninformative attribute value – the result being that the nurse is required to proceed to direct selection of that subsequent data. Seeking direct confirmation was said to occur also in the case of information only weakly suggestive of subsequent data. This, then, begins to resemble the type of inference which the earlier review termed 'abduction' and seems a plausible avenue for exploration in the present experiment.

An example can once again be used to explore and illustrate. On eliciting that a particular patient has a 'urodome' fitted, the reaction of Expert 1 is ..."that sends alarm bells ringing, I'll just check his mental state." Two inferences are made here, that the patient is a male and that urodomes can be pulled off by confused patients. Expert 4 makes the issue clearer...

> "...a urodome fitted...... I'll need to check to see if he's confused otherwise the urodome will be useless for preventing incontinence."
>
> NOTE: a urodome is a condom with a tube leading to a urine collection bag

The first inference, that the patient is male, is an unremarkable yet concrete example of an assumption resulting in no need to check the SEX attribute. The second inference, that disorientated patients often pull off urodomes, is a creative abduction from a set of facts (patient confined to bed with urodome fitted ... may therefore be confused) which is not guaranteed to be sound yet which demands checking for veracity. Notice that the nurse is still seeking to assess the patient's 'moisture' dimension. Notice too the relevance to the careplanning hypotheisis in that the information processing is strongly dictated by the need to plan optimum care.

Unfortunately only one patient in the database had a urodome fitted. Analysis of attribute selection is therefore made difficult by the fact that only those nurses who selected URINARY CONTINENCE before MENTAL STATE can be examined. Nevertheless the results are interesting. Of the 2 Experts in this category, both proceeded immediately to MENTAL STATE from URINARY CONTINENCE. Of the five Proficient nurses who similarly fitted this category, however, none elicited MENTAL STATE immediately after finding out about the urodome.

More concrete results are potentially afforded by examining patients 1,2,5,6,8 and 17) – all of whom had one of the 3 'poor' values for the attribute SKINTYPE. Once again the condition for inclusion in the analysis was restricted to those trials when SKINTYPE was not selected after the attribute of interest. This attribute – URINARY CONTINENCE – denotes on this occasion the generation of a causal hypothesis which became evident from the transcripts. Hence some nurses who elicited that a patient had poor skin remark that this might be due to the irritant effects of moisture. The testing of this 'moisture hypothesis' was to immediately search the URINARY CONTINENCE attribute.

For the Expert group and with respect to this group of patients, there were 7 occasions when SKINTYPE was selected first and URINARY CONTINENCE at some point subsequently. On no fewer than 5 of these 7 occasions URINARY CONTINENCE was selected immediately after SKINTYPE. The total number of attributes available for selection across these 7 trials was 45. Moreover, the probability of 5 chance selections of URINE immediately after SKINTYPE is $= 6.9^{-6}$.

For the Proficient group the pattern is almost equally unequivocable since out of the 21 occasions when SKINTYPE was selected first, URINARY CONTINENCE was selected immediately after on 11 of these occasions when total available attributes was 102. This pattern of selection would have been observed by chance with a probability of 2.53^{-12}.

The relative proportions in each group of nurses (5 out of 7 versus 11 out of 21) were not significantly different. The point, however, is that there is a case for more specific testing to establish the presence of creative abduction. Broadly, the feature of expert nursing cognition which is suggested from the foregoing is that an attribute value which is suggestive of other care-implicating attribute values will result in <u>directed</u> problem space searching.

Specific Testing of Higher Cognition in the Data

In the exploratory analyses it was suggested that expert nursing cognition featured contingent and directed assessment of patients on the basis of knowkedge based inference. An example of assumption based on an 'informative' attribute value was when an unconscious MENTAL STATE led to most Experts but fewer Proficient nurses subsequently <u>not</u> selecting URINARY CONTINENCE since, as it was sometimes made plain in the transcripts, unconscious patients are usually catheterised. Examples of the obverse mode of inference were also offered - for patients with a 'poor' value of SKINTYPE the immediate course was to search the URINARY CONTINENCE attribute for a possible causal explanation.

This more specific analytical section seeks to establish the reliability of these suggestions. If the exercise is successful then a specification of these forms of higher cognition can be made in order to add to the gradually unfolding cognitive model of expert nursing cognition. Aside from the process traces of the Expert subjects, there are two other sources of data available to this specific analysis. Firstly, the record of patients contained in database2, and secondly, the transcripts of subjects' verbalisations.

(note: database1 cannot be used since at a later point the patients therein will be used as a test-set for the cognitive model)

In contrast to the attribute importance analysis, this type of cognitive processing is far from 'quiet'. Higher cognition, particularly abductive inference, is firmly Mode 2 processing. By implication, then, the approach for more fully testing the existence of these goal-directed modes of cognition should be to search the transcripts for all likely instances of knowledge based inference taking place. The problem, however, is that the experiment was designed to chiefly produce process traces as the main source of data. The lengthy nature of the task acts to reduce the reliability of the subsidiary data - the concurrent verbalisations. The procedures for analysis which will be adopted, therefore, will seek to minimise reliance on transcript data.

Inferences based on assumption

In an exercise to establish the reliable existence of this type of higher cognition (henceforth termed 'assumptive inference' for convenience) there would ideally be no reliance on the transcript data. This is not solely due to this data being seen as the weakest available but rather takes account of the fact that assumptive inference is likely to be a much less 'active' form of higher cognition than is abductive inference, with the result that evidence of assumptions is less likely to be available within transcripts. Fortunately, the rationale for the existence of assumptions suggests that real-world patient data along with process traces will be of most usefulness for the exercise of testing for the existence of this type of cognition in Expert subjects. The simple rationale is that reliable co-occurence of attributes is the basis for a 'unit of assumptive knowledge'.

The two phase selection process which was adopted when identifying these units comprised:

A. search through database2 for 'likely' assumptive knowledge units and identify those passing statistical criteria,

B. look for evidence that Expert subjects are employing identified units.

In Phase A progressively rigorous criteria are adopted. Hence, if it is found that all of the 159 patients in database2 who are attribute1/value3 are also attribute 6/value2, then that becomes a likely assumptive knowledge unit – evidence for which can be searched within the process traces. Prior to this going through the database, however, it is necessary to establish a criterion for reliability of co-occurence of attribute value pairings. This must take into account not only the strength of the association and the numbers of patients involved – if both of the 2 patients who are x are also y then the reliability of this potential unit of assumptive inference is unimpressive.

The statistical procedure adopted to screen out unreliable units is the straightforward binomial test with population estimates of proportions in each category (see below). It is also planned that the criteria of strength of co-occurence be taken seriously by setting a rigorous rejection region of p>.0001. In other words, if a pair of attribute values appeared to co-occur then when the test was applied the probability of the observed degree of co-occurence coming about by chance would have to be less than p=.0001. An example calculation will illustrate the approach from the point when a following potential unit of assumption has been identified:

STEP 1: program SPSSx to construct crosstabulation frequency tables of all possible attribute combinations.

STEP 2: scan output for potential units of assumption; pick up the following ....

IF URINARY CONTINENCE -is- fully continent and self-caring

THEN MENTAL STATE -should be- alert and orientated.

STEP 3: calculate the proportion of 'fully continent and self-caring' patients who are 'alert and orientated'. Raw index of association = 51/55 = .93 Note number of patients not 'alert and orientated' (x) = 55-51 = 4

STEP 4: calculate the proportion of patients in the population of N=159 who are 'alert and orientated' (P) = 104/159 = .654 Therefore proportion of patients in population not 'alert and orientated' (Q) = 1-.654 = .346

STEP 5: apply binomial test;

$$z = \frac{(x + .5) - NP}{\sqrt{NPQ}}$$

$$= -16.61$$

STEP 6: interpret result using z tables; if z > 3.8 then accept (as in this case) the unit of assumptive inference.

Steps 1 to 6 were iteratively applied to database2. The outcome was that 12 potential units of assumption passed the selection criteria and were taken forward to phase B – the search for support for each unit within the process traces of the Expert subjects.

Phase B also sought to incorporate rigourous selection criteria in that the basic approach was one of accepting a potential unit if it could be demonstrated that 'consensus support' existed. These units of assumptive inference lead to a subsequent attribute not being selected. Therefore the number of occasions when a relevant attribute could have been selected (but wasn't) requires to be expressed with respect to total occasions.

For example, for the 5 Experts there were 10 occasions when a nurse could have assumed a patient's MENTAL STATE given the cue attribute value 'fully continent and self-caring' had just been elicited. In 6 out of these 10 process traces the subject did not subsequently select that attribute. Since this represents the mode behaviour, this particular unit of inference became accepted. Only the 7 'confirmed by consensus' units of knowledge are set out below along with an illustrative comment from one of the transcripts. The final figure given, as in .93, represents the index of association which was calculated in Phase A. These 7 units of assumption are ......

1. cue attribute + value = URINARY CONTINENCE – fully continent and self-caring
   comment (E2) = "...wont bother with mental state then, should be ok ..."
   consensus attribute assumed = MENTAL STATE (.93)

2. cue attribute + value = MENTAL STATE – heavily sedated or unconscious
   comment (E4) = "...so I can take it that they'll probably be catheterised ..."
   consensus attribute assumed = URINARY CONTINENCE (.86)

3. cue attribute + value = NUTRITIONAL STATE – evidence of protein and/or vitamin deficiency
   comment (E3) = "...that would send alarm bells ringing about emaciation..."
   consensus attribute assumed = BUILD (.92)

4. cue attribute + value = CIRCULATION – poor
   comment (E2) = "...so skin wont be good since it's not adequately perfused.."
   consensus attribute assumed = SKINTYPE (.91)

5. cue attribute + value = BUILD – significantly underweight
   comment (E4) = "...probably means nutrition's bad ..."
   consensus attribute assumed = NUTRITIONAL STATE (.95)

6. cue attribute + value = BUILD – overweight
   comment (E1) = "...so his nutritional state should be ok ..."
   consensus attribute assumed = NUTRITIONAL STATE (.93)

7. cue attribute + value = URINARY CONTINENCE - fully continent and self caring
   comment (G1) = "...so bowels should be no problem ..."
   consensus attribute assumed = FAECAL CONTINENCE (.94)

Inferences based on abduction

The search for evidence of units of abductive inference cannot be primed by a scan of database2 or of the process traces since there is no basis to the existence of patterns of association between the attribute values in either of these sources. That is, the rationale for the suggested existence of this type of higher cognition is that nurses' search of their problem space seems to be directed by care-related hypotheses. The decision to select a specific attribute for display could be, on occasions, contingent on the attribute values which had already been elicited. Since this decision is based on acquired knowledge, the attempt to recover these conditional and consequent attributes must begin, therefore, with the transcripts.

It was argued earlier that abductive inference and hypothesis generation is the form of cognitive processing most likely to be available to consciousness. As such, if a unit of abductive inference reliably existed then verbalisation of its basis would be expected from at least one Expert nurse. However, the cognitive model planned from this project has the stated goal of seeking to emulate group rather than individual expertise. Reliance, therefore, will not be placed solely on transcript data when searching for these units - the second criterion will be required of consensus agreement from the measured behaviour within the group's process traces. Hence, if a unit of abductive inference is suggested by the transcript of Expert n then the procedure will be to examine the process traces of all other Experts to ascertain if their behaviour conforms to that predicted by the unit of inference. That is, even if the transcripts of the other Experts do not support the particular unit of inference, if each behaves as if they are using the unit by selecting the appropriate attribute then that unit will become established.

To illustrate, Expert 4 supplied a potential unit of abductive inference with the example given earlier about eliciting that the patient had a urodome and moving to ascertain if the patient's mental state might lead to it being pulled off. The process traces of the other nurses were searched to see if they proceeded from URINARY CONTINENCE-urodome fitted to MENTAL STATE. Due to only one other patient in the database having a urodome fitted and because MENTAL STATE had to be 'unsearched' at that point, only two of the other nurses' process traces could be inspected. Nevertheless, since each nurse (now 3 in total) moved directly from URINARY

CONTINENCE to MENTAL STATE then this particular unit of knowledge was adopted for subsequent inclusion in the cognitive model.

The units below, therefore, are those which passed these twin selection criteria. Format is largely as for assumptive units;

cue attribute value - the conditional part of an abductive inference unit;

verbal response (given by nurse x) - the segment of transcript which drew attention to the potential of this unit;

consensus next selected - the most popular next attribute selected followed by details; eg '4/7,2' denotes that on 4 out of 7 process trace occasions this attribute was selected and that the next most popular attribute was selected on 2 occasions.

1. cue attribute + value = MOBILITY - bedfast and immobile in bed
   verbal response (E4)= "...just check their mental state to see if they would be thinking to move themselves ..."
   consensus next selected = MENTAL STATE (4/7,2)

2. cue attribute + value = MOBILITY - bedfast but can move freely in bed
   verbal response (E2) = "...is this patient incontinent because that means friction on the skin.."
   consensus next selected = URINARY CONTINENCE (5/9,3)

3. cue attribute + value = MOBILITY - fully ambulant. Restrictions few if any
   verbal response (E2) = ".. mental state to see if they would move around.."
   consensus next selected = MENTAL STATE (11/25,4)

4. cue attribute + value = MENTAL STATE - heavily sedated or unconscious
   verbal response (E3) = "...so there'll be unrelieved pressure on pressure points ... so if there's no padding ..."
   consensus next selected = BUILD (3/3,0)

5. cue attribute + value = URINARY CONTINENCE - occasional incontinence (eg at night)
   verbal response (E1) = "...that would make me think the skin might break down..."
   consensus next selected = SKINTYPE (4/8,3)

6. cue attribute + value = URINARY CONTINENCE – catheterised / urodome – not bypassing (when MENTAL STATE as yet unknown)
   verbal response (E4) = "...a urodome fitted ...I'll need to check to see if he's confused otherwise the urodome will be useless for preventing incontinence.."
   consensus next selected = MENTAL STATE (3/4, 0)

7. cue attribute + value = URINARY CONTINENCE – catheterised / urodome not bypassing (when MENTAL STATE already known)
   verbal response (E3) = "..ah, but we may get moisture from ...if he's incontinent of faeces.."
   consensus next selected = FAECAL INCONTINENCE (3/5,2)

8. cue attribute + value = BUILD – significantly underweight
   verbal response (E1) = "...so circulation would have to good to prevent bruising..."
   consensus next selected = CIRCULATION (3/7,1)

9. cue attribute + value = SKINTYPE –(any value other then 'normal' when URINARY CONTINENCE is already known to be 'dry')
   verbal response (E2) = "...I wonder if that's because his circulation is..."
   consensus next selected = CIRCULATION (5/8, 2)

10. cue attribute + value = CIRCULATION – poor
    verbal response (E3) = "...that may have affected their skin..."
    consensus next selected = SKINTYPE (4/9,1)

11. cue attribute + value = NUTRITION – evidence of protein and/or vitamin deficiency
    verbal response (E3) = "...skin could well be affected by that..."
    consensus next selected = SKINTYPE (6/14, 3)

12. cue attribute + value = FAECAL CONTINENCE – occasional faecal incontinence or diarrhoea
    verbal response (E2) = "...not just the wetness here but also having to use soap often would make me think her skin will be liable to breakdown..."
    consensus next selected = SKINTYPE (3/4,1)

<u>Limitations</u> Given the inevitably partial nature of the database – all possible combinations of attribute values could not hope to be represented in the experimental database – it is reasonable

to suppose that these lists of units of knowledge are not exhaustive. A further limitation of the validity of the eventual cognitve model may lie in the selection method adopted when gathering these 7 assumptive and 12 abductive units of knowledge. Nevertheless, the stringency of inclusion criteria leads to a trade-off; greater stringency lessens the risk of false selection while increasing the risk of false rejection. It should be stressed that chance factors were being fairly tightly controlled - to return to the 'urodome' example, MENTAL STATE was selected when, for the 3 nurses, there were 9, 8, and 8 attributes remaining unsearched. The chance probability that one particular attribute would be next selected on each occasion comes out at $p = .0017$. The hope is that the balance has been about right - a hope which can immediately be put to the test through the point-by-point agreement procedure.

Units of inference- Point-by-point agreement with the data

The comparison of a matrix containing the identified units of inference with each subject's matrix will be the final test of validity of the identification of units of inference. It is not possible, however, to construct a matrix formed from the process traces of higher cognition units alone since only a proportion of any one patient's attribute values will fit the cue parts of the 19 units of inference. For some patients, particularily those with few 'informative' attribute values, there may be no occasions when one of the IF....THEN 'rule' will be triggered. Moreover, it would not be sensible to test these units in isolation since, as shown on the previous section, heuristic search is also important. This discussion on higher cognition has thusfar not incorporated the findings from the previous section on attribute importance. Testing of units of inference cannot therefore be separated from heuristic search. The solution to the problem of point-by-point testing is, therefore, at hand - although it is firstly necessary to explore how this incorporation is to be achieved.

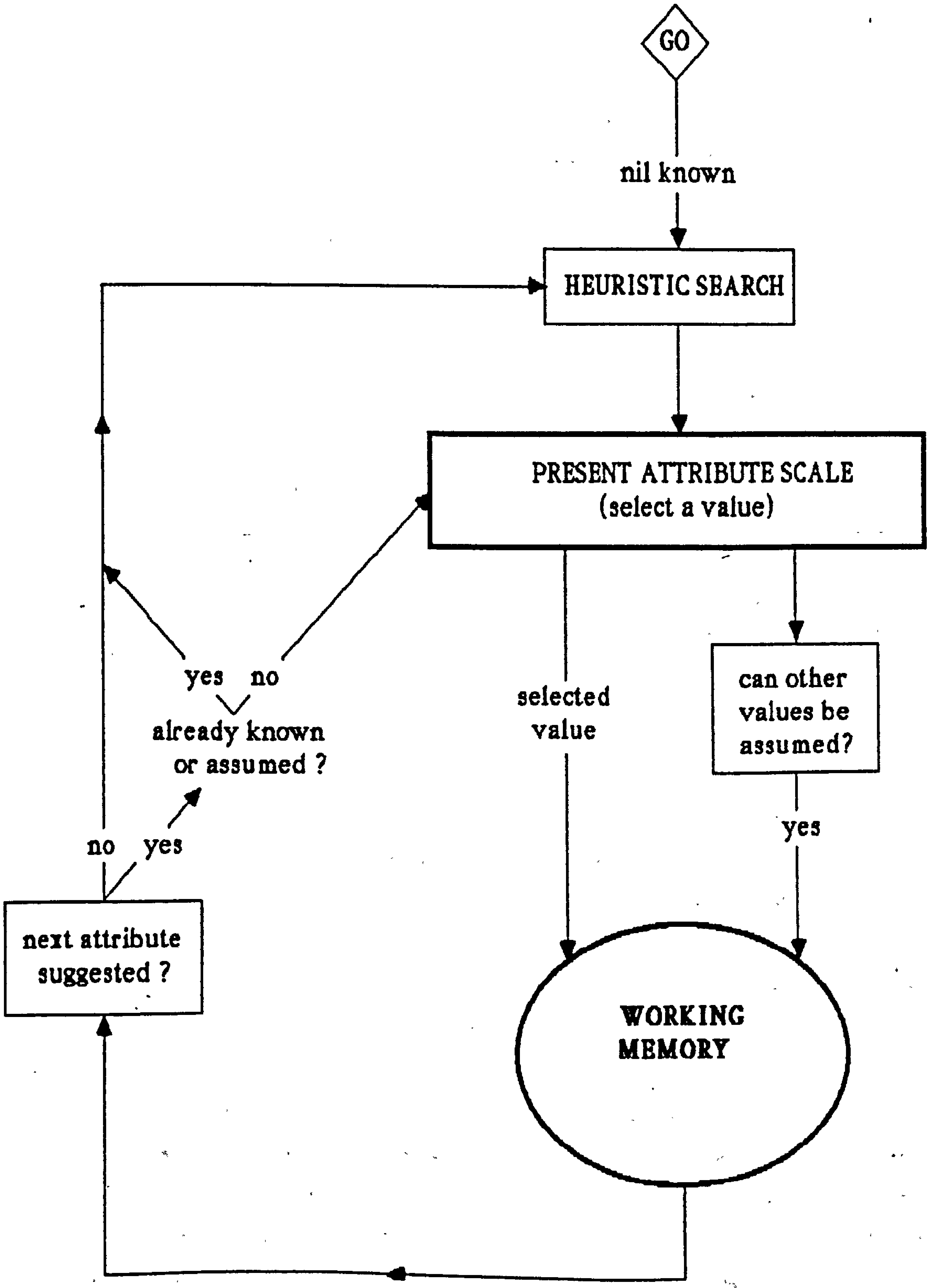
In order to incorporate the attributes in order of importance and the units of cognition, it was firstly necessary to program each feature into computer code. Flow of control between the features must, however, be determined. The basis for control was decided by the nature of each feature and by the data. Hence, each patient assessment begins with header attribute (heuristic search) and will continue through the ranked list of attributes until such time as an elicited attribute value 'cues' a unit of inference. Each type of inference unit also has functions specified - either to assume an unknown attribute value or to direct search.

The resulting program is therefore an update on the 'heuristic search' cognitive model from

Figure 3.2. As illustrated in Figure 3.3 overleaf, the fundamental circuit of this 'Heuristic Search + Inference' model is to chain through the Assessment list ordering of attributes. Higher cognition, however, may serve to alter the resulting process trace:

- firstly, each time an attribute value is elicited a check is made to see if anything can be assumed from this new knowledge, ie the 7 units of assumption are consulted. If a unit is triggered then that attribute is marked as 'known' on the list and subsequently will not be directly searched.
- secondly, after each attribute value is elicited a similar check is made on the 12 units of abductive inference. If one is triggered then heuristic search is bypassed and the search is directed to the appropriate consequent attribute in that unit.

Figure 3.3 Flow diagram for Heuristic Search + Inference model



Lest too much be made of data of undemonstrated reliability, the task now is to construct the matrix representing the model's process trace in order to achieve the point-by-point comparison with the Experts' process traces. This matrix was constructed by running the model as in Figure 3.3 with each of the 18 patients assessed by the Experts. Hence, each time the model presented an attribute on the screen with a choice of values (analogous to the Expert 'clicking' a button beside an attribute), the appropriate value for the patient currently being assessed was entered. The log made of the order of attribute search became the process traces which made up the matrix for the model. Results are presented in Table 3.6 below.

Table 3.6 Process traces of Heuristic Search + Inference model

	patient																	
attribute	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
MOBILITY	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
URINARY CONT	2	3	2	3	2	4	2	2	2	2	2	-	2	3	3	3	3	3
SKINTYPE	4	6	9	9	5	3	9	5	-	9	7	9	8	5	8	8	8	8
BUILD	7	-	5	5	-	6	5	-	5	6	3	5	5	-	5	5	5	5
NUTRITION	6	5	4	4	4	5	4	4	4	5	-	4	4	4	4	4	4	4
AGE	9	7	6	7	6	7	6	6	6	7	4	6	6	6	6	6	6	6
MENTAL STATE	5	2	3	2	3	2	3	3	3	4	2	3	2	2	2	2	2	2
CIRCULATION	8	8	7	6	7	8	7	7	7	8	5	7	7	7	7	7	7	7
LIFT & TURN	10	9	10	10	9	10	10	9	9	10	8	10	9	8	9	9	9	9
SEX	11	10	11	11	10	11	11	10	10	11	9	11	10	9	10	10	10	10
BLOOD PRESS	12	11	12	12	11	12	12	11	11	12	10	12	11	10	11	11	11	11
FAECAL CONT	3	4	8	8	8	9	8	8	8	3	6	8	-	-	-	-	-	-

It can be seen from the matrix for the model that the addition of the units of inference has brought about change from the fixed ordering of attributes seen in the Assessment list. Clearly there were occasions when search became directed toward a particular attribute. Moreover, there are now occasions when an attribute is left unselected (denoted in the Table by –), this corresponds to a unit of assumption becoming activated with the result that the value for these attributes could be assumed and they were therefore not subsequently searched.

The matrices for each subject could now be convolved with the matrix for the model. The results are displayed under 'Heuristic + Inference model' in Table 3.7 overleaf. The number of exact cells in the matrix which match is given under 'n' while the proportion matching cells to total cells (216) is given under 'P'. Also displayed here for reference are the results derived from the Assessment list testing (ie the 'Heuristic' component of the 'Heuristic + Inference model'). Although no data from the Proficient nurses helped form the units of inference, it is interesting to see if the incorporation of Inference brings about an improvement in the prediction of their data points – results also displayed in Table 3.7

Table 3.7 Point-by-point indices of agreement between each subject and the 'Heuristic + Inference' model

subject	total attribs selected	Heuristic Search alone		Heuristic + Inference model	
		n	P	n	P
E1	71	7	.03	15	.07
E2	68	27	.12	38	.18
E3	92	28	.13	45	.21
E4	78	34	.16	48	.22
E5	74	7	.03	15	.07
P1	89	21	.10	28	.13
P2	103	9	.04	20	.09
P3	131	15	.07	20	.09
P4	90	15	.07	19	.09
P5	167	19	.09	21	.10
P6	72	17	.08	31	.14
P7	105	33	.15	30	.14

From Table 3.7 it can be seen that point-by-point agreement is superior for the Heuristic + Inference model versus the Heuristic Search model within all Expert subjects and for all but one Proficient nurse (P7). The scores were analysed with 2 X 2 ANOVA using each group as a level of the between groups factor 'Group' and each explanation as a level of the within subjects variable. Results, as shown in Table 3.8 below, demonstrate that the difference between the mean scores for the two explanations was significant (F=34.6, df1 and 10, p<.001). The interaction term shows that these mean differnences were not, however, relatively superior within one group when compared to the other. Mean scores:

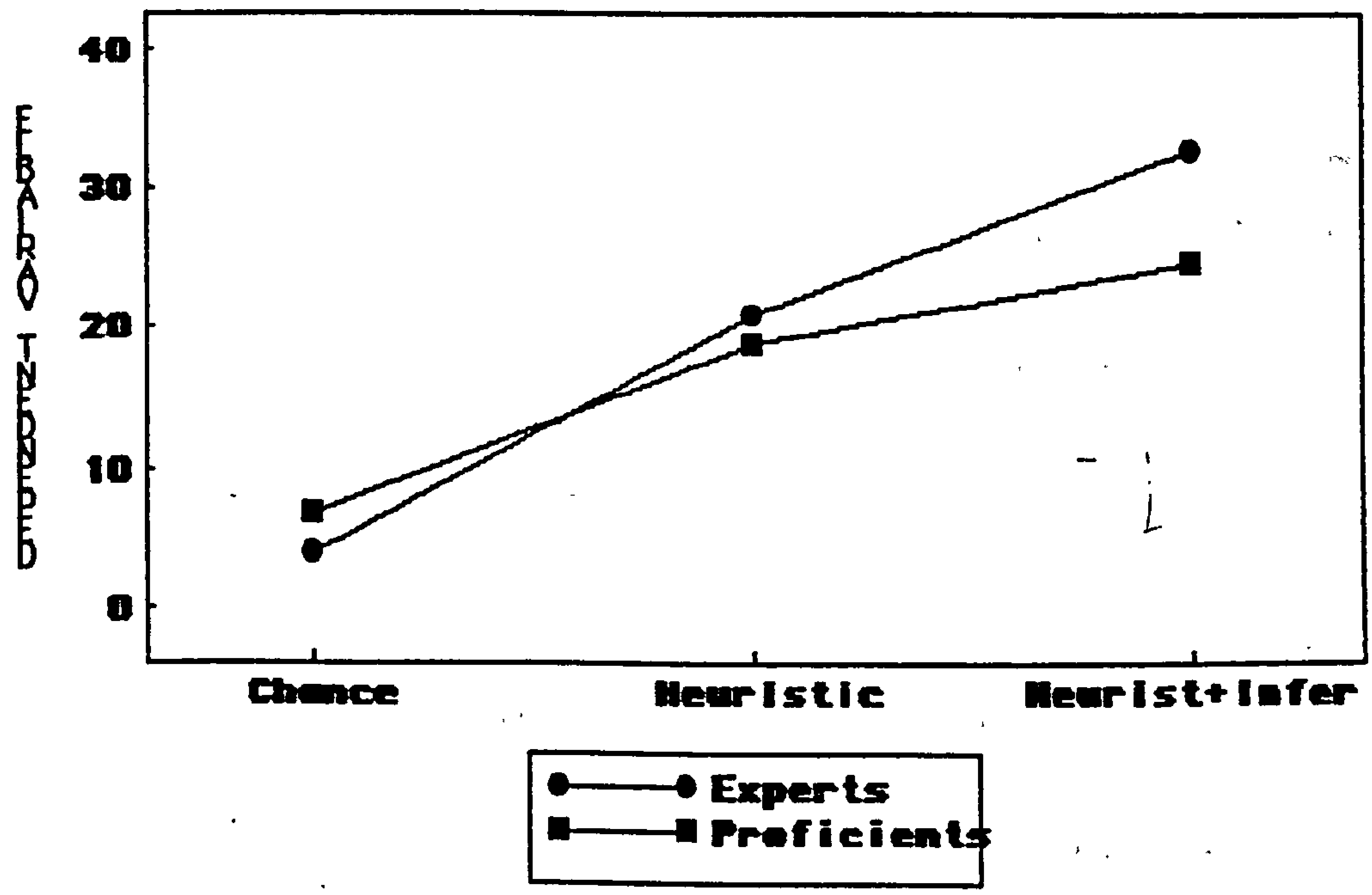
Experts- Heuristic= 20.6, Heuristic + Inference= 32.2
Proficients- " = 18.4, " " = 24.1

Table 3.8 Source table for Group X Explanation ANOVA

Source of Variation	df	SS	MS	F	p
Group	1	152.6	152.6	.74	ns
error	10	2058.2	205.8		
Explanation	1	437.2	437.2	34.60	<.001
Group X Explanation	1	50.5	50.5	3.99	ns
error	10	126.3	12.6		

It is finally interesting to plot the average numbers of data points explained by models based on the Chance ranking scheme, the Heuristic Search ranking scheme, and the Heuristic Search + Inference model. The result, as displayed in Figure 3.4, shows steadily increasing point-by-point agreement.

Figure 3.4 Average number of each groups' data points which are predicted by Chance, Assessment, and Heuristic Search + Inference model



As high a proportion as 61% of attribute selection positions are now predicted by the cognitive model (Expert 4). Bearing in mind the discussion in the introduction to this chapter on the probability of a match occuring by chance, it would seem that this degree of concordance is becoming impressive - until, that is, it is noticed that the model is selecting all attributes save those few whose value are assumed. When <u>all</u> cells in the matrix are considered, the proportion of those explained by the latest model is seen to be only 48/216 = 22%. Clearly the section to follow which analyses the point at which Expert subjects stop gathering information becomes important. Before turning to that analysis, however, some closing comments can be offered on the implications of the analysis of higher cognition as an explanation of expert nursing information processing.

Implications for the Cognitive Model of the Findings on Higher Cognition

It is useful to firstly restate the lines of evidence which support the findings regarding higher cognition:

- the idea that the micro and macro-level goals of processing will direct information acquisition.
- the idea that cognitive economy predicts that nurses will use stored representations of attribute value co-occurence.
- the evidence that level of experience is reflected in patterns of information acquisition.
- the evidence from set patterns in the process traces which agree with a separate database of associations between attribute values.
- the evidence from verbalisations in the transcripts which agree with patterns in the process traces.

The implications from an incorporation of heuristic search and higher cognition is that the basis of assessment will be the 'automatic' or Mode 1 type of processing. Only when goal-relevant information is gathered does the more serial Mode 2 'higher' form of cognition become involved. If, as Schneider and Shiffrin (1977) suggest, this type of cognition is more under conscious control and has greater memory capacity implications, then it is entirely consistent that the default mode of processing used by the nurses is the less capacity limiting one. Nevertheless, although Mode 2 processing has memory capacity limitations it should be pointed out that conversely the use of higher cognition can in fact reduce memory load.

Consider the more usual situation where a time-pressed nurse is assessing a newly-admitted patient with 100+ attribute values. It is reasonable to suggest that this nurse will have developed strategies to optimise effectiveness of the search while maximising cognitive economy. It also seems reasonable to suggest that these same nurses undertaking the proces tracing task will have brought to bear the processing styles developed in the life situation. Higher cognition, in the model, acts to direct the search and afford greater 'sense' from patient details. The result is that the model, like the nurses, has the capacity to alternate between high and low modes of processing on the basis of incoming information interacting with stored knowledge.

No consultation, however, ends without regress to Heuristic Search at a point subsequent to start. This could be argued as reflecting the fact that the data analysis failed to generate a sufficient set of stored packets of knowledge so that default to 'next most important' could be avoided. For reasons to be set out below, however, it is considered that this would have been an incorrect course to have followed.

The impression which may be gained regarding flow of control in the cognitive model is that higher-level cognition has primacy over lower-level processing. To an extent, this seems to contradict the evidence from novice/expert differences in the problem-solving literature which shows that with the acquisition of expertise there is a move from conscious, goal-directed processing to automatic, data-driven processing. This impression, however, is over broad and takes no account of the nature of the two forms of 'higher' cognition identified. Hence units of assumption are evidently more automatic and data-driven than are units of abduction.

The cognitive architecture implication of the findings on higher cognition is that nurses represent this experience-based knowledge in small 'units' or 'packets'. Schank's (1980) idea was of MOPs - Memory Organisation Packets. Clear analogies can also be found in the work of Anderson (1983) - as the above example suggests, condition-action pairs seem almost tangible. It is being suggested, therefore, that the type of higher-level cognition being proposed is being driven by representations which are considerably smaller than those proposed by schema theory.

In the example above, if the fact 'urodome fitted' instantiated a schematic representation of typical such patients, then it might be predicted that subsequent searching of the problem space would be driven by the need to fill slots for which a default value did not reliably exist. 'Might be predicted' is apt given the recent criticisms of schema theory as having little to say about cognitive processes other than the interpretation of input data (eg Anderson 1983). The lack of processing mechanisms, therefore, is the chief reason for this analysis favouring an architecture of high-level cognition based on a collection of smaller units of knowledge.

There are two advantages, then, to the smaller 'packet' type of representation being proposed. Firstly, the goal-directed nature of the cognitive task put before subjects would seem to clearly implicate an internal representation which makes explicit the goal of assessing these patients - to devise the optimal nursing care plan. Secondly, the output of cognition - how responses and actions are created - is taken in this analysis as being perhaps the principal reason for nurses' use of higher-level cognition.

A final argument to support the existence and use of higher-level mechanisms becomes plain when the underlying goal of this type of goal-directed cognition is considered. It must be kept in mind that nurses assess patients in order to achieve several goals. At a macro-level this goal is to plan preventive care for the patient. At a micro-level the goal is to tailor the chosen care to a particular patient. Experience, in these terms, is knowledge of where to look for the necessary information and what implications that information carries.

It is lastly interesting to recall the concerns raised in the introduction regarding the emphasis on higher cognition within a hypothesis testing mode. As Grier (1981) suggests, this emphasis is perhaps due mainly to the development of formal nursing models - although the advent and influence of ward-based information systems cannot be excluded. The term 'formal model' contains two distinctions which are crucial to further discussion. Firstly, as D. Gordon (1984) points out, nursing models can be abstractions *of* reality or abstractions *for* reality. Secondly, models can focus on nursing in a *global* or in a *specific* sense.

Nursing authors, particularly in North America, have tended to move from their global abstractions of nursing in its largest context to prescriptions for specific aspects of nursing. The aim - to achieve better preparation of nurse learners - cannot be faulted. Nevertheless the 'diagnostic' model in particular has become increasingly reified (equated with reality) on the basis of supposition. Furthermore, there is a danger that individual expertise becomes devalued in favour of oversimplified yet mystified complex issues for the dubious benefit of inexperienced nurses (see D. Gordon 1984 for an eloquent discussion of these points). Cognition is but one specific aspect of nursing. To say 'this is how you should process information because it fits with our overall model of nursing' might only be acceptable given valid evidence.

POINT FOR DECISION MAKING

It has been established that on all but exceedingly few occasions the subjects left unsearched a varying number of a patient's attributes. More experienced nurses, moreover, directly select less information than less experienced nurses. The findings from the previous section on inference suggest that nurses may 'know' about more attributes than they have directly searched. Nevertheless, there remains a large discrepancy between the number of attributes searched by the model when compared to either the Expert or the Proficient group. It becomes necessary, therefore, to explore the conditions which describe the point at which the information gained is taken as sufficient. It becomes necessary, moreover, to explore what 'sufficient' might mean.

Selected Literature Review on the Point for Decision Making

The quantitative analysis made plain that the Expert group of nurses asked fewer questions of the database than did the Proficient nurses. What was not established was the conditions of this pattern, why was it that Expert nurses asked less questions yet apparently achieved greater accuracy? One explanation is straightforward - experienced nurses are better able to make a 'risk' decision since the nature of experience necessarily involves familiarisation with the types of example patients presented in the experiment.

An explanation focusing on the decision is, however, partial in terms of a complete model of information processing. It has already been shown that the process of information acquisition can be expected to vary contingent on items of information already elicited. Furthermore, previous work in the area of information acquisition, as reviewed below, has established that stategies of search will seek to maximise goal attainment while minimising processing capacity. It becomes important, therefore, to examine process of decision making rather than outcome when seeking to capture the expertise of knowing when enough information is sufficient information.

There are further points which can be made which argue that the focus should be on process rather than outcome of decision making. One observation from the the field of medical diagnosis (eg Elstein et al 1978, Kassirer & Gorry 1978) is that there has been a continued failure to demonstrate the novice/expert differences in diagnostic skill which have been fairly well documented in other fields. However, virtually all of these medical studies have concentrated on the product of decision making, ie diagnosis as the goal. More recent medical work (Patel and Groen 1986) has shown that experienced physicians have as their initial goal the task of constructing an elaborated understanding (representation) of the patient being assessed. This representation becomes crucially involved in the final goal of making the diagnosis.

This shift in focus from study of product more towards study of process is a feature of the project generally. More immediately, a process focus serves to underline the importance of methodogy and the effect of experimental design on findings. There are, for example, some findings from the nursing literature which apparently contradict the evidence form the present study that expert nurses collected less information than other nurses. Itano (1988), for example, found when studying nurses' interviews of patients that the registered nurses collected more cues than were collected by students. Broderick and Ammentorp (1979) also found that their expert subjects asked more questions than did the novice sample. Each of these studies, however, relied not only on assessment of only one patient but also involved significant demand characteristics where subjects would have been highly aware that their degree of expertise was under scrutiny. The design of the present study went to some lengths to avoid these threats to external validity.

A closer analysis of the study by Broderick and Ammentorp (1979), moreover, shows that the categories of information where significantly more patient details were gathered by experts relate strongly to current vital or neurological signs, and pain being experienced. As the authors comment, this suggests that expert nurses were spending more time seeking relationships between data elements. The familiar picture which is emerging, then, is that the point for and process of decision making is in some way related to the mental representation of the patient. It is not possible, however, to infer from these findings anything regarding the point at which experts' representations are taken to be sufficiently elaborated for the purpose of making the decision.

In the nursing literature the nature of nurses' patient representations continues to be defined as one which involves fairly superficial knowledge such as patient characteristics rather than deeper conceptual knowledge. Nursing theoreticians adopt this idea by stating that the point for making a decision will come when the nurse feels that diagnostic cues of patient characteristics have been noticed (eg Carnevali 1983). And yet, as Baumann and Bourbonnais (1982) show in their study of rapid but complex decision making by critical care nurses, it is possible for nurses to make accurate decisions on the basis of very little data. Moreover, these decisions relate not just to diagnosis but to the 'next step' of patient care management.

If decision making is largely concerned with representation of patient features then, as Corcoran (1986) argues, it is inescapable that working memory capacity limitations are implicated in an understanding of the point for decision making. A further point following from the work of Polson and his co-workers (eg Attwood and Polson, 1976) makes clear that the inferred knowledge about 'attribute to be selected next' (successors) will also be stored in working memory. Working memory load, therefore, would quickly begin to exceed the accepted limitations even though it might be expected that experts would organise the information into some sort of chunks. If, however, patients are represented more pertinently in terms of a deeper conceptual schema then it would be expected that a sound basis would exist for the efficient organisation of the representation. Capacity limitations would not therefore be so directly implicated (Chi et al 1981). It becomes important, therefore, to speculate on and search for a suitable schematic framework of conceptual knowledge.

Such an framework would reasonably be related to the task at hand - pressure sore risk assessment. The work of Bergstrom and Braden (eg Braden and Bergstrom 1987) is apposite since the focus of this research has been to model the biological domain of pressure sore aetiology in terms of a 'conceptual schema'. The argument would be that experienced nurses carry deep level knowledge in the form of this conceptual schema; when they have fitted to their satisfaction the incoming patient details to this biological model then, it would be predicted, the search process would stop. The work of Hawkins (1986) in the expert systems area supports the validity of this idea.

This review, therefore, has suggested two principal contender explanations, either of which might best explain the points at which subjects stopped gathering information. These explanations are firstly that subjects hold deep level knowledge in the form of a biological model of pressure sore aetiology. The point at which a subject will stop information collection will be predicted by the point at which the patient can be 'fitted' to this model and when no further successor attributes are in working memory. The second explanation is based more on superficial representation of patient features, the prediction being that working memory capacity limitations will act to stop information collection when there are no successor attributes and when a capacity limit has been reached.

<u>Exploration of the Point for Decision Making in the Data</u>

The biological model of Bergstrom and Braden, firstly, conceptualises pressure sore development to be a function of the intensity and the duration of pressure and the tolerance the tissues to withstand that pressure. Two main factors relate to pressure, and two factors relate to tissue tolerance. Although this model was introduced in Chapter 2, it is worth restating these four 'dimensions' of pressure sore risk in terms of the attributes from the present experiment which most appropriately 'belong' to each:

1. The extent to which <u>Mobilising</u> results in pressure on the skin. Although chiefly represented in the current experiment by MOBILITY, other attributes which implicate Pressure are MENTAL STATE and BUILD. For example, an unconscious heavy patient will have unremitting and intense pressure.

2. The sensory or perception or <u>Capacity</u> of the patient to relieve pressure or at least tell the nurse of pain experienced. Represented by the attribute MENTAL STATE.

3. The <u>Extrinsic factors</u> impinging on the skin which can lead to breakdown of that skin – notably involving moisture (URINARY and FAECAL CONTINENCE) but also attributes such as MENTAL STATE might be implicated since a restless patient will generate friction on the skin.

4. The <u>Intrinsic factors</u> of the patient which affect their susceptibility to skin breakdown – relevant attributes here are AGE, SKINTYPE, BUILD, NUTRITIONAL STATE, BLOOD PRESSURE and CIRCULATORY STATE.

The task in fitting this model to the data is to determine if the dimensions explain 'stop' points. Questions which require answers relate to the conditions which dictate when each dimension is deemed to have been adequately searched. Dimensions 1 and 2 give no difficulty; each requires only one principal attribute to have been searched. Dimension 3 and 4, however, raise problems since they contain several attributes.

Dimension 4 – Intrinsic factors – is interesting. All assessments included some degree of direct search or value assumption of these six attributes. The uncertainty, therefore, is the point at which search of this dimension will be terminated. The other explanation cannot, however, be ignored – the capacity of working memory to retain an increasing number of information units which together comprise a representation of the patient being assessed.

The most economical solution which might be proposed is that search would terminate when each dimension has been searched and when as much of dimension 4 has been searched as capacity limitations permit. This solution, however, fails to take into account two factors; firstly, that the more usual situation faced by nurses is assessment which involves much more than 12 attributes. Secondly, it is unwarranted at this stage to make assumptions about the nature of the representation which a nurse builds up of a patient she is assessing. Certainly the as yet untested 'careplanning hypothesis' suggests that these representations are considerably more complex than 'tick off each dimension'. Ahead of the further experimentation to be undertaken later, therefore, a more straightforward approach is required to assess the validity of whether a conceptual model has a role to play in predicting that a nurse has reached the decision point.

One suitable approach comes from the idea of trying to specify the minimum conditions which will allow a nurse to stop searching attributes and make her decision. There were 3 patients out of the 18 assessed who satisfied the criterion of having no 'problem' attribute values. It follows that an inspection of the Experts' process traces for these patients will reveal the minimum number of attributes which require to be searched prior to giving a 'low risk' decision. The attributes searched can be grouped by the dimension each belongs to, thereby beginning to explore the potential of the conceptual schema as explanation.

The 3 patients concerned were patients 15, 16, and 18. The process traces of the 5 Experts (E1 to E5) can be summarised using the 4 dimensions in the table below. The numbers represent the number of attributes each nurse searches which belong to each of the four dimensions. The Mobilising dimension, for example, was searched once by every nurse for all 3 patients – in effect this means that the attribute MOBILITY was searched since this is taken to chiefly implicate Mobilising.

	Pt. 15					Pt. 16					Pt. 18				
dimension	E1	E2	E3	E4	E5	E1	E2	E3	E4	E5	E1	E2	E3	E4	E5
MOBILISING	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CAPACITY		1	1	1	1	1	1	1	1				1	1	
EXTRINSIC	1	1	1	1				1		1			1	1	1
INTRINSIC	1		2	1	1	3	1	1	1		1	1	3	1	

This table shows that Mobilising was searched on all opportunities and that Capacity (ie MENTAL STATE) was searched on 10 out of a possible 15 occasions. Extrinsic factors were elicited singly (ie URINARY and not FAECAL CONTINENCE) on 9 out of 15 opportunities. Intrinsic factors, lastly, were elicited on 12 out of 15 opportunities at a rate of 1 only (x 9), 2 (x 1), and 3 (x 2). Looking at the nurses, no subject ignored any one dimension across all 3 patients.

The picture from this analysis of minimum conditions for stopping, therefore, is reasonably clear-cut. When the consensus is taken, the simple rule seems to be that each dimension requires a minimum of one and only one 'member attribute' to be searched. The fact that most process traces reveal that more than 4 attributes are selected can be handled by this rule since there is no limitation implied on the number of attributes to which a search can be directed via the units of abductive inference.

There are, therefore, two contenders which may explain the stop points observed in the data. Firstly, limitations on the capacity of working memory. Secondly, the application by subjects of an internal conceptual model of pressure sore risk. The task now becomes one of evaluation of these contenders, by adding each explanation in turn to the current model (the Heuristic + Inference model) and measuring goodness of fit to the data.

Specific Testing of Explanations of the Point for Decision Making in the Data

Nurses have been observed to limit the amount of information they elicit from the pool available for patient assessment. The goal, therefore, is to identify the most adequate explanation of the cognitive basis to these 'stop-points' which characterise the nurses' information processing. The explanations which are being tested are the Conceptual Schema (Concept.), and the Capacity Limitations model (Capacity). Matrices required to be prepared which incorporated a version of each model; to achieve this it was necessary to incorporate new subroutines into the computer code of the 'Heuristic + Inference Model' which would provide a facility for limiting the number of attributes searched. The procedure for each version prepared was a follows:

1. Conceptual schema

The computer based model was modified to include a simple routine which checked off each of the four dimensions outlined above as soon as a single member attribute was searched or assumed. As soon as each dimension becomes flagged in this way the process of searching attributes stops. The only way in which attribute searching can proceed from this point is if a unit of abductive inference has been triggered.

It is possible, therefore, that if the value elicited from the header attribute triggered sufficient units of deductive inference then no more attributes would be searched, ie if member attributes for each dimension had been assumed. It would also be theoretically possible for all 12 attributes to be searched if units of abductive inference continue to direct the search to successor attributes. In practice, however, the range of attributes directly searched when the 18 patients were 'put through' this model was 3 to 6. A matrix based on the process trace of the model over these 18 patients was therefore prepared prior to the comparative exercise.

2. Capacity model

In preparing this model the goal is to emulate a nurse limiting the number of attributes she searches on the basis of number of patient features in working memory. As the earlier review shows, there has been no sufficiently specific previous nursing research in this area, although the work of Attwood and Polson (1976) is helpful. It is therefore not clear how to specify a 'capacity limit'. The least problematic issue is the actual number of items, which could vary around seven. It is the unresolved issues which force significant assumptions to be made about the nature of nurses' representations of patients. These issues critically affect the notional capacity limit, for example, are attribute values organised by active processes on some basis? are other concepts (such as appropriate care) brought into working memory along with attribute values? is some cognitive capacity devoted to maintaining and updating a decision of pressure sore risk?

Attwood and Polson (1976), their model of a human solving the Three Waterjug problem, had to initially make a rather arbitrary assumption about number of successor moves which can be held in working memory. The main aim was the later comparison to observed human behaviour. This exercise, ahead of some later experimentation (Chapter 4) which may go some way to resolving these issues, must therefore be seen as only a crude test of whether capacity limitations explain stop points. In an effort to provide reference points, two Capacity versions will be prepared based on crude implementations of Miller's (1956) findings - one based on a 7 item limit, the other on a 9 item limit. 'Item' is taken to refer to an attribute value (whether directly elicted or assumed) or to a successor attribute. In practice, the 7 item model stopped 'processing' after between 5 and 8 attributes had been searched. The 9 item model had a range from 7 to 10.

Results

Matrices of 18 patients x 12 attributes were prepared for the Concept and two Capacity models. Each matrix was then convolved with those of each Expert. Results, given in Table 3.9 overleaf, follow the same format for each column reporting point-by-point agreement. The exact number of cells in the matrices which match is given under 'n' while the proportion of matching cells out of the total (216) is given under 'P'. Separate results are plotted for the Conceptual Schema matrix and for the two Capacity matrices (Capacity <= 7 and Capacity <= 9).

Table 3.9 Point-by-point indices of agreement between each subject and the 3 explanations of stop-points

subject	Conceptual schema		Capacity <=7		Capacity <= 9	
	n	P	n	P	n	P
E1	107	.50	86	.40	63	.29
E2	142	.66	118	.55	91	.42
E3	132	.61	112	.52	90	.42
E4	143	.66	120	.56	95	.44
E5	102	.47	84	.39	56	.45
P1	111	.51	89	.41	72	.33
P2	94	.43	84	.39	70	.32
P3	79	.36	71	.33	65	.30
P4	92	.43	65	.30	61	.28
P5	46	.21	40	.18	39	.18
P6	130	.60	101	.47	77	.36
P7	107	.49	85	.39	68	.31

From Table 3.9 it can be seen that a perfect rank ordering of superiority of explanations exists across all subjects - from least cells explained (Capacity <= 9) via Capacity <= 7 through to Conceptual Schema. It should be recalled, however, that the high scores must be interpreted in terms of the the contribution of the earlier analyses - each 'stop-point' explanation has been built onto the the existing Heuristic + Inference model. The scores were analysed in order to test this observation for reliability with 2 X 3 ANOVA using each group as a level of the between groups factor 'Group' and each explanation as a level of the within subjects variable.

Results, as shown in Table 3.10 overleaf, demonstrate that highly significant differences exist between the mean scores for the 3 explanations (F=97.7, df2 and 20, p<.0001). The mean scores for each explanation and by group are:

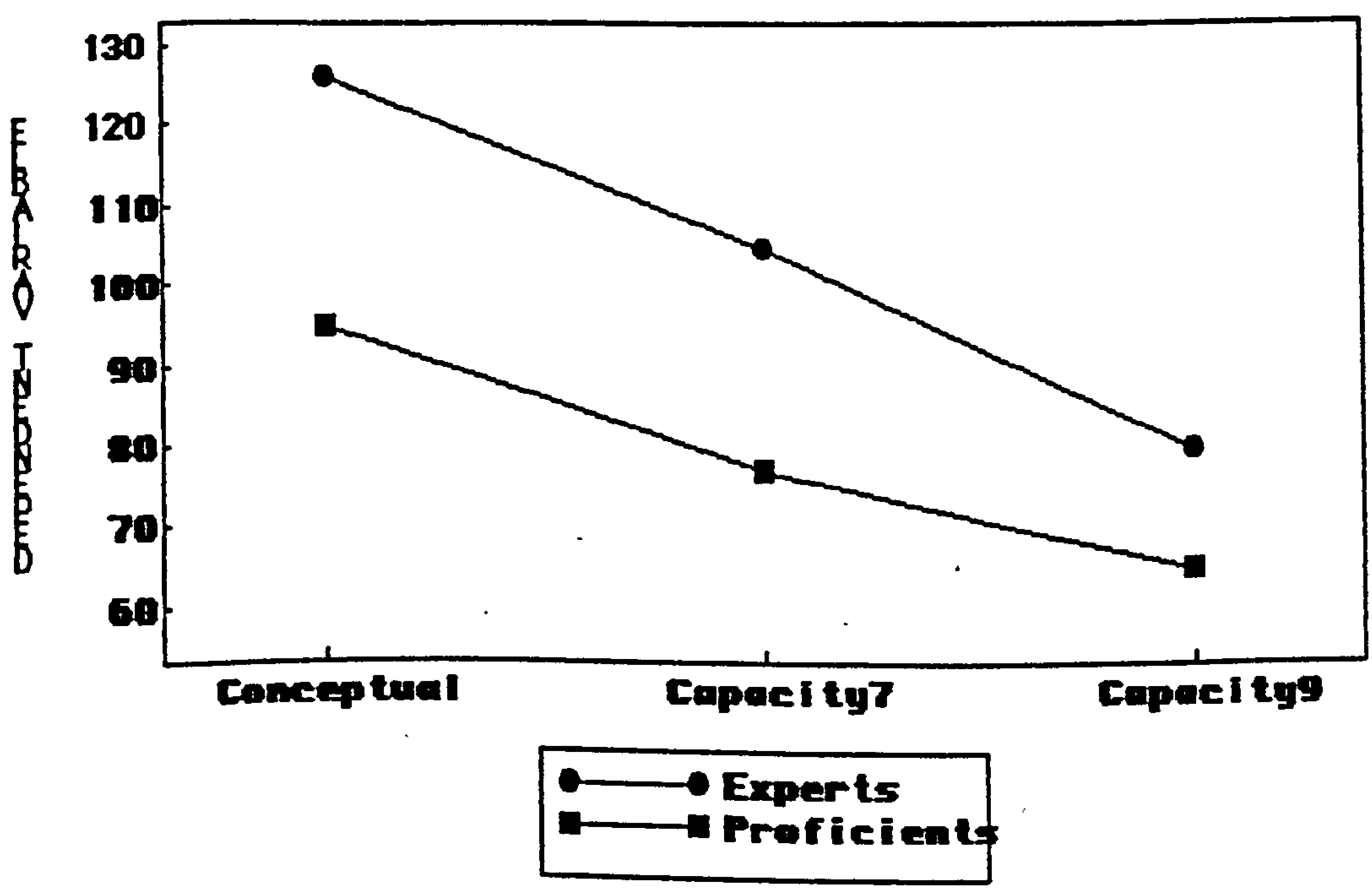
	Conceptual Schema	Capacity <= 7	Capacity <= 9
Experts	125.2	104.0	79.0
Proficients	94.1	76.4	64.6
overall	109.7	90.2	71.8

Table 3.10 Source table for Group X Explanation ANOVA

Source of Variation	df	SS	MS	F	p
Group 1	5189.1	5189.1	4.81	.053	
error	10	10788.5	1078.8		
Explanation	2	8374.8	4187.4	97.78	<.0001
Group X Explanation	2	448.6	224.3	5.24	<.05
error	20	856.5	42.8		

Although the main effect for Group narrowly failed to show a significant difference, the Group X Explanation interaction was significant (F=5.23, df2 and 20, p<.05). It was not possible, however, to analyse the simple effects which contribute to this interaction due to the unequal group sizes. A graphical illustration of the interaction is given in Figure 3.5

Fig 3.5 Number of data points explained by 3 stop-point explanations



The conclusion which can be taken, therefore, is that the most powerful explanation of stop-points is that offered by the Conceptual Schema. It is not, however, permissable to state that this explanation favours one based on Capacity limitations since, as previuosly discussed, these Capacity models must be considered to be crude. The sensible conclusion, then, is that an explanation based on a Conceptual Schema of pressure sore aetiology was found to improve point-by-point agreement of an existing cognitive model beyond the level of two competing explanations based on number of attribute values known.

A final test of goodness of fit to the Expert subjects can be performed using 'number of attributes selected' as a dependent variable. When the 18 patients were run through the model, the number of attributes searched for each of the 18 patients is set out in Table 3.11 along with the mean numbers of attributes selected by both the Expert and the Proficient groups.

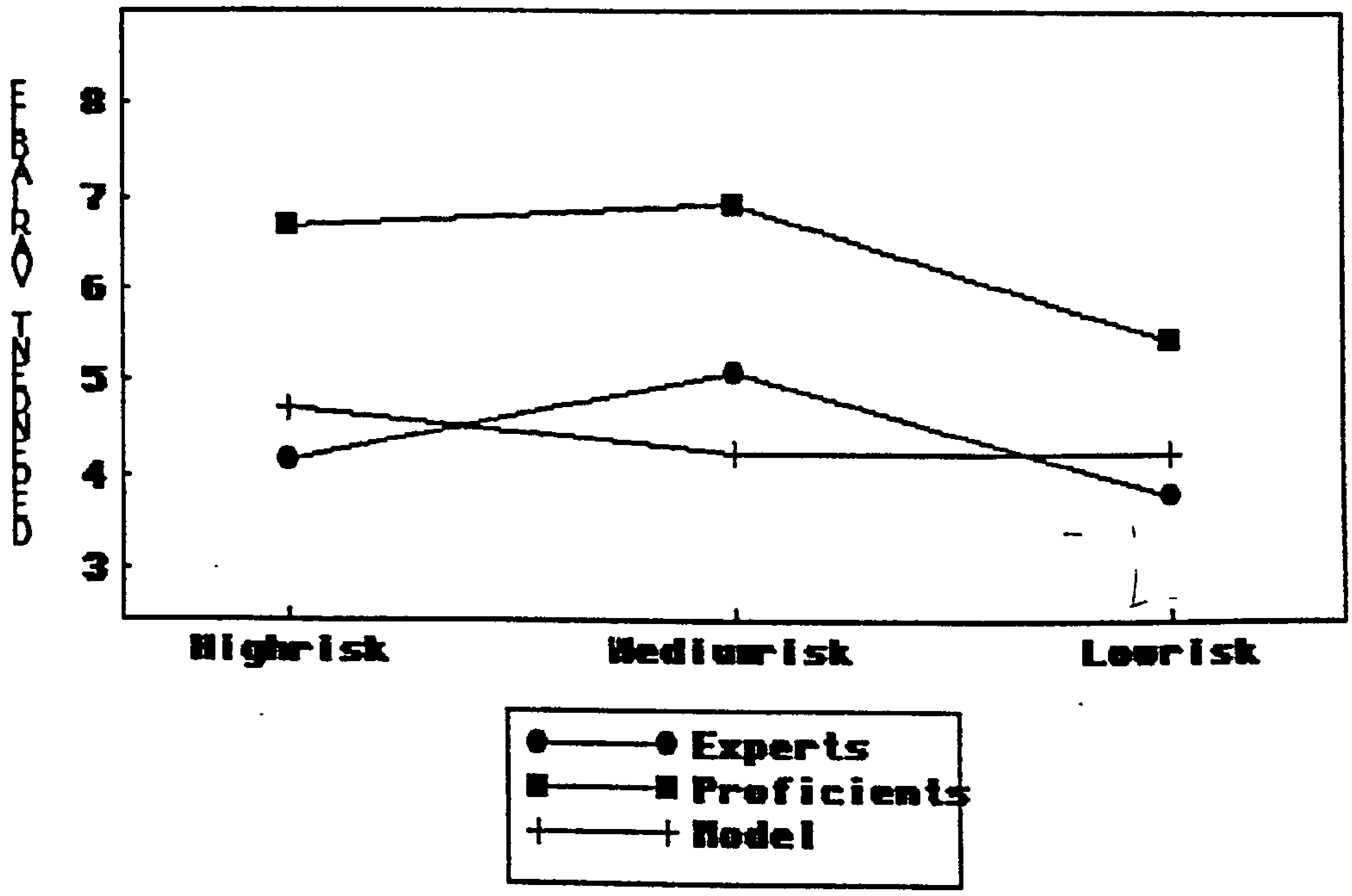
Table 3.11 Numbers of attributes selected by Model, and mean numbers selected by Expert and Proficient nurses

patient	Expert nurses	Proficient nurses	Cognitive Model
1	3.8	6.3	5
2	3.6	6.1	5
3	6	6.7	4
4	4.6	8.3	4
5	4.4	6.7	5
6	2.2	5.7	5
7	4	6.4	4
8	4.6	6.7	5
9	6.8	7.8	4
10	6	7	5
11	3.2	5.4	3
12	5.6	7.8	4
13	3	4.6	4
14	3.8	5.4	5
15	3.8	6	4
16	3.6	4.1	4
17	4.4	6.7	4
18	3.8	5.7	4
mean	4.29	6.30	4.33
st. deviation	1.17	1.08	.59

It can be seen from Table 3.11 that, on average, the Proficient nurses selected about 2 attributes more than both Experts and model. If the model compares most closely to the Expert group then there should not be significant differences between the scores for the Expert group and the scores for the model. Results of t tests confirm that only the mean score for the Proficient group can be considered to be significantly different from other means (t values as follows: E v. P=5.36, P v. M=2.82, each $p<.01$, paired samples test).

It is also possibe to represent the mean number of attributes selected for the 3 categories of pressure sore risk. Inspection of the data, as depicted in Figure 3.6 overleaf, displays graphically the correspondence between Experts and Model with the Proficient group, by contrast, selecting more attributes across each risk category.

Figure 3.6 Number of attributes selected by Model and mean numbers selected by Expert and Proficients over 3 categories of Risk of patient

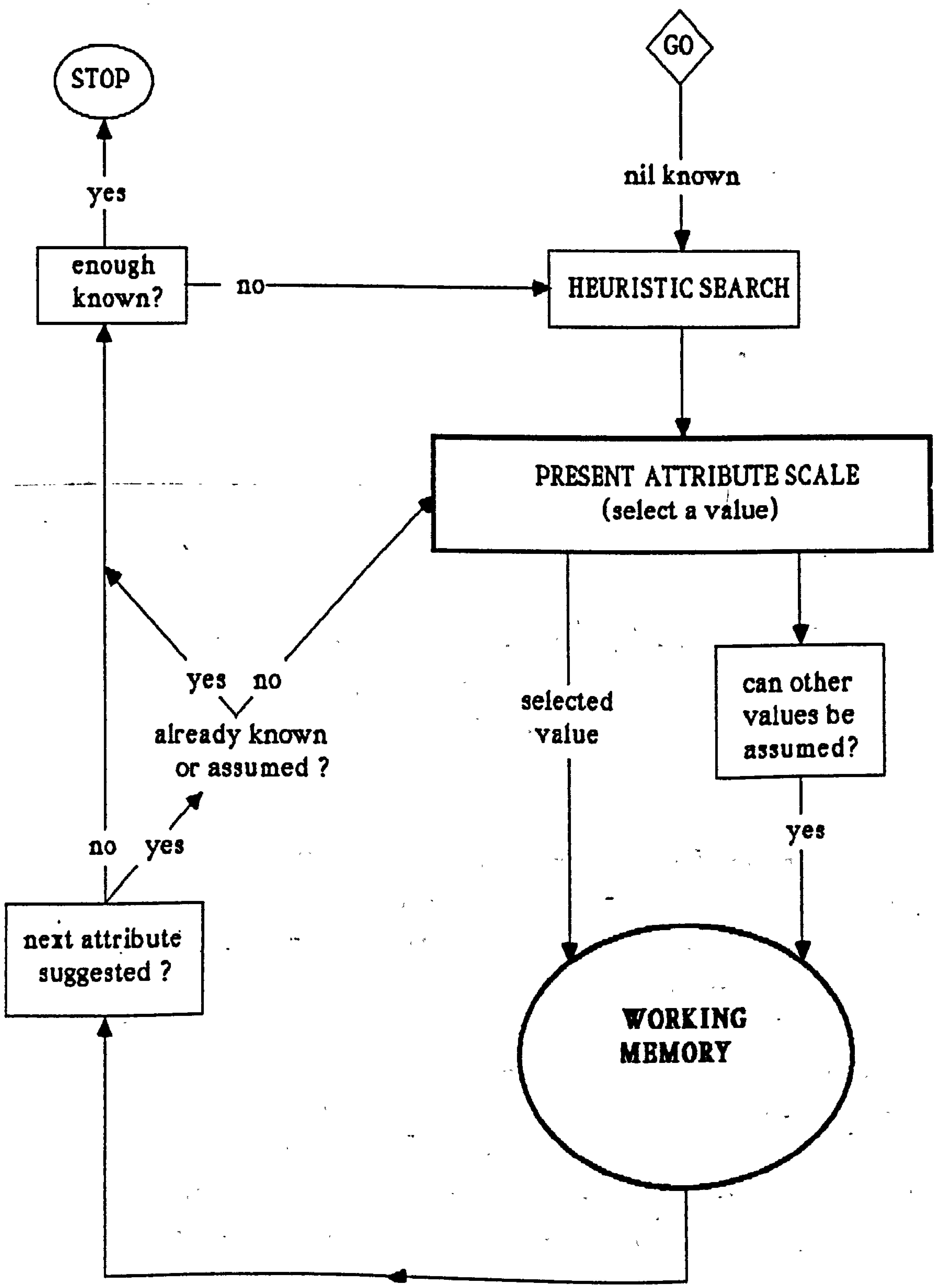


A more specific anlaysis could be performed to establish whether the Model selected different numbers of attributes than the Experts in any one of the categories of Risk. Figure 3.6 seems to suggest that some difference might exist between the Experts and the model with respect to Medium risk patients. In order to explore this possibility, an ANOVA was performed with the Proficient group scores omitted. On this occasion, the18 patients were considered in between subject blocks of high, medium, and low risk categories. Hence there were 3 levels of the between subjects variable Risk with 6 patients at each level and 2 levels of the within subjects variable Group (Experts and Model). Results confirmed the t test above with no significant main effect for Group (F=.02, df2 and 15). Results also failed to demonstrate a significant main effect for Risk (F=1.7, df2 and 15). Analysis of simple effects was performed to test the suggestion that Medium risk scores differed significantly - once again the effect was not beyond that expected by chance.

This finding, albeit using the rather coarse measure of number of attributes selected, supports the earlier finding of correspondence between the model and the Expert nurses in terms of numbers of attributes selected. In addition, these findings do not show a close explanation by the model of the number of attributes selected by the Proficient subjects.

With the adoption of the Conceptual Schema into the cognitive model, the flow diagram (which has evolved through the incorporation of Heuristic Search and then Inference) requires to be updated. The current model, then, is displayed in Figure 3.7 overleaf. Aside from the facility to make the decision regarding pressure sore risk (see next section), the main unspecified part of this model is the part labelled 'Working Memory'. The attempt to illuminate this part will occupy two experiments which are reported in Chapter 4.

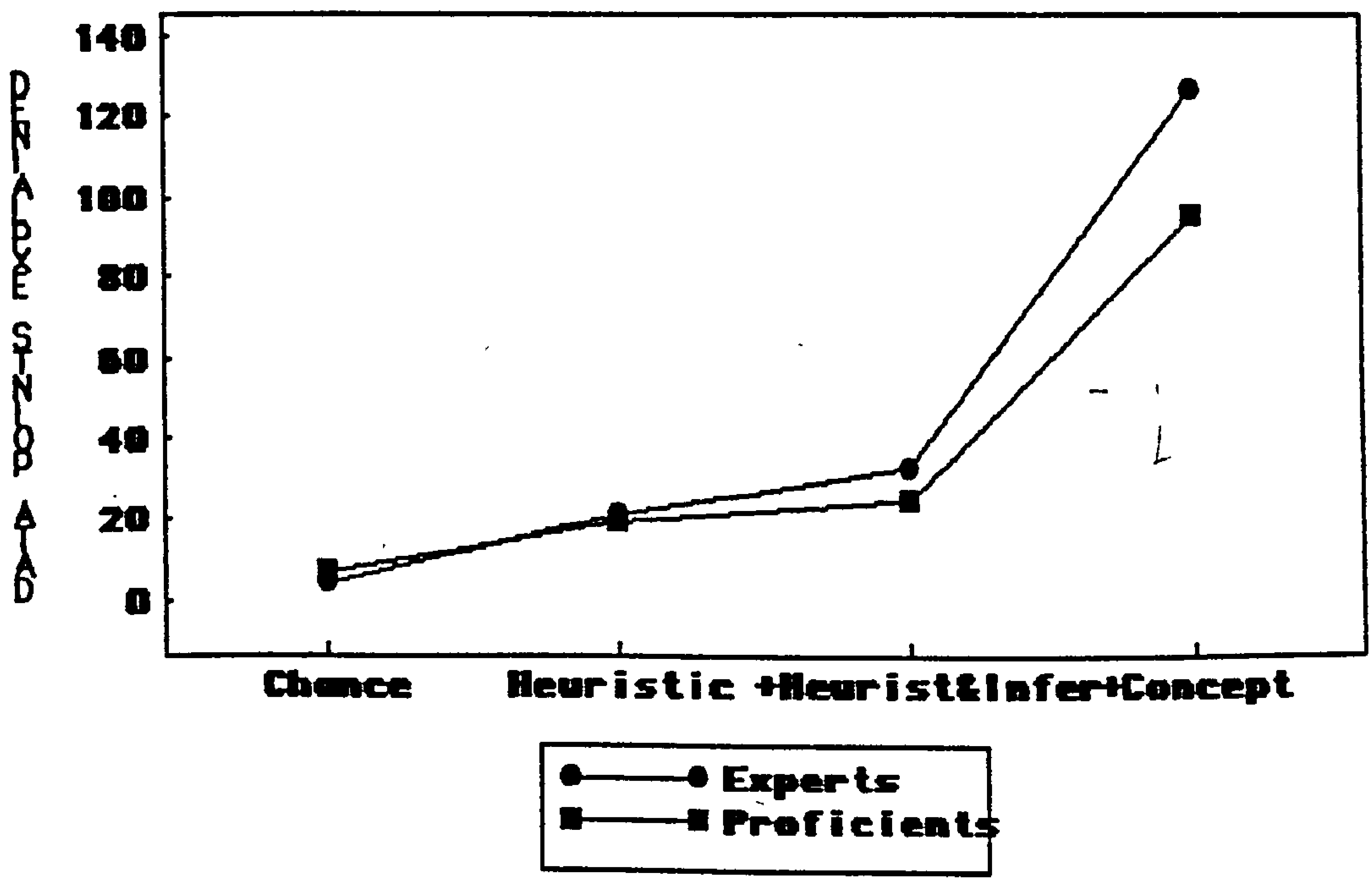
Figure 3.7 Cognitive model flow diagram incorporating Conceptual Schema



As undertaken in the previous sections, then, it is finally interesting to plot the average numbers of data points explained by the developing models from the reference point of the

Chance ranking scheme. Hence, the model has evolved from Heuristic Search ranking scheme through to Heuristic Search + Inference, and now finally (with the adoption of the Conceptual Schema) to a 'complete' model which self-starts and self-limits. The result, as displayed in Figure 3.8, shows steadily increasing point-by-point agreement.

Figure 3.8 Average number of each groups' data points which are predicted by the evolving cognitive model



The indications, then, are that the model of cognition has gone some way to the attainment of the goal of explanation of information gathering by the Expert nurses. The complexity of this task, never treated lightly, has largely been due to the difficulties of 'averaging' expertise. Hence this project has set out to construct a model which is representative of nomothetic rather than idiograhic cognitive expertise. One final test is clearly required which will evaluate the extent to which this aim has been realised. Such a test might follow from a simple observation that can be made. This observation, derived in the present context, states that the model must approximate to the group of subjects more closely than any one subject can approximate to that group. If the one subject outperforms the model in this respect, then that one subject should more properly be considered to represent the group's cognitive expertise.

This challenge to the adequacy of explanation within the model is perhaps not entirely justified given that the model contains specified information about its processes in excess of the specificity of the information which could have been gained from any one subject. Moreover, the ability to 'explain itself' is of crucial importance to the eventual educational function of the model. The challenge, however, remains. The test of it is relatively simple - each individual Expert's matrix must be convolved with each other matrix and with the current model's matrix. The results, appropriately presented in matrix form, are given in Table 3.12.

Table 3.12 Matrix of point-by-point agreement between each Expert and model

	E2	E3	E4	E5	model
E1	119	95	110	110	107
E2		128	137	124	142
E3			124	94	132
E4				107	143
E5					102

The matrix comprises 15 pairings. These pairings can be ranked in order of magnitude of the index:

		pairing
lowest index -	1	E3 with E5
	2	E1 with E3
	3	E5 with model
	4=	E5 with E4
	4=	E1 with model
	6=	E1 with E4
	6=	E1 with E5
	8	E1 with E2
	9=	E3 with E4
	9=	E2 with E5
	11	E2 with E3
	12	E3 with model
	13	E2 with E4
	14	E2 with model
highest index -	15	E4 with model

It becomes straightforward to identify if a specific subject or if it is the model which offers the closest approximation to the group. Inspection of these ranks reveals that the 'best match' to each subject is as follows:

E1 - best match = E2
E2 - " " = model
E3 - " " = model
E4 - " " = model
E5 - " " = E2

The model, therefore, can be safely taken as providing the closest approximation to a notional 'average' expert nurse – within the limitations, of course, of the present sample, experimental task, and knowledge domain. The adoption of the Conceptual Schema into the cognitive model is satisfying on both theoretical and inspirational grounds. The theoretical basis, as reviewed above, lends epistemological credibility to the cognitive model in that the nature of nursing knowledge about pressure sore aetiology can now be accounted for. The inspiration which follows from the demonstration of the power of this explanation is, however, potentially more satisfying in that a basis is offered to the crucial issue surrounding this entire project – the nature of nurses' mental representations of their patients. It is towards the exploration of that issue that the discussion can now proceed.

DECISION MAKING PROCESS

In this section a study will be made of the process of decision making by experimental subjects. The conclusions which will be drawn will establish the method whereby decision making will be accomplished within the cognitive model.

In order to build the decision making component of the cognitive model it is firstly necessary to more closely study the decision making of the subjects. In a major sense, this has already been partly accomplished in that the evidence thusfar points to the importance of a nurse properly carrying out the information gathering phase of decision making. Decision making, fundamentally, is an information processing task. Furthermore, the suggestion has been made that fundamental to the information processing task which this project studies is the finding that nurses are not just gathering information in order to make a risk decision, rather they are driven by the imperative of planning care. Nevertheless, assuming the expert nurse (and the emulation of the expert nurse) arrive at the point where a risk decision should now be made, what then is the method by which that decision is to be made?

<u>Selected Literature Review on Decision Making Processes</u>

The type of decision involved in the present project fits with the class of decision which has received most attention in both the psychological and Artificial Intelligence (AI) literature - decisions which involve judgement and uncertainty. It will therefore be useful to outline the major approaches which have been taken in these fields and to relate these models to nursing. These approaches can here be characterised as probability models, cognitive processing models, computational models, and finally as knowledge-based models. The conclusion offered will be that the knowledge-based approach possesses most potential for implementation within the current model.

<u>Probability models</u>.

Probability models are taken here to represent the substantial field of normative statistical models (eg see Hammond 1980) which assume that there is an optimal mathematical way of weighing pros and cons and reaching a judgemental decision. The consistent finding from a great deal of research has been that humans depart from these 'rational' solutions through the use of various biases, deficiencies, and cognitive illusions (Slovic, Fischoff and Lichtenstein 1977). The response, for example in nursing, is to construct prescriptive statistical models of the decision process (Grier 1976). Nevertheless, as Einhorn and Hogarth (1981) point out, prescriptions vary - humans have been urged to adopt evaluation strategies such as conjunctive, disjunctive, lexicographic, elimination by aspects, additive, additive difference, multiplicative, majority of confirming instances, or random.

More recently, Simon (1978) and Einhorn and Hogarth (1981) have argued that the 'rational' basis for comparison should be redefined to take more account of the functional goal-directed nature of human decision making behaviour. A 'linear' statistical model will become part of a later comparison with the cognitive model (Chapter 4), however for present purposes the point which can be taken is fairly simple - if humans are not using a statistical probability model when making decisions then there is no basis for incorporating such an approach within the present cognitive model.

Cognitive processing models

Tversky and Kahneman (1974) pioneered the more direct attempt than offered by mathematical models to understand the cognition of humans when making judgements under uncertainty. In terms of levels of understanding, the findings that people rely on a limited number of heuristic principles such as ease of recall are fairly high level. More recent attempts have been made to achieve a lower level of description by building computer models such as production systems (Anderson 1983) which simulate the processing errors which can affect decision making. Fox (1980), for example, simulated the effect of the 'availability' heuristic in a computer decision making model and gave an apparently good account of how judgement under uncertainty can be affected simply by how the information is processed.

The importance of how information is processed and the component cognitive structures is clearly a position with which the findings of the present analyses would concur. On closer inspection, however, it becomes clear that the central issue of knowledge about uncertainty has not been explicitly dealt with in the approaches which, as Fox (1987) argues, treat management of uncertainty as a side-effect of cognitive mechanisms. To take a nursing example, nurses have been shown by Cohen and Strantz (1976) to be able to choose between actions which minimise risk to patients. The crucial issue is that knowledge of degrees of uncertainty must somehow be represented. For an outline of the variety of approaches to achieving this representation it is appropriate that the discussion now turns to the third category - Artificial Intelligence models.

Artificial Intelligence models

Cohen and Gruber (1984) point out that there are three general sources of uncertainty in decision making. A person or system can be uncertain about the quality of the evidence, about the adequacy of the model of knowledge, and about strength of beliefs about each. One major way in which the problem of uncertainty has been tackled is to take a 'control approach' in the system which aims to recognise where uncertainty will arise and reduce it through information processing strategies (eg Pople 1977). Nevertheless, the last source of uncertainty - how much something is believed - has proved to be a crucial issue in the construction of expert systems which reason within inexact domains.

AI systems designers have responded to the problem of representing beliefs about uncertainty, it can be seen, by adopting essentially numerical solutions. Hence two of the most common but related approaches which incorporate probabalistic reasoning mechanisms are based on either 'certainty factors' (eg Shortliffe 1976) or Bayes Theorem (eg Duda 1978). This latter system gives an example of another way round the problem which incorporates the user being asked for 'confidence ratings' whenever strength of belief is implicated. An essentially numerical approach is proposed in expert system for nursing reported by Ozbolt et al (1985). However, whether or not the complex mathematical algorithm involved can be seen to operate satisfactorily is missing the point, for this discussion, that there is no evidence that this is how humans cope with uncertainty.

Although Cohen and Gruber (1984) discuss one or two number-free methods based on collaborative evidence, Fox (1987) has made the valid point that AI research has missed the opportunity of representing uncertainty knowledge explicity in favour of implicit representations within an algorithm or abstract representations as numbers. As such, expert system methods which rely on numerical calculus must be placed alongside Probability models for the present purpose - of no demonstrable correspondence to humans and therefore of little use in the construction of a cognitive model.

Knowledge-based models

This approach rests on a simple but important observation by Fox (1987) that since people are clearly not using mathematics to cope with uncertainty then they must be representing it explicitly in some other way. We can all 'feel uncertain', moreover in our vocabulary there are many specialist terms to describe degrees of uncertainty such as possibility, plausibility, doubt, conceivability and so on. Any and more of these terms are associated by us to each of the types of uncertainty outlined by Cohen and Gruber (1984). Moreover, there are different distinctions within humans' strength of belief associated with any fragment of knowledge (denoted by a proposition P). Principally these are *possibility* (P is possible if no conditions which are necessary for P are violated), and *probability* (P is probable if P is possible and the balance of evidence is in favour of P). In terms of the task at hand, it is probable that a particular patient has a high risk of developing pressure sores if a high risk decision is possible and the balance of attibute values favours a high risk decision.

The chief difference between the reformulation proposed by Fox (1987) and classical probability theory is that 'balance of evidence' is not necessarily represented by humans in terms of numerical weight. This reformulation fits with Simon's (1978) idea of bounded rationality in that people, unlike mathematical decision models, may for example see something as more probable if the evidence is observational rather than circumstantial. Moreover, people well understand the distinction between possibility and probability which Bayesian theorists discount. As Adams (1976) argues, one of the principal problems with existing mathematical models is that they subsume all aspects of belief into a single concept.

Until this point the review of human and expert system approaches to reasoning with uncertainty has showed that quantitative models predominate in the absence of sound evidence that humans make decisions in what Fox (1987) terms a manner where "qualitative knowledge simply used in an ad-hoc combination with some numerical calculus" (p.203). Fox offers a fresh approach which rests on the belief that knowledge of uncertainty is held by domain experts in the form of representation of beliefs about the degree to which events are related and therefore the logical possibility and probability of conclusions. In terms of an model of expert decision making the arguement is that reasoning is a knowledge-intensive activity which does not easily lend itself to formalisation in rules or maxims - a position which fits closely with Benner's (1984) calls for a return to respect for the context-specific intuition of excellent nurses.

The conclusion, once again, is that the issue of mental representation is all important if the present cognitive model is to remain loyal to the goal of emulating human cognition rather than adopting ad-hoc solutions of little or no validity. It would not, however, be sensible to ignore the importance of information processing approaches such as that of Tversky and Kahneman (1974). With these conclusions in mind, the exploration of the present data can proceed.

<u>Exploration of Decision Making Processes within the Data</u>

If the knowledge-intensive approach has validity then by implication the nurses are forming a mental model during the assessment of the patient which incorporates relations between the patient's attribute values and a risk judgement. To take an example, 'experience' with patients who are bedfast can be thought of as mental representation of belief that such a patient will be in a particular risk category. If a nurse has never encountered such a patient who was of low risk of developing sores then in terms of her knowledge it will not be possible for this patient to be low risk but probable that the patient will be high risk. What this study has arbitrarily termed 'medium' risk may be possible and could end up becoming more probable than high risk if the nurse were to go on to find that the patient is young and fit.

The attempt to establish the validity of this approach can be carried out through a closer analysis of decision errors made by the subjects. The idea of looking at errors in order to test a hypothesis is an established one in cognitive psychology. It is apposite that two leading exponents of the approach - Tversky and Kahneman (1974) - are cited at this point since it is sensible to test their approach to understanding decision making in terms of goodness of fit to the data before making any conclusions about the usefulness of Fox's model. Therefore, before evaluating the 'knowledge representation' approach as a contender for implementation in the cognitive model, some attention can now be paid to the Tversky and Kahneman model.

The basis to Tversky and Kahneman's model is that people do not attempt to use 'proper' decision procedures but rather rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities to simpler judgemental operations. It is perhaps important to note that the great bulk of experimental work carried out to establish the heuristic model has been carried out with quantitative tasks and subjects who were inexpert in the domain.

When looking at the present (qualitative) task, there is an apparent immediate difficulty for the model in that in the ALLUP condition there were more errors in total than in the SELECT condition. Although this difference was not significant (74 versus 69 respectively out of 252 decisions in each condition), the point is that there was no evidence to support the view that rapid heuristics were being used on minimal data in the SELECT condition with the result that errors were being made when compared to the case when there was a great deal of additional information at hand in the ALLUP condition. Nevertheless, it is difficult to know to what extent this additional information was heeded.

Continuing with the Tversky and Kahneman model, then, perhaps the most appropriate 'rule of thumb' which subjects in the present experiment might have been using would be the anchoring and adjustment heuristic, where an initial 'guessimate' of risk is made followed by progressive but inadequate adjustment as new facts are uncovered. Once again, an apparent difficulty arises when comparing just two patients who were assessed - patients 8 and 9 who each were MOBILITY - bed/chairfast with short assisted walks. It is possible to estimate the probability of patients with this attribute value being of either a High, Medium, or Low risk using Bayes Theorem for determining conditional probabilities.

Bayes Theorem is given as:

$$p(r \mid v) = \frac{p(r)\, p(v \mid r)}{p(v)}$$

where $p(r \mid v)$ is the conditional probability of, for example, Low risk given the existence of the attribute value 'bed/chairfast with short assisted walks'; $p(r)$ is the baserate probability of risk being Low; $p(v \mid r)$ is the probability of the attribute value within the Low risk population; $p(v)$ is the baserate probability of the attribute value.

The task now requires a database of attribute values. Since the reliability of the risk judgements in that database is important for the analysis to be undertaken below, for this reason the database2 will be employed. Nevertheless, the potential for error is not being disregarded - judgements about risk can never be reliable at the level of irrefutable facts. For this and other reasons to be explained below the decision making process must necessarily retain a probabilistic element.

Application of Bayes' formula to database2 with respect MOBILITY - 'bed/chairfast with short assisted walks' results in probability estimates which are almost unequivocal - .30 for High risk; .79 for Medium risk; and .05 for Low risk. Indeed this attribute value (and therefore patients 8 and 11) has been chosen for this preliminary analysis simply because of these unequivocal probability estimates. Hence, if a nurse begins the assessment of these patients in the SELECT condition by finding out what the value of MOBILITY is, then if that nurse is using a heuristic approach the rule of thumb will point strongly to 'Medium risk' and further adjustment from this judgement will not be radical.

In fact, 13 nurses selected MOBILITY either first or second when assessing these two patients in the SELECT condition. It is possible to categorise the final decisions of these nurses for these patients and then to look at decisions arrived when the same patients appeared in the ALLUP condition.

The picture which emerged was:

pt.	SELECT CONDITION final decision = medium	final decision = high	ALLUP CONDITION final decision = medium	final decision =low	final decision = high
8	10	3	2	0	11
9	8	5	8	2	3

What is apparent is that for patient 8 there were no less than 8 nurses who changed their decision from Medium to High between SELECT and ALLUP conditions. Figures for patient 9, however, remained fairly constant. Given that all SELECT condition assessments began with searching the MOBILITY attribute, an explanation is required of this apparent contradiction. The key differences between these patients were that although both patients were 'bed/chairfast', patient 8 had further 7 'problem' attribute values whereas patient 9 had only another 2 risk-enhancing values.

For patient 8, clearly many of these were missed in the SELECT condition. However, when all were there to see in the ALLUP condition there were only 2 nurses who arrived at a Medium judgement. Anchoring and failure to adjust, on these examples, is not apparent. Moreover, there is some evidence that nurses are heeding all of the attribute values on display in the ALLUP condition. The task now is to undertake a more complete analysis of these points.

Specific Testing of Decision Making Process Explanations in the Data

More detailed analysis of errors can be undertaken by identifying all the occasions when a subject gave a different risk decision in the SELECT condition as compared to the ALLUP condition. It is necessary to define 'errors' in this fashion since it could be that a decision given in the ALLUP condition disagrees with the decision pre-determined by the nurses who originally cared for these patients, nevertheless the subject firmly believes her judgement to be correct on the basis of the available information in the ALLUP condition. There were 24 such cases of SELECT condition error. The task now becomes one of trying to find the most likely explanation for these within-subject deviations.

Three principal explanations for these errors will be entertained:

1. The decision made a subject is **rational** given the information gathered. By this it is meant that, for example, a subject elicits only the non-problem information about a patient and, having missed the problem attribute values, gives a 'correct' Low risk decision.
2. The decision is best explained by **anchoring**, ie the final risk judgement given by a subject

closely corresponds to the risk associated with the first value elicited and is not subsequently adjusted to the extent that new information gathered should demand.

3. The decision is a 'genuine' **error** in that it defies explanation by other means and could possibly correspond to factors such as inattention, pressing the wrong button, or faulty knowledge held by the nurse.

In order to determine the appropriate decision based only on the values selected (explanation 1 above) or at the point of only 1 value having been selected (explanation 2), a method based on Fox's (1987) approach will be adopted. Fox would argue that the nurses hold logical beliefs about probability of risk. The key to this approach is that attribute values will be associated with risk categories in a 'frequentistic' fashion – the extent to which these values when encountered in the past have been present in, for example, a Low risk patient. Hence, by reference to the database2, the most probable decision given the values selected can be identified and compared to the decision arrived at by the subject.

The method of achieving this will be as follows. First, if the initial attribute(a) value(v) elicited is A1v3 then all patients in the database2 who are A1v3 will be isolated and the greatest risk category membership identified. If the next attribute value elicited is A5v2 then the procedure is simply to isolate all patients who are A1v3 and A5v2. This proceeds until only the patients who possess exactly the attribute values elicited by a particular nurse are isolated – whereupon the most frequently occurring risk judgement will be identified. In a sense, database2 is being likened to a nurse's memory store of patients previously encountered.

The 24 cases of apparent error can now be considered individually. Results, presented in Table 3.13 on page 194, are of great interest in terms of deciding between competing explanations of the process of decision making. From 24 apparent errors the number of *genuine* errors has been reduced to only 7 since one or both of the other explanations seems to fit the data. The *rational* explanation was found to posess most power in that it alone explained 8 'errors' and was a possible explanation in a further 7 cases. The *heuristic* (anchor) explanation, the other contender in these 7 cases, could nevertheless achieve sole explanation in only 2 cases.

A good example of the rational explanation is B5 *2 – where the Medium risk decision given by the subject seems at first sight to be indefensible given that the patient is both bedfast & immobile in bed and mildly disorientated. The point, however, is that the subject failed to elicit these very high risk attribute values. On the basis of the information she did gather, therefore, her rather disastrous Medium decision turned out to be quite rational.

The columns in Table 3.13 overleaf which require elaboration are:

subject/patient – the codes of the subject and patient involved. E=expert, B=beginner, and S=the subjects who were not put into either of these groups

pre-risk – the predetermined risk judgement for that patient

problems missed – the number of risk enhancing attribute values which the subject failed to elicit

subject's decision – risk judgement given by the subject in SELECT condition

MOST LIKELY RISK, after Av1 – most likely risk after subject's first attribute value, determined through the procedure described above,

MOST LIKELY RISK, finally – most likely risk given all the attribute values elicited by the subject for that patient,

comment / best explanation – which of the 3 explanations can be reasonably taken as an explanation of the observed error.

Table 3.13 Details and best explanation of subject's decision errors

subject/ patient	pre-risk	subject's decision	MOST LIKELY RISK after Av1	finally	problems missed	comment/best explanation
E1 #4	H	L	L	L	3	all Avs elicited were low risk, s. missed all high risk values. Explan. rational or anchored
E1 #7	M	L	L	L	2	as for E1 #4
E3 #1	H	L	H	M	4	explanation = error
E4 #1	H	M	H	M	5	despite values missed, best explanation = rational
E4 #10	M	H	M	H	0	explanation = rational
B2 #4	H	L	M	M	2	explanation = error
S1 #4	H	M	M	M	0	explan. rational or anchored
S1 #17	L	M	L	M	0	explanation = rational
B4 # 4	H	M	M	H	0	explanation = anchored
B3 #4	H	M	M	M	1	explan. rational or anchored
S2 #9	M	H	M	M	0	explan. rational or anchored
S2 #12	M	H	H	M	1	explanation = anchored
B1 #1	H	M	H	H	3	explan. = error, also missed serious attributes
B1 #3	H	M	H	M	2	explanation = rational
B1 #9	M	H	M	M	1	explanation = error
B1 #12	M	L	M	M	1	explanation = error
B5 #2	H	M	M	M	2	explan. rational or anchored
B5 #9	M	H	L	M	0	explanation = error
B6 #1	H	M	H	H	5	explanation = error
B6 #3	H	M	H	M	2	explanation = rational
B6 #10	M	L	L	L	1	explan. rational or anchored
B2 #14	L	M	L	M	0	explanation = rational
B2 #17	L	M	L	M	0	explanation = rational
B7 #12	M	L	M	L	2	explanation = rational

Implications for the Cognitive Model of the Decision Making Process Explanation

The implications from these findings are considerable. Firstly, the heuristic model of decision making has failed to make a significant impact as an explanation in this task. Secondly, the knowledge-based representation approach of Fox has demonstrated power as an explanation of errors, and, therefore, deserves to be further evaluated as the decision making mechanism within the cognitive model. Representations of patients seem not to be constructed around single 'diagnostic' cues, rather an elaborated model constructed form deeper knowledge of the patient as a 'whole' is suggested. There is considerable theoretical (eg Benner 1984) and empirical (eg Broderick and Ammentorp 1979) support for this conclusion in the nursing literature.

This decision making technique can be seen as representing a 'weak' method compared to, for example, expert systems which are based on Bayes Theorem. The principal difference between the cognitive model and a conventional expert system is that the cognitive model does not assess patients solely with the goal of making a risk decision. The nurse, and the model of the nurse, has been conceptualised as processing information in order to construct a 'care-implicating' mental model of the patient. It follows, therefore, that the co-existing representation of this patient's risk of developing sores will not be a fine-grained statistical representation of gradually altering probability values (as in a Bayes model) or altering certainty factors (as in the AI approach).

At a general but nevertheless important level, the implications of the findings above are yet more support for the view that patient assessment and decision making is an information processing task. Three components are both necessary and sufficient for expert performance of this task - the right data, the methodical gathering of these data, and the possession of an adequate knowledge base from which the relationships between these data can be recovered. Given that the first two components of this information processing task have been specified in the current cognitive model (Figure 3.7), the task is now to incorporate the knowledge based method of decision making. Assessment of the power of this explanation of decision making processes can then be undertaken.

Procedure for specifying decision making by the cognitive model

The adoption of the knowledge based logical probability approach requires firstly that the 'paths' through patient assessment which the model takes are specified. This has been achieved by running the current computer-based cognitive model exhaustively through every possible attribute value combination (or 'paths') until the model announces it has gathered sufficient patient details, ie each dimension of the Conceptual Schema has been adequately searched. The resulting paths, set out in Appendix 5, amount to some 612 separate paths but in fact only 32 principal paths are followed. This is due to the occasions when the same successor attribute is searched following more than one value of another attribute, hence if BUILD is being searched and the user selects 'within average limits' or 'slightly underweight' then the next attribute searched will be NUTRITIONAL STATE. This counts, however, as one path.

These 32 paths, therefore, were then written into an SPSSx program which in turn analysed database2 of n=159 patients. The method adopted was to identify the frequency of risk decisions at the 'stop points' in the assessment. Hence, on the first pass through the database, all patients whose attribute values fitted Path 1 were identified. These patients were then categorised by risk decision. The procedure was then repeated until each path had been covered.

As explained above, the key to the knowledge based approach is that attribute values will be associated with risk categories in a 'frequentistic' fashion - the extent to which these values when encountered in the past have been present in, for example, a Low risk patient. Hence, when the model's paths were followed through database2, the frequency of risk decisions of the identified patients could be further classified using only two terms - 'probably' and 'possibly'.

It would have been possible to proceed no further with the analysis and simply take this information to the cognitive model in the form of a set of IF (attribute values) ...THEN (risk) rules. Although the model would then possess a decision making mechanism, it is clear that the spirit of the knowledge based approach would not have been followed, ie a solution which tended toward the 'AI pragmatic' had been adopted. The term 'cognitive model' implies model of cognition, therefore if proper claim is to be made to the term it becomes necessary to fully incorporate the findings from the analysis of database2 into a model of knowledge as held by the expert nurses.

The model of knowledge of pressure sores which has been adopted into the cognitive model is the Conceptual Schema discussed in Part 3 of this chapter. This model demonstrated a powerful explanation of the necessary and sufficient conditions for information search; it follows that an exploration of its potential to explain the decision making process should be undertaken. What is being suggested is that beliefs about the degree of risk are represented in reference to deep level knowledge of pressure sore aetiology. The procedure for accomplishing this exploration was straightforward in that each of the 4 dimensions of risk within the Schema could be 'scored' with respect to all 32 paths to stop point.

Hence, for example, MOBILITY is the header attribute for all paths. Of the 159 patients in the database there were 59 who were MOBILITY - 'bed or chairfast'. At this point the only dimension which can be scored is Mobilising; risk frequenticity which was found for these patients was High (45 patients) and Medium (14 patients). The model, if asked to give a risk judgement at this point, could therefore answer 'probably High, but possibly Medium'. When the other 3 dimensions are considered, only one set of circumstances subsequent to 'bed or chairfast' was found to reverse this frequenticity order - of the patients who were mentally alert (ie Capacity dimension ok) and had no risk-enhancing problems within the Intrinsic Factors dimension (skin, nutritional state etc), there finally resulted in a majority of Medium risk and a minority of High risk.

Following similar analysis of all 32 paths, a set of conditional rules were compiled to represent pressure sore risk within the cognitive model. The conditions related to the 'state' of each dimension. When implemented in the model, the rules can output a risk judgement at key junctures during the various information-seeking 'routes' which the model can follow. These junctures are the points at which there is a shift in the current decision probabilities. A judgement may be in effect from after the first dimension searched (Mobilising) until the conclusion of assessment of a patient. On some occasions, however, a shift in judgement occurs after, for example, the Extrinsic Dimension has been searched if the attribute value elicited was sufficiently informative to alter the current judgement. An example would be the following route:

<u>after eliciting that</u>	<u>risk</u> =
MOBILITY = bedfast with free movement in bed	prob M, poss H
+ URINARY CONTINENCE = continent with nurses' help	unchanged
+ MENTAL STATE = alert and orientated	unchanged
+ NUTRITIONAL STATE = seems adequate	unchanged
+ SKINTYPE = fine & delicate - 'papery'	prob H, poss M
	final decision = <u>High risk</u>

This example reflects the associations which were derived for 'bedfast with free movement in bed' patients within database2. Most frequently, such patients were judged as Medium risk. However, it was found through the analysis of these patients that Medium risk was most likely <u>conditional</u> on the Capacity and Intrinsic Dimensions being favourable. Hence, only if these patients were alert mentally and had no adverse Intrinsic Factors was the Medium classification most probable. If any other circumstances prevailed then the most likely judgement was High risk.

The set of conditional rules based on Dimensions which were derived are set out below:

CONDITIONS FOR DIMENSION	LOGICAL PROBABILITY
IF MOBILISING = bedfast & virtually immobile in bed	prob H, poss M
+MENTAL CAPACITITY = alert and orientated	
+no EXTRINSIC problem (ie patient not incontinent)	
+no INTRINSIC problems	prob M, poss H
else	prob H, poss M
IF MOBILISING = bedfast with free movement in bed	prob M, prob H
+MENTAL CAPACITITY = alert and orientated	
+no INTRINSIC problems	prob M, prob H
else	prob H, prob M
IF MOBILISING = bed or chairfast with short assisted walks	prob M, poss H or L
+MENTAL CAPACITITY = alert and orientated	
+no EXTRINSIC problem	
+no INTRINSIC problems	prob L, poss M
IF MOBILISING = bed or chairfast with short assisted walks	prob M, poss H or L
+ EXTRINSIC problem exists	
+MENTAL CAPACITITY = alert and orientated	
+up to 1 INTRINSIC problem	prob M, poss H
IF MOBILISING = bed or chairfast with short assisted walks	prob M, poss H or L
+ no EXTRINSIC problem	
+up to 2 INTRINSIC problems	prob M, poss H
else	prob H, poss M

IF MOBILISING = fully ambulant. Restrictions few if any	prob L, poss M
+MENTAL CAPACITITY = alert and orientated	
+up to 2 INTRINSIC problems	prob L, poss M
else	prob M, poss L

With the analysis complete it becomes necessary to incorporate the findings and suggestions into the current cognitive model in order to evaluate, once more, the goodness of fit to the data.

Evaluation of the explanation of decision making by the cognitive model

Whereas testing of fit between model and data had up until now employed measures relying on process comparison, the requirement now is for a measure of product or outcome of decision making. A measure must be made of decision concordance between the model and the subjects - the extent to which the model agrees with the risk judgements arrived at by the subjects. In the next chapter a product evaluation will be performed using a more realistic test - when a completely unseen set of patients are assessed by the model. For now, however, the focus remains on the goodness of fit between the model and the subjects from which the model was constructed.

The literature on measures of decision concordance, as reviewed by Kazdin (1983), reveals several approaches to establishing indexes of agreement between raters. The index used, ideally, would take account of the relative frequecies of each risk judgement - if 'Low risk' is very common among the raters decisions then high agreement is less impressive than if Low risk was rare. A simple 'percentage agreeing' index can therefore be misleading. An index of concordance, then, would ideally have a component built in which takes into account agreement expected by chance. The index of choice in this circumstance has been provided by Cohen (1965). The coefficient, known as Kappa (K), has been generalised by Fliess (1971) to give a statistic and reliability measure for agreement over and above that expected by chance when there are several categories for jugements. The following calculations were performed using a computer program based on a listing supplied by Jackson (1983).

Decision concordance with the model can be calculated firstly with respect to the predetermined decisions for the 18 patients (6 each of high, medium, and low). Results, as set out in Table 3.14 below, show the number of judgements from each source in each risk category, the number of judgements agreeing by category, and the Kappa values with corresponding probability of the observed agreement occuring by chance.

Table 3.14 Decision concordance - Model with predetermined risk decisions

	number of judgements		n judgements		
Risk	predetermined	Model	agreeing	Kappa	p
LOW	6	5	5	.87	ns
MEDIUM	6	4	3	.45	ns
HIGH	6	9	6	.66	ns
overall	18	18	14 (77.8%)	.66	<.001

It can be seen from Table 3.14 that the model agreed with predetermined decisions on 14 out of 18 occasions, which represents a Kappa of .66 ($p<.001$). It is interesting that the model tended to 'overpredict' when compared to the predetermined decisions. Hence, the 6 predetermined High risk patients were given 'High' by the model and in addition a further 3 patients were judged as High. That each category of risk failed to achieve agreement at a level beyond that expected by chance is a reflection principally of the low numbers.

The next, and more interesting, step is repeat the analyses for agreement between the model and the groupings of nurses. To achieve this it was firstly necessary to establish the mode decision arrived at by each group. Hence if one Expert subject gave a low risk judgement for patient x and the other 4 gave a medium risk then the medium risk judgement was taken as the mode. On no occasion was a mode decision endorsed by less than 3 out of 5 Expert nurses or 4 out of 7 Proficient nurses. Thereafter, concordance was measured between model and Expert group and between model and Proficient group. Results of these calculations are given in Tables 3.15 and 3.16 respectively (see overleaf).

Table 3.15 Decision concordance - Model with Expert group

	number of judgements		n judgements		
Risk	predetermined	Model	agreeing	Kappa	p
LOW	6	5	5	.87	ns
MEDIUM	4	4	3	.68	ns
HIGH	8	9	8	.89	ns
overall	18	18	16 (89%)	.82	<.0001

Table 3.16 Decision concordance - Model with Proficient group

	number of judgements		n judgements		
Risk	predetermined	Model	agreeing	Kappa	p
LOW	7	5	5	.75	ns
MEDIUM	6	4	3	.45	ns
HIGH	5	9	5	.53	ns
overall	18	18	13 (72%)	.58	<.001

It can be seen from Tables 3.15 and 3.16 that both percentage agreement and Kappa coefficients demonstate higher agreement between the model and the Expert group than between the model and the Proficient group (Kappa = .82 versus .58). Unfortunately there is no procedure for testing if this difference is reliable, hence these findings should properly be taken as descriptive data. It is also noteworthy that the agreement between model and Experts was stronger than between model and predetermined decisions. The model is after all aimed at being an emulation of the Expert nurse subjects; therefore in identical conditions of sequential attribute selection it is satisfactory that model and Experts perform similarly. A last point concerns the finding that each category of Risk failed on any occasion to show agreement beyond that expected at the 5% level. No conclusion can be taken from this given the low number of items.

The cognitive model, therefore, is now complete in the sense that it possesses all main components necessary to allow it to stand alone in the form of an ICAL consultative teaching system. That is, a learner nurse could input patient details as requested by the system and receive a commentary on the processing that is being used and the rationale for the decision made. To a lesser extent, information has been gathered about the possible less-than-expert information processing styles which could eventually form a student modelling module.

There are, however, issues to explore which are more immediate for the present project before considering an effort to achieve these ambitions. These issues are the broadening of the evaluation of the model and exploration of the knowledge representation question. Before moving to a chapter which sets out to accomplish this, however, the opportunity can be taken at this point to give a brief overview of some of the limitations of the cognitive model as it currently stands.

It was mentioned earlier that one of the justifications of using a top down approach of looking for 'principles' in the data was that the problem could be overcome of not having a database in the process tracing experiment of patients representing every legal attribute value combination. There are, to be blunt, a possible 100,000 or so attribute value combinations (ie patients), and this model is based on 18 (although 18 patients each assessed twice can be regarded as an improvement on the more usual case when single patients are used in simulation studies).

No experiment, however, could attempt to incorporate the full set of patients. The top down approach has considerably reduced this problem - for example goal-directed processing results in not nearly all attributes being selected. However there remains a group of 'possible' patients who are not directly responded to by subjects. It is perfectly possible that the representative sample of patients used will result in failing to uncover valuable 'special case' knowledge.

Several other but probably not all objections have been anticipated at various points in this document. The experimenter is only too aware of these limitations. The most important defence of the analysis and of the resultant model is the exploratory nature of the work. In Expert Systems terms, the model could be analogous to 'rapid prototyping' where a system is quickly built (to impress the customer) and later refined. From the point of view of the experimenter this analysis has been far from rapid, but it is possible that the entire project may be a valid exercise in hypothesis generation - if so then much in the way of continued work is required. More tightly-controlled and narrowly-focused experiments could be designed to test and refine the predictions of the model; two of these will be reported within Chapter 4.

In defence of the model it could be argued that the nature of the real-life task and more particularly the nature of the experimental processing tracing task seems to demand a conscious mode of processing from any subject, no matter their expertise. Nevertheless, the author is aware of the validity of these possible weaknesses and would offer once more the defence that the work has been exploratory. Furthermore, if the entire model were to be recast into a Parallel Distributed Processing framework (McClelland et al 1986) then the distinction between higher and lower level processing could become less important - deductive assumptions, for instance, could be explained in terms of interaction between mutually-activated nodes. This shift would be interesting and doubtless worth pursuing, however it is felt that the ultimate purpose of the model (an Intelligent Teaching System) will be more ably served by preserving the distinction - students may benefit from being able to witness 'expert' processing demonstrated in lower and higher level terms.

Finally, mention should be made of the facility where an attribute scale is presented with the range of possible values offered for the user to make a choice. This would seem to be an artificial and perhaps unsatisfactory aspect of the model - analogous only to the situation where an expert was being consulted about a patient and where the questioner was unable to generate an answer which could be mapped on to a single value held by that expert. The expert may at that point run through all possible values in order to offer choices. These choices would conform to the expert's representation of possible values that the attribute concept possesses. An intelligent front end with natural language interface could fulfil the role of mapping user inputs to appropriate attribute value. A future development, perhaps.

For this (and other) reasons, it becomes vital that the next phase involves a multi-dimensional evaluation.

CHAPTER 4 EVALUATION OF COGNITIVE MODEL PERFORMANCE

INTRODUCTION

When evaluating performance of a model it is important to specify the benchmarks against which measurement is taken. The earlier literature review found that principal measure used in expert system evaluation has been 'classification accuracy' while the principal benchmark employed has been expert human performance. Hence in the classic case the diagnosis arrived at by a system when presented with a test set would be compared to the diagnosis decided on by expert physicians. Nevertheless, the system which has been constructed in this project differs from the classic system in the important respect that it has been designed as a model (or simulation) of expert nursing cognition. It follows, therefore, that the present evaluation must not only employ additional benchmarks but also look at additional facets of evaluation.

In this section, therefore, a more complete evaluation of the performance of the cognitive model will be undertaken. 'Complete' refers to the position in this project where it is argued that an evaluation must take into account all phases of an ICAL project (Hyslop, Jones and Ritchie 1987). One of the main justifications for this approach arose from the review of literature where the focus was on the problems inherent in evaluation of educational benefit. As one means of getting round these difficulties, the use of reliable methods in the knowledge acquisition, experimentation, and analysis phases of the present project have each been considered as integral to the overall evaluation. For this reason, efforts have been made throughout the analysis sections to report, with reliability measures where possible, the extent to which the derived model 'fits' expert cognition. Since these other crucial phases of evaluation have been completed, therefore, the requirement becomes that comparative evaluation between the model and other models is undertaken (see Part 1).

The second focus of this chapter will be on the prediction which has arisen from the analyses within Chapter 3. The prediction focuses on the suggestion that there is a 'careplaning' basis to much of the expert cognition studied. Alongside this prediction there were instances identified which called for research into the area of nursing representation of patients and of assessment knowledge. It was suggested that the mental representation of a patient which the nurse acquires seems to be constructed around 'care concepts' and that this ran counter to much of the prevailing nursing theory on patient assessment. Experiments will therefore be reported which explore further this area (see Part 2).

COMPARISON OF COGNITIVE MODEL WITH ALTERNATIVE DECISION MODELS

In keeping with the theme of this project, comparison will be undertaken of both process and product of decision making. Process refers to 'route' taken through a given patient's assessment details while product simply denotes the outcome of that route – the decision arrived at.

Process comparison

As previously, this will be undertaken at both a coarse and at a more specific level.

Coarse level of comparison – the measure employed will be the familiar one of number of attributes selected.

Specific level of comparison – two measures can be employed to compare the order in which information is selected by models. Firstly, the point-by-point index will be employed to provide some continuity with the earlier analysis section. Secondly, a different index based on correlations between process traces of models will be employed. This approach, suggested by an authority on non-parametric statistics (Ray Meddis, Senior Lecturer at Lancaster University; see Meddis 1984), involves computation of Spearman's Rho correlation between each model's ranked order of information selection where non-selected attributes are designated with the high rank of 100.

In practice, process comparison can only be undertaken between the cognitive model and the decision model built from the Automated Rule Inducer (ARI). This is due to the Discriminant Function model (DFA) being a classic 'proper linear model' in Dawes' (1979) terms. Hence, the DFA model not only utilises all available attributes but also makes no distinction between the order in which the attributes are input to its statistical formula. As will be discussed more fully below, this lack of correspondence between linear models and human cognition in terms of process of decision making acts to seriously reduce any claim that nurses might process information in linear fashion.

One final point should be made concerning process comparison. Since data is required of human performance against which comparison can be made, the performance of the cognitive and ARI models with respect to the 18 patients assessed by the experimental subjects will be utilised.

Product comparison

This form of evaluation, as discussed above, corresponds to the more traditional comparison of the 'decision accuracy' of each model when presented with a test set of cases. The basic plan is to take a test set of patients which are entirely new to each model (ie these patients are distinct from the training set of patients who were used in the construction of the models) and to measure and compare the performance of each model in terms of classification accuracy. An appropriate test set is simply those patients who comprise database1 and who were not used in the process tracing experiment. Hence, the 18 patients used were withdrawn from the cohort of 152 in database1 – leaving a test set of 134.

Background and Construction of the Decision Models for Comparison

In this section an account will be given of the two models chosen for comparison and the rationale for their choice. Firstly, however, it is necessary to discuss the 'training set' of example patients who were used to construct these models.

Establishment of the training set

Some important points affect the decision of how best to construct the alternative decision models. Firstly, in the ideal situation all three models would have been constructed using the same training set. This is not possible since the central model to the project – the cognitive model – was constructed from a basis of only 18 patients. In the situation where there are perhaps 100,000 possible combinations of attribute values (ie 100,000 different patients), both ARI and DFA would require a much larger representative sample from which to construct models.

Given that ARI and DFA models require a larger training set, therefore, a second point can be made about the possibility of 'unfair' advantage being afforded one or other models. Hence, if the ARI / DFA training set comprised a large number of 'atypical' patients which the cognitive model would fail to accurately classify then the resultant decision models would themselves be atypical. The possibility would therefore exist that a test set of 'typical' patients would be poorly classified by such models.

The solution which overcomes these potential problems is to construct a training set for ARI and DFA of patients who are correctly classified by the cognitive model. Hence, the 159 patients in database2 were input one by one to the cognitive model. The risk decisions given by the model for each patient were then compared to the decisions recorded by the nurses who cared for these patients. As Table 4.1 shows, the cognitive model correctly classified 123 of these patients. Also given in Table 4.1 is the percentage agreement and Kappa values overall and by risk classification. The cohort of 123 patients, therefore, became the training set for the construction of the ARI and DFA models.

Table 4.1 Decision concordance - cognitive model with patient database2

	number of judgements		model agree with db			
Risk	database	Model	n	%	Kappa	p
LOW	42	44	36	85.7%	.81	<.001
MEDIUM	56	43	35	62.5	.57	<.01
HIGH	61	72	52	85.2	.62	<.01
overall	159	159	123	**77.4%**	**.66**	<.0001

With the composition of the training set now established, an outline of each model and its rationale for selection in this comparison can be given.

Model 1 - Discriminant Function Analysis (DFA)

Discriminant Function Analysis, first introduced by Sir Ronald Fisher, is a statistical technique for predicting group membership of cases on the basis of 'predictor' variables. The basic approach is to find an optimal statistical relationship between cases for whom group membership is known. In other words, a formula is arrived at using data which are the values of numerical predictor variables for each known case. Group membership of unknown cases can then be estimated simply by inputing the values of predictor variables of these cases to the formula. The linear model formula usually comprises the additive combination of numerical values for each variable which have been transformed by weighting coefficients. Put crudely, the formula to predict size of house (small versus large) might be 'number in family' X 0.345 + 'income' X 0.79 + 'proportion who are same sex' X -0.46. It becomes possible to predict the size of Family X's house by running this formula and finally comparing the output with a predetermined cut-off. Hence a final value of greater than 3.5 might predict 'large house'.

This approach has been chosen since it represents the type of 'linear' decision models which have been popular (at least until fairly recently) since Meehl's classic book on clinical versus statistical prediction published some 34 years ago (Meehl 1954). The main argument of the approach is that a linear model (eg DFA, regression analysis, Bayes Theorem) will outperform in classification tasks the humans who are skilled in this type of prediction but who use clinical intuition. Dawes (1979) has defended the view that "human judgement can't be systematised" by listing the superior performances of linear models over humans and urging that a greater number of 'important social decisions' are taken on the basis of statistical decision models.

As many of the papers reviewed elsewhere in this thesis testify, however, there has been a general shift away from linear models and toward knowledge based (or symbolic reasoning) approaches to decision making. Shortliffe et al (1984) note this trend in medical decision support computing and conclude that it is only partly explained by movements in performance goals toward the more qualitative and fuzzy medical reasoning problems. The other aspect of strong importance is the issue of human acceptance of these systems.

Nursing seems to have come to numerical decision models around the time when the trend was tending more toward psychological approaches. Hence Grier (1976) tested a Decision Theory approach to selection of appropiate nursing care plan actions and concluded that the performance of the linear model was sufficiently sound to justify greater use on the wards. A similar mathematical approach has been used in a reported expert system in nursing project (Ozbolt et al 1985). Bennett (1980), moreover, has claimed that nurses are 'intuitive Bayesians', although they fit a normative Bayes model only when dealing with unfamiliar tasks. More recently, Grier (1981) has formalised her ideas into a complex model for decision making in nursing practice which would be quite incomprehensible for the nurse witnessing its operation - which is perhaps a crucial point. That is, the recent unpopularity of linear models relates to resistance to this prescriptive approach and to the incomprehensibility of the process of decision making. These related criticisms can now be explored.

Prescriptive decision models state how people *should* behave while descriptive models are of how people *do* behave. A descriptive model which is expert and descriptive might possess strong training potential - the present cognitive model is such a model. Albert's (1978) point is not only that these models are comprehensible but also that one is more likely to put one's faith in a decision made by a model which can be followed. A related point is that statistical linear models require computers (always readily mistrusted) to carry out the complex numerical operations. Shortliffe et al (1984) discuss this 'acceptability' issue in relation to medical computer based decision aids. The observation is made that essentially none of such systems built have been adopted outside the research environment - even when their performance has been shown to be excellent. They conclude with a call for some of the research effort to be switched to understanding and overcoming the evident bias.

Other major criticisms of linear models have been discussed (eg Elstein et al 1983) such as their inability to take into account the configural nature of data. That is, data are not always weighted independently but sometimes in the light of other predictor variables. The importance of this has been shown in studies by Edgell (1978). The example given is when a cue which is usually a symptom of pathology is in fact only a side effect of treatment. A doctor would disregard a headache as evidence in the process of making a diagnosis when it was known that the headache was a result of the drug treatment the patient was receiving.

There are, therefore, substantial reasons for utilising a linear prescriptive model in the evaluation of the cognitive model which has been constructed in the present project, just as there are substantial reasons for doubting the practical utility of such a model. The construction of this model can now be outlined.

Construction of DFA model As outlined above, the construction of the DFA formula requires a 'training set' of example cases for whom the risk decision is known. This training set, of 123 patients, was input to the SPSSx DISCRIMINANT procedure in order to build the linear formula which will be used to classify the test set of patients. The training set patients were coded according to each of the 12 attributes, for example a patient could take one of 5 values on the variable MOBILITY:
bedfast & immobile=1; bedfast with free movement=2; bed or chairfast=3; walks with assistance=4; fully ambulant=5.

The default assumption of equal prior probability of risk category membership was overridden to reflect the proportion of cases actually falling into each group (low risk=37, medium=36, and high risk=50). All variables which satisfied the tolerance criteria (.001) were force entered simultaneously. The first computation is to check for violations of the assumptions of variables coming from multivariate normal distributions and the covariance matrices being equal. The test used in DISCRIMINANT is Box's M where a very small probability leads to rejection of the null hypothesis that covariance matrices are equal. Results:- M= 121.17, F(approx)=1.31, p=.036.

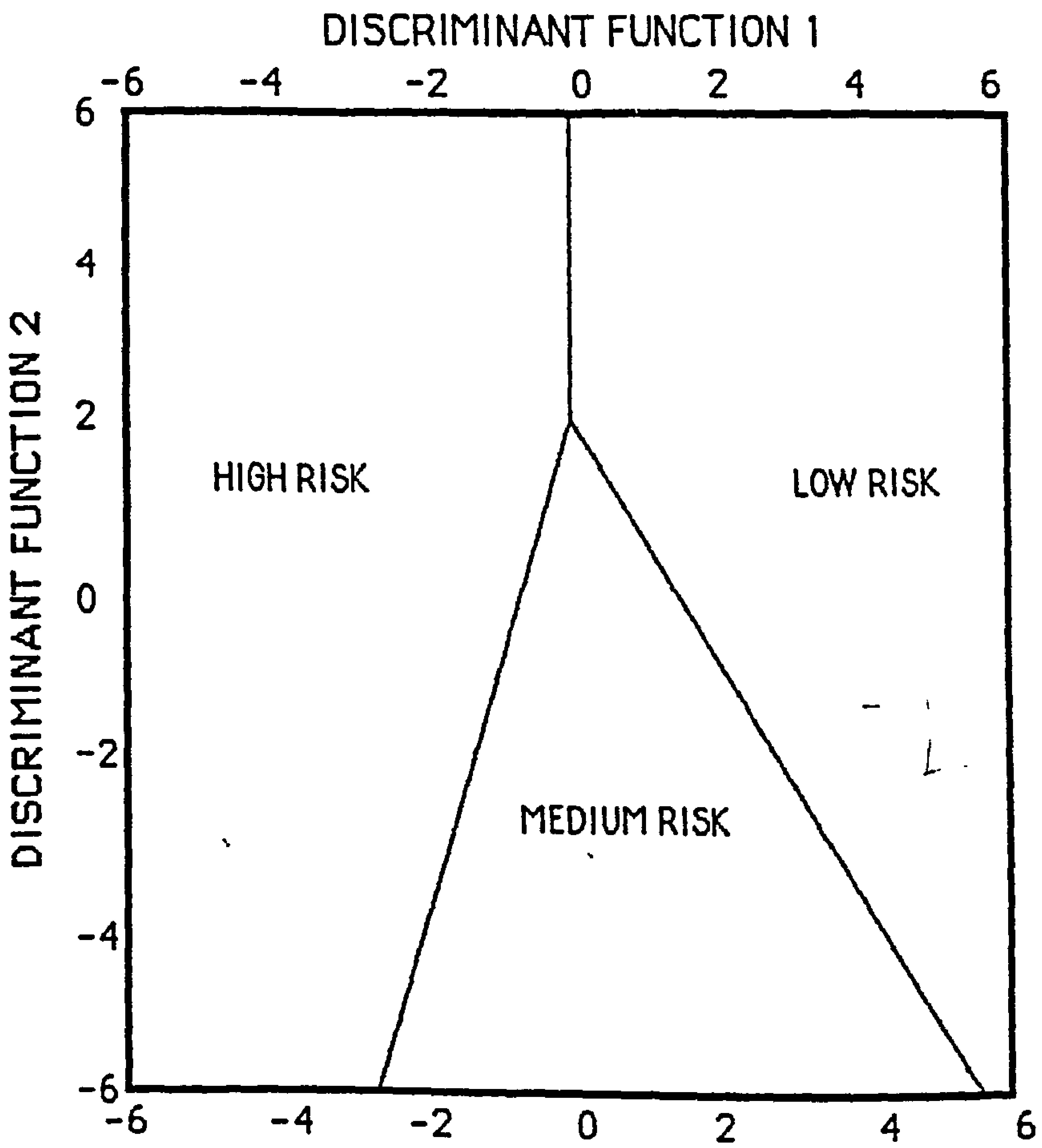
With this reassurance about the assumptions, the procedure then computes, for the 3 group situation, two sets of unstandardised discriminant function scores which act to maximise the ratio of between-groups to within-groups sums of squares. It is these coefficients, along with a constant, which are used to classify each patient. The coefficients are set out in Table 4.2 below.

Table 4.2 Unstandardised discriminant function coefficients

attribute	function 1	function 2
SEX	.40	.32
MOBILITY	.58	-.30
MENTAL ST.	.25	-.98
NUTRITION	.91	-.54
BUILD	-.30	.009
URINARY	.19	.45
LIFT & TURN	-.57	-.94
AGE	-.25	-.24
CIRCULATION	-1.08	1.18
FAECAL INCONT.	-.14	.64
BP	-.19	.61
SKIN	.15	.43
(constant)	1.33	.67

The procedure multiplies a patient's values for each variable by the coefficient for that value. In this way, the constant is added to give 2 discriminant scores for each case. Classification proceeds with reference to a 'territorial map' where the boundaries of each of the 3 groups are marked out with respect to each function. Figure 4.1 overleaf reproduces the territorial map printed by DISCRIMINANT for this data. It is possible, for example, to plot a patient with a function 1 value of -.75 and a function 2 value of -3.2. This patient would be classed as a Medium risk of developing pressure sores.

Figure 4.1 Territorial map for classification of cases by each function



As pointed out earlier, however, the comparison can only be of product of information processing rather than of process and product. For a suitable model which affords opportunity for both process and product comparison the discussion can now turn to Model 2 - the Automated Rule Induction system.

Model 2 - Automated Rule Induction (ARI)

Automated Rule Induction provides a radically different and potentially more appropriate decision model which can be compared with the cognitive model. The principal reason for this potential is that ARI, unlike DFA, has been claimed to proximate to aspects of human cognition. Hence, while the only claim that could be made for linear models is that classification accuracy is of an expert standard (see above), some claims have been made that the process of decision making of ARI is at least 'humanesque'. These claims will be reviewed below. Firstly, however, an outline will be given of the background and operation of ARI.

The origins of rule induction, as reviewed by Jones and Ritchie (1987), can be traced to the concept formation research which was popular in psychology in the 1950s and 60s (Bruner, Goodnow and Austin 1956, for example). Around that time, Hunt, Marin and Stone (1966) developed a computer program known as Concept Learning System (CLS) which was designed to simulate the process whereby a human learns concepts. Later, the basis of this technique was redesigned by Quinlan (1979 and 1983) into the more efficient and automated information theoretic algorithm called ID3. From here the method was taken up by commercial interests and developed to the point where Mowforth (1986) could report that some 2500 ARI software packages had been sold by the IT Ltd (Glasgow) company alone.

The basis of CLS/ID3/ARI is that a mathematical treatment of a set of 'example decisions' can be made to produce a parsimonious summary in the form of a decision tree. Each element of this statement requires elaboration. Example decisions corresponds to training set; therefore this element is at least familiar. The idea here is that a domain expert or experts can provide a set of cases each specified in terms of their attributes and values and each classified in terms of decision - as in databases 1 and 2 for example. The important point about these example cases is the accuracy of classification of attribute values and decision. Ritchie (1987) makes the crucial observation that the decision performance of the resulting system is limited by the quality of the initial expertise.

The decision tree output of the system, if not familiar, is nevertheless straightforward in that it comprises rules of the IF....THEN type which completely summarise the training cases. Hence, if all training cases who are High risk are 'bedfast' and no Medium or Low risk cases are 'bedfast' then one part of the decision tree output would simply denote this by stating 'IF patient is bedfast THEN risk is High'. It is clearly more usual that such rules will be more complex since some bedfast patients will be not High risk while some non-bedfast patients will be High risk; in this event the rules would involve more attribute values along the lines of 'IF....AND.....AND....THEN'. The point, however, is that the decision tree is much more parsimonious than rules corresponding to individual cases in the training set. This is achieved by maximally exploiting the 'links' between the attribute values and the decisions. The manner of this achievment is mathematical based on information theory; the final element of the elaboration.

Information is taken here to denote 'reduction of uncertainty' - the extent to which it *informs* about the liklihood of an event occuring. In the example above, 'bedfast' unambigously informed about the decision 'High risk'. ARI, put simply, embodies a mathematical way of measuring the information of the attribute values in the training set. It should be stressed that there is no semantic component in the way in which an attribute value is treated - 'bedfast' does not "mean" anything other than its power to discriminate between decision categories (see Blois 1983 for more on this point). Corso (1967) gave an early report of the mathematical formula which forms the basis of the ID3 algorithm implemented in the ARI model used below. A full explanation of the basis and operation of the algorithm is beyond the scope of this discusssion, however see Ritchie (1987) if that explanation is required.

ARI, therefore, works by measurement of information. Its first step is to measure the amount of information given by each attribute in the training set. As explained above, this refers to the power of each attribute to inform about the risk classifications. The one which is selected will ultimately become the 'header node' of the decision tree; this is, it will be always asked about first when the tree is being used to classify a new case. After the invariant header attribute, however, the route (or process of decision making) will be contingent on the attribute values selected for this new case. Since the branches of the tree are only as long as necessary to classify training set cases, the path to decision point is as short as possible. This is similar, apparently, to the cognitive model derived during this project - information search which is not invariant and which does not utilise all available information. It is therefore appropriate to offer a brief elaboration of the claims that ARI has some proximity to human cognition.

Among AI workers who are involved in 'machine learning' there has been some interest in the development of ARI. This interest has led to some perhaps injudicious claims being made for proximity to human cognition. For example, McLaren (1985) comments that the rules which are output can "often effectively model the decision making behaviour of the expert" (p.159). These claims, reviewed extensively by Jones (1987), are more partial than based on empirical evidence. Moreover, as Jones (1987) makes clear, the similiarites focus more on the overt processes of decision making than on the covert processes. More on product than on process. Hence Michie (1984) states that it is well known that experts use examples to explain complex concepts to apprentices. Since ARI learns by examples there is therefore similarity.

Nevertheless, the important point is being missed - do humans use discriminative power of information in their decision making process? There is no evidence that they do, indeed the findings of this project are that task-related concepts are more important than power of information. Even if human problem solving is based on an information theoretic approach, can this cognition be modelled mathematically? Once again there is no evidence. It is planned that the comparative exercise which is being conducted here might begin to substitute evidence for opinion in this debate.

Construction of ARI model A commercial implementation of the ID3 algorithm supplied by Intelligent Terminals Ltd (Glasgow) was used to build the decision tree which could then be used for both process and product comparison. The training set input to the software was the same as that used to build the DFA model (see above). For convenience, attributes and values were frequently abbreviated during the input of data - as explained above there is no semantic basis to these words; species of insect corresponding to each attribute or value would have produced the same result. The output was the decision tree printed by the software package which is displayed in Figure 4.2 overleaf.

The decision tree depicts 57 separate routes to a decision point, each depicted by 'high', 'medium', or 'low'. It can be seen that the header attribute is MOBILITY. To read the tree as rules (ie routes), simply put IF before MOBILITY followed by 'is' on a dotted line to a value of MOBILITY. Subsequently, put 'and' before the next attribute and so on until substituting 'then risk is' for the colon which appears to the left of 'high', 'medium', or 'low'.

It can be seen that on occasions not all of an attribute's values appear in this tree. For example, the highest value of AGE (ie 'older') does not appear in the branch where MOBILITY is 'immobile'. This reflects the operation of the algorithm and the training set. Hence, when the 'best' attribute was being located which discriminated between the risk of all the 'immobile' patients, it was found that AGE could achieve this without reference to any further attributes. There were, however, no patients in the training set who were both 'immobile' and 'older'. Potentially, this could give problems when it comes to the input of the test set patients - there could well be an immobile older patient. In fact, this occurred on only four occasions. The conservative solution adopted on these occasions was to default to the 'next most serious' attribute value.

Figure 4.2 Decision Tree output from Rule Inducer

```
SuperExpert Rule listing.
Problem: Risk

88 nodes.

Mobility
├─immobile : Age
│   ├─old : high
│   ├─middle : high
│   └─young : medium
├─bedfree : Age
│   ├─middle : medium
│   ├─older : medium
│   ├─young : Mental
│   │   ├─alert : low
│   │   └─marked : medium
│   └─old : Gender
│       ├─female : high
│       └─male : Lifting
│           ├─two : medium
│           └─one : high
├─fully : Skintype
│   ├─normal : Mental
│   │   ├─mild : medium
│   │   └─alert : low
│   ├─shiny : Gender
│   │   ├─female : high
│   │   └─male : medium
│   ├─papery : Nutrition
│   │   ├─ok : low
│   │   └─deficient : medium
│   └─drythin : Urinary
│       ├─helped : medium
│       ├─occasional : medium
│       └─fully : Faecal
│           ├─full : low
│           └─diarrhoea : medium
├─bedchair : Lifting
│   ├─one : medium
│   ├─self : high
│   ├─onetwo : Build
│   │   ├─slightunder : medium
│   │   ├─signifunder : high
│   │   ├─slightover : Gender
│   │   │   ├─female : high
│   │   │   └─male : medium
│   │   └─average : Age
│   │       ├─old : medium
│   │       ├─middle : medium
│   │       └─young : high
│   └─two : Nutrition
│       ├─deficient : high
│       └─ok : Urinary
│           ├─catheter : high
│           ├─helped : Build
│           │   ├─average : high
│           │   └─slightover : medium
│           ├─frequent : Gender
│           │   ├─female : high
│           │   └─male : medium
│           └─occasional : Gender
│               ├─female : medium
│               └─male : high
└─walks : Circulation
    ├─poor : Build
    │   ├─slightover : medium
    │   ├─obese : high
    │   ├─slightunder : high
    │   ├─signifunder : Urinary
    │   │   ├─fully : high
    │   │   └─frequent : medium
    │   └─average : Skintype
    │       ├─drythin : medium
    │       ├─papery : high
    │       └─normal : Gender
    │           ├─female : medium
    │           └─male : high
    └─ok : Build
        ├─slightover : low
        ├─obese : medium
        ├─signifunder : medium
        ├─slightunder : Faecal
        │   ├─full : medium
        │   └─diarrhoea : high
        └─average : Lifting
            ├─onetwo : medium
            ├─self : low
            └─one : Gender
                ├─female : medium
                └─male : Mental
                    ├─mild : medium
                    ├─alert : high
                    └─marked : medium
```

Results of Comparisons with Cognitive Model

As discussed above, the results of the comparisons between the three models can now be set out under the headings of Process and Product of decision making.

Process comparisons

The DFA model was omitted from analysis of process of decision making since, as explained earlier, a linear model is based on a formula which requires that all attributes are treated mathematically in any order. It has been demonstrated in this project that human information processing not only 'ignores' large numbers of the available attributes but also processess them in a highly contingent fashion. Furthermore, in Chapter 3 the section on 'attribute importance' found that a linear ordering of the attributes possessed only weak explanatory power of human cognition. It follows that statistical comparison of the DFA model to the cognitive model cannot be meaningful. The following analyses, therefore, focus only on the ARI and cognitive models. As explained earlier, it should also be borne in mind that the requirement for measures of correspondence with human Experts means that all process comparisons utilise the 18 patients from the process tracing experiment.

Process measure 1 - number of attributes selected. This measure, although coarse, is being carried forward from the analyses of human cognition in order to lend coherence to the evaluation. Hence, in Table 3.11 from Chapter 3 it was shown that the 18 patients input to the cognitive model resulted in very similar numbers of attributes being selected when compared with the Expert nurses. That table can now be augmented by inputing the same 18 patients to the ARI model and counting the number of attributes searched prior to decision. The results are given in Table 4.3 overleaf.

Table 4.3 Numbers of attributes selected by cognitive and ARI models, and mean numbers selected by Expert and Proficient nurses

patient	Expert nurses	Proficient nurses	Cognitive Model	ARI Model
1	3.8	6.3	5	4
2	3.6	6.1	5	2
3	6	6.7	4	4
4	4.6	8.3	4	2
5	4.4	6.7	5	2
6	2.2	5.7	5	2
7	4	6.4	4	2
8	4.6	6.7	5	3
9	6.8	7.8	4	3
10	6	7	5	3
11	3.2	5.4	3	2
12	5.6	7.8	4	4
13	3	4.6	4	3
14	3.8	5.4	5	3
15	3.8	6	4	3
16	3.6	4.1	4	3
17	4.4	6.7	4	4
18	3.8	5.7	4	3
mean	4.29	6.30	4.33	2.89
st. deviation	1.17	1.08	.59	.74

The earlier analysis in Chapter 3 revealed, using paired-sample t tests, that only the number of attributes selected by the Proficient nurses was significantly different from the numbers selected by the cognitive model. The numbers selected by the ARI model, however, are significantly less beyond the $p<.001$ level than the numbers selected by Proficients, Experts, and cognitive model. (ARI vs Experts: $t=4.26$, ARI vs Proficients: $t=10.96$, ARI vs cog model: $t=6.36$).

The ARI model was introduced earlier as being maximally parsimonious in terms of its procedure for discriminating between the risk decisions. On this evidence, the parsimony significantly exceeds that of both human experts and the cognitive model constructed to emulate those experts. More important, however, will be the other measures of process agreement which begin to compare which attributes are selected rather than how many.

Process measure 2 - point-by-point agreement This measure can also be carried forward from Chapter 3. In Table 3.12 of that chapter, a final test of the power of the cognitive model to explain the information processing of each Expert was undertaken by constructing a matrix of point-by-point agreement indices. The cognitive model was seen to approximate most closely to the Expert subjects as individuals. It becomes sensible, now that an alternative model has been made available, to repeat this exercise with the inclusion of a matrix derived from the ARI model when assessing the same 18 patients. Results are given in Table 4.4

Table 4.4 Matrix of point-by-point agreement between each Expert, cognitive model, and ARI model

	E2	E3	E4	E5	model	ARI
E1	119	95	110	110	107	119
E2		128	137	124	142	137
E3			124	94	132	123
E4				107	143	142
E5					102	125
model						116

A similar procedure can be followed as in Chapter 3 in order to identify the individual or model which most powerfully explains each individual or model. The matrix now comprises 21 pairings. These pairings can be ranked in order of magnitude of the index:

	pairing		pairing
lowest index -	1 E3 with E5		12= E3 with E4
	2 E1 with E3		12= E2 with E5
	3 E5 with model		14 E5 with ARI
	4=E5 with E4		15 E2 with E3
	4= E1 with model		16 E3 with model
	6= E1 with E4		17= E2 with E4
	6= E1 with E5		17= E2 with ARI
	8 ARI with model		19= E4 with ARI
	9= E1 with ARI		19= E2 with model
	9= E1 with E2	highest index-	21 E4 with model
	11 E3 with ARI		

As before, it becomes straightforward to identify the subject (or model) which offers the closest approximation to the group.

Inspection of these ranks reveals that the 'best match' to each subject is as follows:

E1 – best match = E2 or ARI
E2 – " " = model
E3 – " " = model
E4 – " " = model
E5 – " " = ARI

The models, therefore, seem each to have some claim to closest approximation to the subjects. If the cognitive model is taken as 'winner' then the margin is unconvincing due to the nature of this data being at the descriptive level only. For two reasons, then, the requirement is for a more sensitive test which can incorporate reliability measures. Firstly, the high agreement observed relates only to exact match of data points – there is a sense in which a slight difference in order of attribute selection may be less important than the <u>relative</u> orders of, for example, a subject and a model. Secondly, there is no basis for a <u>reliable</u> statement to be made on the power of each model to correspond to each Expert's process trace. This more sensitive test – of <u>order</u> in which information is selected – can now be undertaken.

<u>Process measure 3 – order of information selection</u> This measure, as introduced earlier, is the correlational index suggested by Ray Meddis of Lancaster University. The process traces of both cognitive and ARI model were recorded as each of the 18 patients were input. Spearman's Rho values between these traces and the traces of each Expert were then computed. Results are given in Table 4.5 overleaf.

Table 4.5 Correlational indices between process traces of each Expert and the cognitive and ARI models

	ARI model					cognitive model				
patient	E1	E2	E3	E4	E5	E1	E2	E3	E4	E5
1	.10	.24	.35	.24	.15	.64	.77	.70	.60	.95
2	.52	.49	.79	.79	.67	.72	.58	.77	.77	.18
3	.38	.15	.17	.43	.59	.37	.73	.83	.72	.37
4	.56	.62	.90	.61	.59	.04	.73	.76	.85	.35
5	.30	.69	.69	.61	.68	.59	.68	.73	.70	.25
6	.45	.69	.90	.90	.99	.48	.80	.71	.71	.48
7	.52	.54	.61	.54	.65	.39	.79	.98	.52	.65
8	.35	.15	.63	.35	.35	.70	.69	.68	.92	.74
9	.74	.46	.45	.54	.21	.30	.88	.83	.70	.25
10	.45	.56	.34	.29	.14	.55	.53	.67	.73	.18
11	.65	.79	.90	.73	.65	.62	.92	.83	.38	.73
12	.30	.56	.07	.51	.63	.04	.65	.86	.84	.06
13	.64	.88	.78	.68	.75	.81	.86	.90	.56	.70
14	.48	.88	.68	.99	.33	.27	.81	.83	.88	.38
15	.23	.77	.57	.66	.52	.37	.93	.91	.81	.42
16	.57	.77	.70	.99	.69	.31	.93	.99	.74	.64
17	.77	.47	.50	.65	.14	.52	.70	.91	.56	.46
18	.72	.69	.65	.86	.52	.69	.61	.81	.81	.42

These correlations can be considered as 'indices of proximity' between the Experts information search behaviour and that of each model. As such, the indices can be treated in an ANOVA with 2 within group factors. Factor 1 (Model) has 2 levels - cognitve model and ARI model. Factor 2 has 5 levels, each corresponding to Experts 1 to 5. The results of this ANOVA are set out in Table 4.6 below.

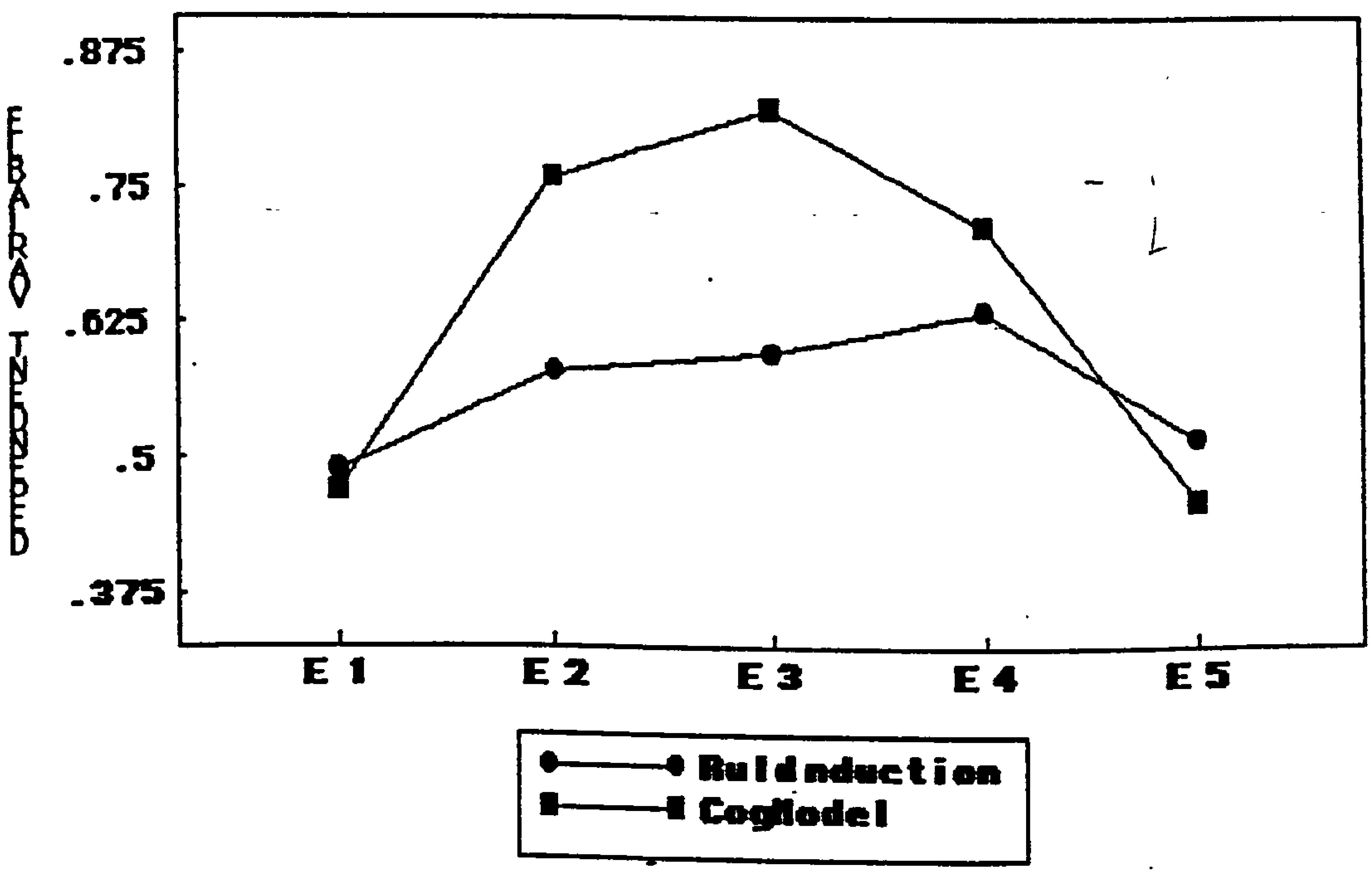
Table 4.6 ANOVA of Proximity indices between Experts and cognitive or ARI models

Source of Variation	df	SS	MS	F	p
Subjects	17	1.64	.10		
Model 1	.29	.29	4.04	.06	
error	17	1.24	.07		
Experts	4	1.77	.44	14.99	<.0001
error	68	2.00	.03		
Model X Experts	4	.53	.13	4.64	<.01
error	68	1.93	.03		

Although the mean index for the cognitive model exceeded that of the ARI (.64 versus .56), as Table 4.6 shows the effect for Model was not found to be significant. There were, however, significant differences suggested within the interaction between Model and Experts. Figure 4.3 overleaf, which plots the indices for each model's proximity with each Expert, suggests that there may be significant simple effects within the main effect for model since for some of Experts there appears to be fairly large discrepancies between the proximity of each model.

Testing of these simple effects revealed two significantly different indices, each in the cognitive model's favour. (Effect of Model at E2: $F=12.7$, df1 and 17, $p<.01$. Effect of Model at E3: $F=13.04$, df1 and 17, $p<.01$) The more specific conclusion from this analysis, therefore, is that on the two occasions when there were significantly higher correlations with Experts then on each of these occasions it was the cognitive model which showed greater proximity.

Figure 4.3 Mean proximity index for each Expert(E) by cognitive and ARI models



Some conclusions can be made with respect to the three comparisons of process proximity.

- using number of attributes selected, the cognitive model approximates significantly more closely to Experts than does the ARI model.
- using the descriptive measure of point-by-point agreement, both cognitive model and ARI were shown to provide better 'averages' of the Experts than did any one individual Expert. A further tentative conclusion was that the cognitive model succeeded in this better than the ARI model.
- using the proximity measure, reliable improvements in mean indices were found in the cognitive model's favour for two out of the five Experts.

In general, therefore, the findings are in a consistent direction - the cognitive model emulates human cognition more closely than does the ARI model. Having drawn this conclusion from empirical evidence, the luxury of more qualitative comment can now be briefly indulged. The first point to be made is that the header attribute – MOBILITY – agrees with cognitive model and Experts. This seems reasonable; when humans have no information then the attribute which most powerfully affects whether a patient succumbs to a pressure sore will be the priority for search. However, the ARI model, from that point onward, seems to depart from reason on occasions in its relentless use of discriminative power of attributes. The most obvious example is the use of GENDER in the decision tree, nearly always at the 'terminal' point so that that a very few cases can be finally discriminated. This situation would be analogous to a human saying ..."I've found enough to decide that this patient is either X or Y decision; now I only need to GENDER in order to decide between X or Y".

This seems superficially reasonable except for the crucial point that there is no <u>conceptual</u> reason for males being higher (or lower) risk than females for developing pressure sores. It is simply fortuitous that GENDER is a two-value attribute which potentially carries high information power to discriminate cases. The mathematical model, at least at the terminal nodes, seems therefore to depart from human cognition. Also weakened is Michie's (1984) claim that ARI permits the expert to "transfer to the machine a judgemental rule which he already had in his head but had not explicitly formulated" (p338). On the basis of the evidence above and on these comments, therefore, there is insufficient evidence to reject the null hypothesis that ARI is a *process* model of human cognition. The task now becomes one of evaluation of the product of the decision making.

<u>Product comparisons</u>

For product comparisons – the outcome of decision making – the measure of comparison to be used is decision accuracy with the test set. As explained earlier, this test set comprises the 152 patients in database1 with the exception of the 18 of these patients who were used in the process tracing experiment. According to the rationale explained in Part 4 of Chapter 3, the outcome measures used will principally include Kappa coefficients both overall and by risk category. In addition, percentage agreement figures will be given.

<u>Procedure</u>. Unlike the process comparison, all three models become eligible for product comparison.

<u>For the cognitive model</u>: using the computer implementation of the cognitive model, each of the 134 patients was 'run' through, ie each time a value was requested for an attribute the

appropriate value for that patient was input until a risk classification was given.

For DFA model; an ad-hoc computer program was written within SPSSx in order that 2 discriminant scores (one for each function) were computed for each of the 134 test patients. Each patient was then classified into a risk category by plotting these scores on the territorial map reproduced in Figure 1 above.

For ARI model: each of the 134 patients were input through the decision tree depicted in Figure 4.2 (page 213) until a risk classification was reached.

Results Comparison of the test set decisions with each model's classifications was then undertaken. The results of these exercises for the cognitive model, the DFA model, and the ARI model are given in Tables 4.7 to 4.9.

Table 4.7 Decision concordance - cognitive model with test set (n=134)

	number of judgements		model agree with test set			
Risk	test set	Model	n	%	Kappa	p
LOW	50	36	34	68.0%	.69	<.01
MEDIUM	39	38	18	46.1	.25	ns
HIGH	45	60	41	91.1	.64	<.01
overall	134	134	93	**69.4%**	**.54**	<.0001

Table 4.8 Decision concordance - Discriminant Function Rule(DFA) with test set (n=134)

	number of judgements		DFA agree with test set			
Risk	test set	DFA	n	%	Kappa	p
LOW	50	40	34	68.0%	.65	<.01
MEDIUM	39	58	27	69.9	.20	ns
HIGH	45	36	26	57.8	.49	<.05
overall	134	134	87	**64.9%**	**.47**	<.0001

Table 4.9 Decision concordance - Rule Induction(ARI) with test set (n=134)

	number of judgements		ARI agree with test set			
Risk	test set	ARI	n	%	Kappa	p
LOW	50	39	34	68.0%	.65	<.01
MEDIUM	39	56	26	66.6	.30	ns
HIGH	45	39	26	57.7	.44	<.05
overall	134	134	94	**70.1%**	**.46**	<.0001

Interpretation of the decision concordance results in Tables 4.7 to 4.9 can be accomplished with the aid of the summary table below (Table 4.10) where the cognitive model is denoted by CM.

Table 4.10 Summary results of decision concordance calculations

	n pts	n agreements with test set			Kappa values with test set		
Risk	test set	CM	DFA	ARI	CM	DFA	ARI
LOW	50	34	34	34	.69	.65	.65
MEDIUM	39	18	27	26	.25	.20	.30
HIGH	45	41	26	26	.64	.49	.44
overall	134	93	87	94			
Kappa		.54	.47	.46			
%		69%	65%	70%			

Table 4.10 shows:

a) using the simplest measure of number of agreements per risk category, the models are equal for Low risk patients only. The cognitive model is seen to perform well for High risk patients in achieving 41 out of 45 agreements. This, however, must be interpreted in the light of apparent 'overprediction' since there seems to have been a general trend for the cognitive model to increase risk (50 test set patients in Low, CM put 36 in Low; 45 test set patients in High, CM put 60).

b) for slightly greater clarity, the total number of agreements can be inspected. This shows that ARI and CM achieve similarly high levels of 94 (70%) and 93 (69%) respectively while DFA achieves somewhat less at 87 (65%). Overall, this level of 'correct' judgement is impressive if chance probability is taken into account; only around 44 cases would be placed in each category on a random basis.

c) using the only measure for which there are reliability measures (Kappa), it can be seen that the cognitive model achieves the highest values for each risk category. Each model showed non-significant concordance with the test set's Medium risk patients. This is the group predicted as most variable since judgements can most easily 'go either way'. All other values were highly significant ($p<.01$) with the exception of DFA and ARI with respect to High risk patients ($p<.05$ on both occasions).

d) finally, using the most important statistic - overall Kappa values - it appears that the cognitive model performance exceeds that of the DFA and ARI (.54 versus .47 and .46). Regrettably, however, there is no method whereby these values can ever be deemed 'significantly different' from eachother.

Conclusion

In conclusion, it appears that in terms of decision accuracy with a test set the cognitive model approximates most closely with the nurses who originally cared for these patients and evaluated their risk of developing pressure sores. The ARI model also performs well; however there are grounds for additional satisfaction with the cognitive model given the manner in which the training set for ARI and DFA was constructed. In that exercise (described above) the cognitive model was used to screen out 'potential error' patients from database2.

This result should be added to the favourable result for the cognitive model with respect to the earlier process analysis. The cognitive model seems to not only process information with greater similarity to humans but also seems to concur more closely with decisions made by humans. In terms of an evaluation of statistical versus cognitive appoaches, the findings reported above can be seen to concur with those of Fox (1980). In that study, the comparison undertaken was in several respects similar - human information seeking and decision processes evaluated against an ARI-like model and against a simple cognitive model. It was found that the traditional prescriptive account of decison processes gave a moderately good account of observed human behaviour but that the cognitive model showed equal or better quantitative approximation.

The ultimate evaluation, however, will concern the utility of the cognitive model as a teaching tool. Nevertheless, the earlier literature review on evaluation pointed out at length the considerable problems associated with such an evaluation. An experiment might be designed in which learner nurses witness the system assessing their own patients and periodically were tested with one of the stored patients in the program. The extent to which learners themselves use the information processing principles which they have observed when assessing patients presented to them by the system will form the basis for measurement of degree of 'expertness' of performance - the hypothesis being that exposure to the Intelligent CAL will bring about 'drift' of their performance toward that of an 'emulated' expert.

Such plans are tantalising but premature with regard to the present thesis. As discussed in the earlier literature review, it would be more appropriate to follow the sober course outlined there whereby the evaluation can now be completed by turning now to testable predictions made by the cognitive model.

REPRESENTATION OF PATIENT INFORMATION

Introduction

During the analysis of the information processing behaviour of subjects which was reported in Chapter 3, several references were made to the suggestion that a 'care orientation' apparent in the performance of Expert nurses. Thus in the Part 1 it was found that a list of attributes ranked in order of importance for planning the care of patients had stronger explanatory power than other lists of ranked attributes. Similiarly, the discussion on the findings of the analysis of higher cognition (Part 2) suggested that knowledge representation in long term memory might be organised on a functional care-planning basis which reflects the goal-directed nature of the observed cognition.

Some of these suggestions can be illustrated by the following excerpt from the transcript of the verbalisations of Expert nurse 2. For convenience, the transcript has been broken down into 12 lines (L1 to L12); 'click' refers to the mouse being operated (ie a button on the screen being pressed in order that an attribute might be made to reveal its value).

L1 ... mobility ...
L2 click ... bedfast but can move freely in bed, so that's a bad sign ...
L3 click ... and they are mildly disorientated, I what ...
L4 click ... age, it's 70 - 89. Risk seems high so far
L5 click ... build is significantly underweight, and it takes ..
L6 click ... 2 nurses to turn him, patient unable to assist
L7 click ...
L8 click ... but circulation is good and nutrition seems adequate. Skin might be ok ...
L9 click ... no, it's fine and delicate ... so I better see about urine ..
L10 click ... oh, a urodome fitted and faecal ...
L11 click ... is occasional faecal incontinence.
L12 well, bedfast but can move freely again immediately highlights a possible risk, especially since they are mildly disorientated. Build was underweight and poor skin ... they had a urodome fitted but if this failed with occasional faecal incontinence then the skin would be wet. He is 70 - 89.

This transcript can be divided into two parts. From Line 0 to L11 there are information search verbalisations which, though fairly unilluminating, do contain the suggestion that risk is being assessed on an on-going basis. L12 contains a post-hoc review where this nurse recalls/reads out the attribute+values which she adjudged to represent problems.

Two interesting points can be made about this transcript. Firstly, it is notable that only problems are recalled, and secondly, there is no clear correspondence to the order in which the problems were recalled when compared with the order in which these problems were uncovered. The suggestion is that patient assessment is driven by the need to plan care. It therefore becomes interesting to speculate on the apparent relationship between the order in which problems are reviewed and the appropriate care for this patient. Thus the first four problems reviewed - bedfast, mildly disorientated, underweight, and poor skin - map directly to the planned relief of the patient's positional pressure. Problems reviewed in positions 4 to 6 (poor skin, urodome, faecal incontinence) in turn map directly on to the second aspect of appropriate care - moisture prevention.

This and earlier (Chapter 3) findings suggest that expert nurses organise the information they elicit and that there is a functional basis to this organisation. While the organisation of information into chunks has been established since the seminal work of Miller (1956), the interest in the effect of expertise on chunking has grown since de Groot's (1965) observation that there was a functional basis to chess masters' recall extensive configurations of board pieces. This basis reflected functions such as attack or defensive potential. Chase and Simon (1973) confirmed the finding and extended the analysis to a demonstration that novice chess players were more likely to represent configurations on the basis of surface features rather than hierarchically organised on a conceptual basis.

The strong evidence for functional effects in memory organisation has more recently led to more specific interest in the implications for skill of differing representations. The representation issue, furthermore, is becoming increasingly related to manner in which information is processed by novice and expert performers. In fact, it would possibly be appropriate to use the term 'cognitive architecture' (Anderson 1983) to denote a unified cognitive explanation of mental representations, memory structures, and processing mechanisms. This is of central interest to the present project, hence the goal of this Part of Chapter 4 is to carry out more formal empirical investigation into the basis to nurses' representations. Before reporting the experiments, however, it is necessary to introduce and review the literature relating to the nature of memory organisation, particularly with respect to expert practitioners.

Literature Review – Knowledge Representation and the Expert Practitioner

Up until this point the present project may have seemed to be solely concerned with the *process* of patient assessment. It is, however, impossible to separate information processing from what Rumlehart and Norman (1983) deem to be the central issue in cognitive psychology – the nature of knowledge representation. By this term there are in fact two inseparable issues denoted. Firstly, the mental representation which the nurse constructs of the patient she is assessing, and second, the memory structures which she employs – although for a complete account of cognition it is necessary to include processing mechanisms. Although this section introduces two experiments which fairly specifically look at the first aspect above – the mental representation of the patient being assessed – there are in fact evaluation issues which extend both to the nature of how knowledge is stored and to processing mechanisms. These issues implicated by study of representation of the problem or patient are central both to cognitive psychology and to nursing. Issues such as whether cognition is conscious or automatic, limited or unlimited in capacity, and the extent to which incoming information or stored knowledge drives the assessment. Comment, where appropriate, can be made on these issues throughout this outline of the literature.

It is instructive to begin a look at the literature by drawing lessons from findings from other areas of psychology. In visual perception, for example, the dominant model for recognition of objects formerly emphasised the incoming information. It was argued that recognition depended on stored prototypes or on lists of defining features (eg Gibson 1969) which perceivers held of such objects. More recent evidence, however, has demonstrated the role of the semantic nature of stored knowledge so that what we perceive is to an extent determined by prior expectations generated by a conceptual representation of a scene (eg Goodman 1980). Representation depends, therefore, on an interaction between bottom-up <u>and</u> top-down processing.

The prototype approach to patient classification by nurses has been suggested by Abraham (1988) and by Tanner and colleagues (eg Westfall, Tanner, Putzier and Padrick 1986). This model is similar to that put forward in medicine. For example, Rubin (1977) suggests that physicians have stored 'disease templates' of defining signs and symptoms which are activated early in the diagnostic process and attempted to be fit to incoming patient data. Clusters of attributes which are highly correlated in the real world are represented as typical of the category. However, this model contains no explicit suggestion of a deeper 'conceptual' classification of patient details in this classic hypothetico-deductive model. Unlike the case of object perception, there seems to have been no shift toward a recognition of the importance of the functional nature of the representation. Classification, rather than action following from classification, is taken to be the goal of patient assessment.

There are similarly useful lessons to be drawn from looking at another area of psychology – the person perception domain. Here once again the dominant model is one which emphasises a prototypical representation. However, there has been greater attention to a theoretical model of the *purpose* of representation. Hence in the work of Cantor and Mischel (1977) it is assumed that we categorise people into stored stereotypes such as 'extrovert'. The finding has been that recall of a person's attributes are a function of the centrality of these features to a prototype.

The theoretical suggestion is that this mechanism fits with the principle of cognitive economy in that rapid categorisation can be achieved which permits the perceiver to utilise the 'person schema' to infer beyond the information on the basis of expectations contained in the stored knowledge. Schneider and Blankmeyer (1983) support this prediction of facilitation of quicker and more efficient processing by showing that inferences about a person are a function of the salience of prototypes – if some features of a person are disconforming of the prototype then it is harder to make inferences about that person.

Research in the area of concept categorisation within psychology can be seen to have followed a similar progression to that of object perception and to have developed the theoretical basis to representation. Hence the earlier debate on the nature of stored representations was between prototype versus lists of features (see Mervis and Rosch 1981). A chair would be represented as a 'classic kitchen type' or as a list of defining features such as 'has legs; seat; ...'. More recently there has been a recognition that different models of representation may be necessary depending on the type of concept, individual knowledge, and the purposes of the representation. How well-defined the concept is, how expert the perceiver, and what goal underlies the perception all become important. The more flexible position which has been taken in order to account for these points is the *dual representation* model. Armstrong (1983) provides some evidence to support such a model – a fuzzy representation constructed from the properties of typical members and, if necessary, a more precise definition known as a conceptual core.

The key distinction, then, is between a 'shallow' representation focusing on surface features and deeper level knowledge based on a conceptual categorisation. The idea of a deeper representation of conceptual knowledge is familiar in the literature on Nursing Models. Roper, Logan and Tierney (1985), for example, maintain that the conceptual basis to nursing assessment relates to the patient's activities of daily living (ALs). On this basis the prediction would be that mental representation and categorisation of patient information would reflect the various ALs.

Nevertheless, the conceptual representation model puts stronger emphasis on the function of perception or assessment. It might therefore be helpful to shift the emphasis to the care planning purpose which underlies the purpose of assessment.

A functional basis to representation is supported by some recent nursing literature. For example, Stainton (1988) complains that clinical judgement is not only the formulation of a diagnosis. Furthermore, she suggests that "the meaning (of patient cues) for the nurse will be found in the way that it then directs CARING .."(p.275). Although this position is not contradicted by the conceptual model of ALs outlined above, it nevertheless stands in contrast to the dominant North American model of nursing diagnostics (eg Kim, McFarland and McLane 1984). This approach, owing much to the medical hypothetico-deductive model of diagnosis, can be seen as fitting more with a shallow level representation of the patient where the goal of assessment is to fit incoming information to a predetermined set of necessary and sufficient criteria.

The suggestion, then, is that nurses will potentially represent patient information on shallow and/or deep levels. The strongest support for this position will be cited presently when the experimental literature on expert and novice differences in problem solving is reviewed. However it will be instructive to look in some more detail at the nursing and medical literature prior to that review. Firstly, the significant comment can be made that the 'patient prototypes' which comprise the set of nursing diagnoses have been derived from studies asking nurses to list patient properties rather than studies of how nurses actually think. The major support for the model is drawn from the medical literature where, for example, the pioneering work of Elstein and colleagues (eg Elstein, Shulman and Sprafka 1978) showed that physicians appear to hypothesise diagnoses which are then used to guide further search for confirmatory data.

Similarities between nursing and medical cognition are seen by Carnevali (1983) to be many, while differences seem largely to imply that nurses possess much larger capacity working memories in that her prescription is that many more aspects of functional and dysfunctional patient information is to be simultaneously represented by nurses. There are three main objections to this alleged similarity. Firstly, the challenge to the weak evidence for the hypothesis testing model in medicine (see Patel and Groen 1986) which includes failure to demonstrate expert/novice differences in the use of hypothesis testing and the evidence that expert knowledge representation is much more elaborated than simple prototypes of diseases (Feltovitch, Johnson, Moller and Swanson 1984). Secondly, the general absence of evidence from nursing studies alongside the findings of Corcoran (1986) and Benner (1984) which suggest that processing styles of expert nurses are not only different to novices but also highly situation specific. Thirdly, the objections based on the 'curing-caring' distinction (eg Altschul 1978). The moment has arrived, therefore, when Stainton's (1988) call for experimental investigation of expert and novice cognition should be heeded.

Finally, then, this review can conclude by returning to some of the most apposite literature - knowledge representation and cognitive processing differences by experts and novices. It was mentioned earlier that novice chess players were found to be more likely to represent board configurations on the basis of surface features rather than hierarchically organised on a conceptual basis. More recent work has comfirmed this finding in applied knowledge domains. Hence in a comprehensive series of studies of expert and novice physicists Chi, Feltovich and Glaser (1981) showed that novices tended to categorise a physics problem presented to them in terms of the the literal properties while experts represented problems according to the type of physics principles which fitted the problem. More critically, the representation by experts was found not only to be conform to 'deeper' concepts but also to include procedural knowledge necessary to solve the problem.

This idea that experts group information into an internal 'model' of the problem is consistent with the foregoing findings concerning representation. Hence this model will reflect both an initial categorisation process of incoming cues and a completion of the representation based on stored knowledge of problem type which contains procedures for solving the problem. Similar findings have been reported by Mckeithen and Reitman (1981) with respect to computer programmers where skilled programmers were shown to organise lines of computer code presented to them in terms of common functions. Novices, on the other hand, tended to construct a more shallow internal representation which grouped code more by superficial similarities.

Some methodological points can finally be made before reporting the experiments to investigate the nature of representation by novices and expert nurses. Chiefly, methods used have utilised free recall tasks, the rationale being that grouping of items in memory will be preserved in the recall protocols. Items within one chunk will be recalled before moving on to the next chunk. Various statistical techniques can then be applied to the recall data in order to abstract the details of memory organisation. Mckeithen and Reitman (1981), for example, employed Multidimensional Scaling, but concluded that this technique was more suited to answering questions about the relationships between items of represented knowledge.

The technique which has been most commonly used to analyse the organisation of cluster membership is Hierarchical Cluster Analysis (Adelson 1981). However, although the present investigation will employ this technique, there is a basis for criticism of the commonly-used method in that subjects have typically expected a recall test. As Cohen (1986) points out, such laboratory-based experiments have typically been concerned with the mechanisms rather than the content of memory. Since this investigation aims to make a statement concerning each of these aspects, it follows that a methodolgy which includes an unexpected memory task might avoid artifactual effects on the measured content of memory.

A suitable intervening task between problem presentation and recall task is suggested by the 'person perception' literature reviewed earlier. This task would be a variant on the method which is associated with semantic network models of memory. Hence Harris and Hampson (1980), assuming that the strength of relationship (or 'distance') between any two items in memory can be quantified, set up a reaction time experiment to estimate grouping of items.

However, while the 'distance' principle is apposite to the present hypothesis, the reaction time paradigm cannot easily be incorporated. Nevertheless, there has recently been a resurgence of interest in metamemory (knowing what you know) methods (eg Lachman, Lachman and Thronesberry 1981). It follows that a suitable intervening task between problem presentation and recall would be an exercise where subjects were asked to 'rate their own database' in terms of the extent to which items are grouped.

Two experiments, therefore, have been introduced with this review. It is intended that the final discussion will of the combined findings from the reports of each experiment. The basic design of the experiments is to give tasks to subjects which follow from a presentation of patient in case history form. The entire exercise, including instructions, was contained within a booklet (see Appendix 6). The separate experiments undertaken by the subjects after they had read the patient description were:

1. self-rating exercise on how closely pairs of items from the case history were 'grouped' within memory, and
2. recall task when subjects were asked to write down all items from the case history which they could remember.

Experiment 1 - Self Rating of Item Grouping

The hypothesis is that more experienced nurses will employ a 'functional' basis for grouping (or chunking) items. That is, the greater the relevance a pair of items has for patient care then the higher the degree of relatedness these two items wil have. Alternate hypotheses will be considered in that three other bases for grouping items will be tested. These alternate bases involve both superficial and deeper level representation.

Method

The basic design of the experiment involved firstly exposing subjects to a brief description of a patient at-risk of developing pressure sores and secondly asking subjects to estimate through introspection how closely 'grouped' were the facts in this description within their own mental representations of the patient. The case history description read by each subject was:

<u>A patient at risk of pressure sore development</u>

> Mrs Ritchie (73) has a protein deficient nutritional state. She is mildly disorientated and bedfast and immobile in bed. Her circulation is good, blood pressure is normal, and she has good control of her bowels. However she is occasionally incontinent of urine.
>
> The nurses caring for Mrs Ritchie, who is lying on a sheepskin to relieve pressure, are ensuring that roughly every two hours her position is changed and she is encouraged to use a bedpan.

To operate the self rating exercise, all subjects were presented with all possible pairings of the 8 main facts in the description - 28 combinations in total. Possible order effects were controlled for through counterbalancing of blocks of pairs in a Latin Square. Four blocks were established through random allocation of the 28 pairs. The counterbalancing operated both between and within 3 groups which were tested. Hence, for example, Block A (pairs 1-7) appeared as either the first, second, third, or last block of pairs for equal numbers of subjects within each group. The 28 pairs, with further information to be explained below, are set out in Table 4.11 on page 235.

Analysis was planned largely to be between groups of subjects, although comparison of ratings given to different pairs of items within each subject would be incorporated into the analysis. Of interest in the analysis was investigation of the basis for variability within and between subjects ratings of different pairs. Four factors were hypothesised prior to the experiment as possessing the potential for explaining variability:

1. <u>Co-occurrence</u> Any given pair of facts may co-occur in reality with varying degrees of regularity of infrequency. Reliable co-occurrence might lead to economical grouped storage of these facts. Very rare co-occurrence, paradoxically, may have the same effect, although the

construction of the description was careful to avoid this. This test for the presence of this factor corresponds to prototypical representation.

2. Textual Distance How close or distant a given pair of facts appeared in the recently-read description may affect strength of grouping. This factor corresponds to a representation based on superficial features.

3. Risk Salience The extent to which a pair of facts represented a problem for the patient in terms of risk of developing pressure sores might be expected to provide an explanation of strength of grouping. Risk as a factor corresponds to a deeper level representation which is not as functional (in terms of care planning) as Factor 4 below. A further aspect of the Risk factor, however, is the degree of salience of facts (how noticeable).

4. Care applicability If a pair of facts held strong implications for the nursing care of that patient then the prediction would be that this pair would be more strongly grouped than a pair of facts which do not interact to represent care implications. This factor corresponds most directly to deep level functional representation.

Subjects 3 groups of subjects with equal numbers were recruited to the experiment. Each group represented a different level of years of experience:

- the Naive group had no experience in caring for this type of patient,
- the Beginner group were 2nd year Degree nursing students, and
- the Experienced group had each worked (and were currently working) with this type of patient for a minimum of 2 years since basic qualification.

A subsequent section of the questionnaire asked subjects to answer questions about length of experience in case, for example, an apparently naive subject had in reality cared for such a patient at home. No attempt was made to control for expertise rather than experience, for age or for sex differences. The numbers (n=17) in each group were determined by the size of the Beginner group since they were tested first.

Materials A booklet (appendix 6) was compiled which contained all instructions necessary, the patient description, and the response sections for both Experiment 1 and Experiment 2. Booklets had identical content and varied only with respect to the blocks of facts appearing in different positions for the self rating exercise. Experiment 1, completed first by all subjects, asked for indications on 8.5 cm. analogue lines of the degree to which each pair of facts was grouped. Only the extremes of the analogue lines were anchored - with 'not grouped at all' and 'strongly grouped'.

Procedure The Beginner group was tested in a class situation, they had agreed to meet and give 10 minutes to 'a nurse researcher looking at the process of nursing'. Subjects in the other two groups were recruited individually - Naive subjects in University study areas; Experienced subjects while at work in wards of a Glasgow general hospital. All subjects were asked to work at their own pace in order to avoid artificial constraints on cognition.

Results

Prior to analysis it was necessary to define each of the 28 pairs in terms of the four predicted explanations for variability outlined above. An index was required which would indicate the extent to which a given pair was 'loaded' with respect to each of the four factors. This was achieved largely by referring to a separate database of n=154 General Hospital patients described and evaluated in terms of pressure sore risk and care plan. The procedure for each factor was:

Factor 1 - Co-occurrence (O) The number of patients who satisfied both facts in a pair was calculated as a proportion of all patients. For example, of 154 patients there were 83 who had both a good circulation and good control of bowels (pair 3) - a proportion of .52. Index of Co-occurrence for pair 3 therefore = .52.

Factor 2 - Text Distance (T) A simple index of 'distance between facts' in the patient description was identified. If any two facts were neighbours in the description then the index was 1 - the facts comprising pair 3 had an index of 2 since there was one intervening fact.

Factor 3 - Risk (R) As for Factor 1 except that ther number of patients who satisfied both facts in a pair was expressed as a proportion of the number of patients who were designated by their nurses to be a high risk of developing pressure sores. For example, there were only 8 patients satisfying pair 3 facts out of the 83 in the whole sample who were high risk. Index of Risk for pair 3 therefore = 8/83 = .09.

Factor 4 - Care (C) As for Factor 3 except that the proportion was calculated of patients who required the most intensive care category.

Indices for each factor were calculated for each of the 28 pairs (see Table 4.11 overleaf).

Testing of the predictions required that a multivariate analysis was performed which considered ratings given to pairs which were either strong or weak on each of the 4 factors. In order to identify candidate pairs which satisfied these criteria, the median index for each factor was calculated. If a pair was 'strong' on a given factor then the index for that pair would be above the median for that factor, similarly a 'weak' loading on a factor would be indicated by a below-median index.

Table 4.11 28 pairs with indices calculated for 4 factors

	INDEX OF			
pair	co-occur	text dist	risk	care
1. aged 73 / mildly disorientated	.17	2	.42	.65
2. bedfast and immobile in bed / normal blood pressure	.12	2	.69	.64
3. good circulation / good control of bowels	.52	2	.09	.13
4. normal blood pressure / occasionally incontinent	.08	2	.31	.39
5. bedfast and immobile in bed / aged 73	.18	3	.95	.81
6. mildly disorientated / good circulation	.10	2	.47	.27
7. occasionally incontinent / bedfast and immobile in bed	.03	4	.99	.99
8. aged 73 / good circulation	.29	6	.17	.12
9. mildly disorientated / bedfast and immobile in bed	.07	1	.99	.79
10. good control of bowels / aged 73	.29	6	.17	.63
11. good circulation / occasionally incontinent	.04	3	.14	.71
12. mildly disorientated / normal blood pressure	.18	2	.55	.52
13. bedfast and immobile/protein deficient nutritional state	.09	2	.81	.87
14. normal blood pressure / protein deficient nutritional state	.26	4	.16	.31
15. good control of bowels / normal blood pressure	.62	1	.08	.11
16. aged 73 / normal blood pressure	.44	5	.37	.21
17. protein deficient nutritional state / mildly disorientated	.18	3	.67	.62
18. good circulation / bedfast and immobile in bed	.08	1	.67	.61
19. occasionally incontinent / good control of bowels	.03	2	.20	.92
20. protein deficient nutritional state / aged 73	.19	1	.62	.71
21. normal blood pressure / good circulation	.57	1	.16	.16
22. protein deficient nutritional state / good control of bowels	.15	5	.31	.15
23. mildly disorientated / occasionally incontinent	.03	5	.75	.95
24. bedfast and immobile in bed / good control of bowels	.06	3	.62	.81
25. aged 73 / occasionally incontinent	.07	7	.45	.55
26. good circulation / protein deficient nutritional state	.12	3	.25	.31
27. occasionally incontinent /protein deficient nutritional state	.02	6	.66	.75
28. mildly disorientated / good control of bowels	.07	4	.42	.62
median	.11	2.5	.44	.62

16 pairs were needed for full multivariate testing of the predictions since there were 2 levels (above and below median) of each of the 4 factors - 2x2x2x2 = 16. The task then became one of

identifying a suitable pair which satisfied each of the following 16 combinations:

(O = Co-occurrence; T = Text; R = Risk; C = Care)
('+' = above median; '-' = below median)

combination	combination
1. O+ T+ R+ C+ .... pair 5	9. O- T+ R+ C+ .... pair 27
2. O+ T+ R+ C- .... pair 17	10. O- T+ R+ C- .... pair 25
3. O+ T+ R- C+ .... pair 10	11. O- T+ R- C+ .... pair 11
4. O+ T+ R- C- .... pair 14	12. O- T+ R- C- .... pair 28
5. O+ T- R+ C+ .... pair 20	13. O- T- R+ C+ .... pair 13
6. O+ T- R+ C- .... pair 12	14. O- T- R+ C- .... pair 18
7. O+ T- R- C+ .... pair 1	15. O- T- R- C+ .... pair 19
8. O+ T- R- C- .... pair 21	16. O- T- R- C- .... pair 4

(note: for 5 combinations there was more than one pair which satisfied the criteria. As a serious test of the null hypothesis, on these occasions the pair which closest approximated to the median of the Care index was chosen)

Analysis could now proceed using the ratings given by subjects for each of these 16 pairs. The 4 factors became within subject variables, each with 2 levels. Subjects were analysed in their 3 groups of n=17, hence there was one between subjects variable (Experience) with 3 levels. The ANOVA, therefore, was a complex 3x2x2x2x2 multivariate set-up. Despite the complexity, however, the effects of interest are confined almost exclusively to the possible interactions between each within subjects variable and Experience. For clarity, therefore, it is intended to focus mainly on of these interactions. The full source table for the ANOVA is given in Appendix 7.

The first finding which can be noted is that there was no significant main effect for the between subjects factor of Experience (F=.257, df2 and 48, p=.77), indicating that the three groups did not vary across the overall range of the rating scale. Each within subjects factor can now be inspected in turn.

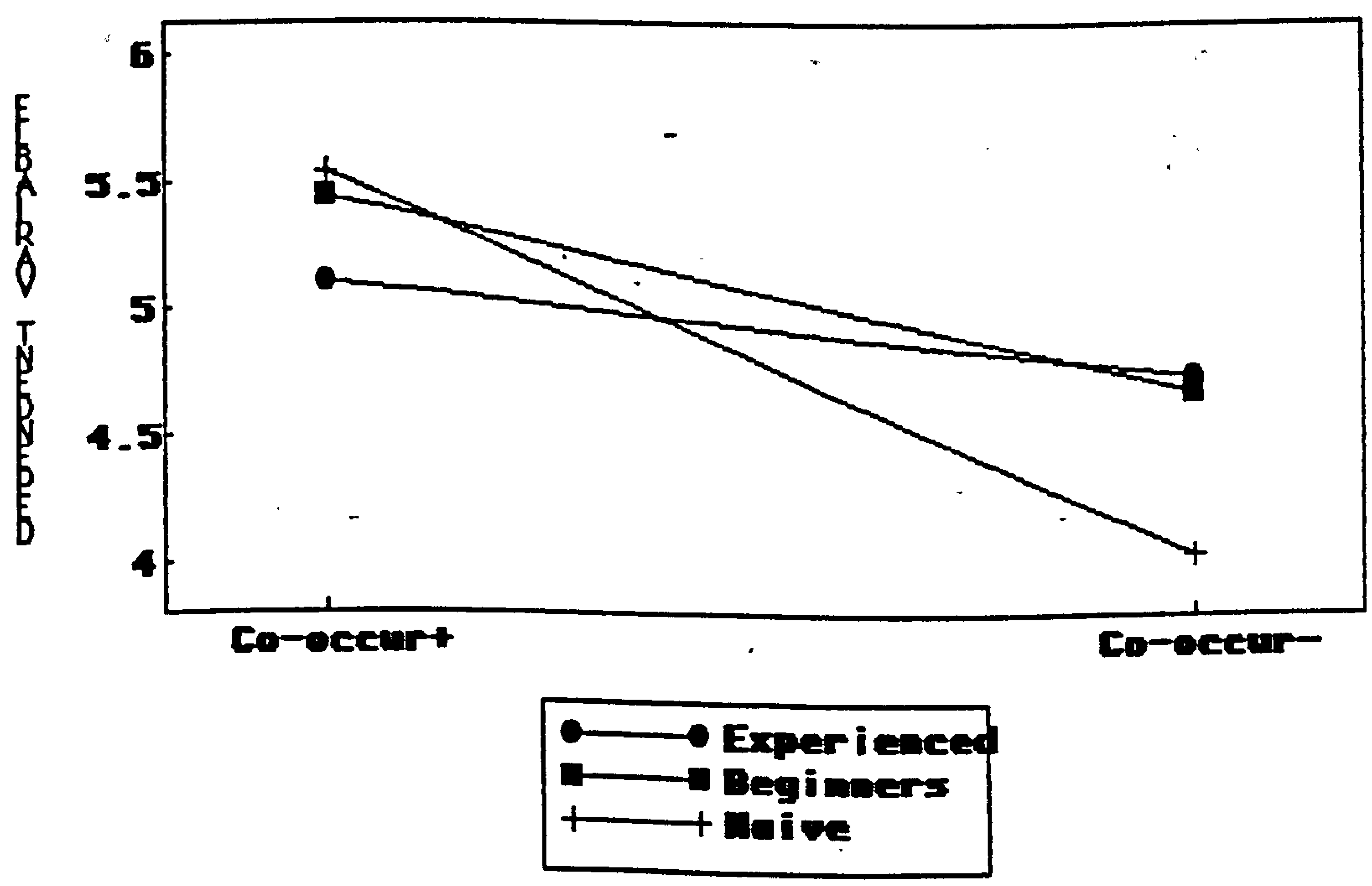
Factor 1 - Co-occurrence   The mean values for each level of this factor by group were as follows:

subjects	Co-occurrence+	Co-occurrence-
Experienced	5.09	4.73
Beginners	5.42	4.68
Naives	5.53	4.02
overall mean	5.348	4.475

The main effect for Co-occurrence was highly significant (F=34.54, df1 and 48, p<.001). The Experience:Co-occurrence interaction, shown graphically in Figure 4.4, was also significant (F=5.1, df 2 and 48, p<.01). Simple effects analysis of this interaction showed that the effect of Experience at both Co-Occur+ and at Co-Occur- was not significant. There was, however, a differential effect of Co-occurrence at each level of Experience, as the following summary results indicate:

Co-occurrence at Experienced - F=2.04, df 1 and 48, not significant
Co-occurrence at Beginners - F=8.34, df 1 and 48, p=.006
Co-occurrence at Naives - F=34.4, df 1 and 48, p<.0001

Figure 4.4 Mean values for Co-occurrence factor by Group



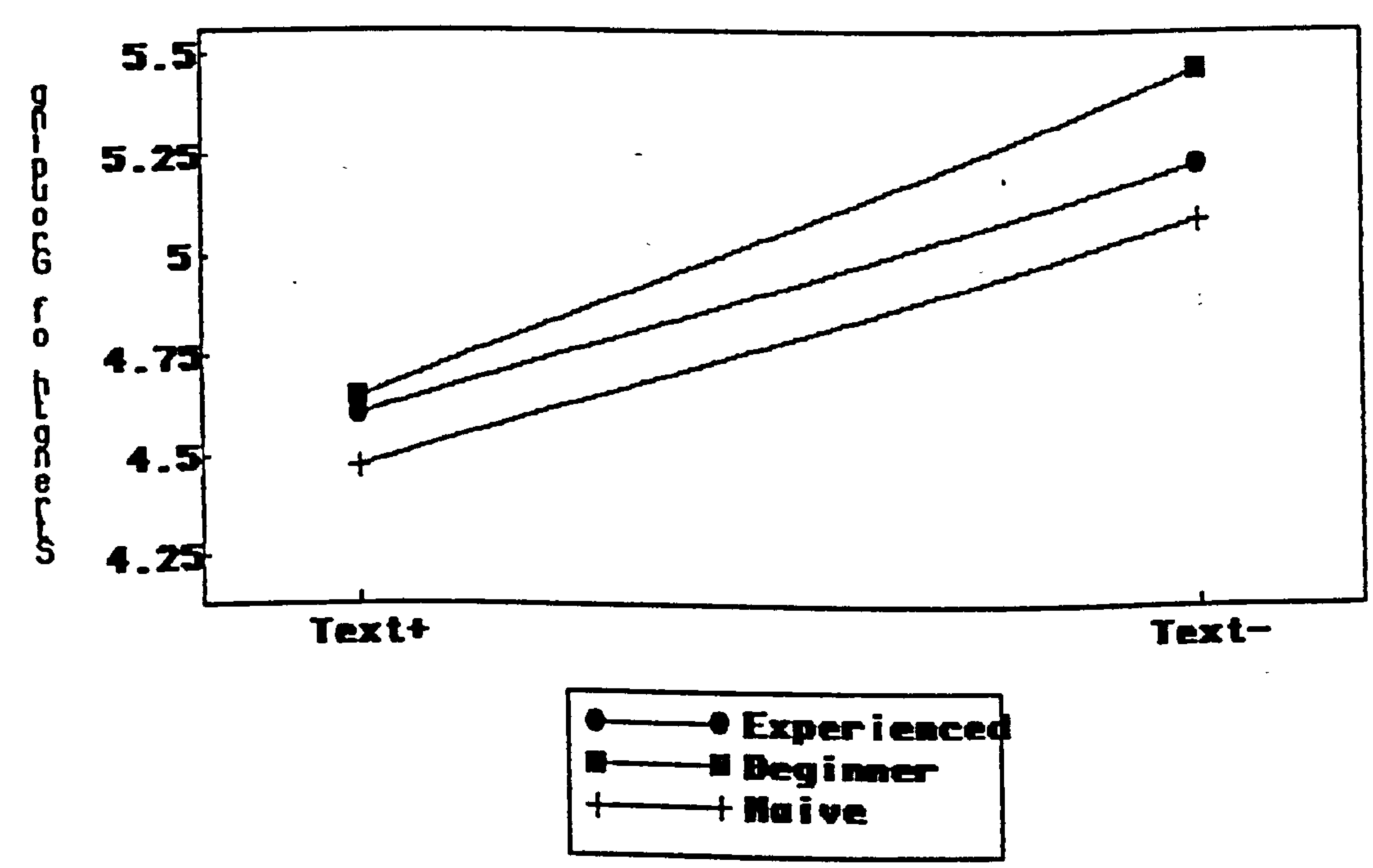
Factor 2 - Text The mean values for each level of this factor by group were as follows:

subjects	Text+	Text-
Experienced	4.59	5.23
Beginners	4.64	5.45
Naives	4.46	5.09
overall mean	4.567	5.257

The main effect for Text was highly significant (F=25.78, df1 and 48, p<.001). The Experience:Text interaction, shown graphically in Figure 4.5, clearly was not significant (F=.21, df 2 and 48, p=.81). Simple effects analysis of this interaction showed that the effect of Experience at both Text+ and at Text- was not significant. The significant main effect of Text, therefore, seemed to be fairly uniform at each level of Experience, as the following summary results of simple effects indicate:

Text at Experienced - F=7.19, df 1 and 48, p=.01
Text at Beginners - F=11.98, df 1 and 48, p=.001
Text at Naives - F=7.03, df 1 and 48, p=.011.

Figure 4.5 Mean values for Text factor by Group



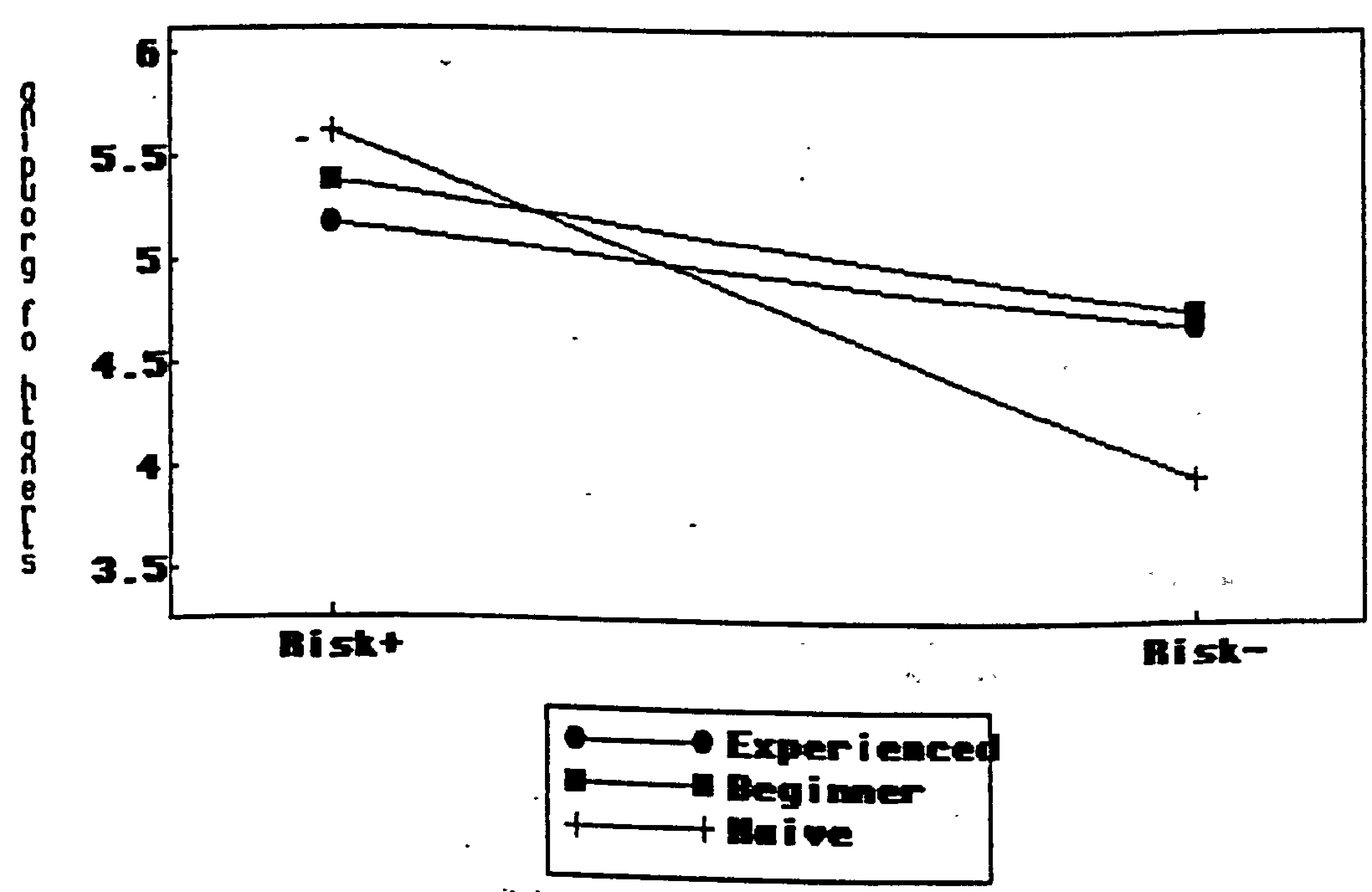
Factor 3 - Risk The mean values for each level of this factor by group were as follows:

subjects	Risk+	Risk-
Experienced	5.15	4.67
Beginners	5.36	4.73
Naives	5.60	3.95
overall mean	5.372	4.451

The main effect for Risk was highly significant (F=54.6, df1 and 48, p<.001). The Experience:Risk interaction, shown graphically in Figure 4.6, was also significant (F=8.74, df 2 and 48, p<.001). Simple effects analysis of this interaction showed that the effect of Experience at both Co-Occur+ and at Co-Occur- was not significant. There was, however, a differential effect of Co-occurrence at each level of Experience, as the following summary results indicate:

Risk at Experienced - F=5.05, df 1 and 48, p=.03
Risk at Beginners - F=8.37, df 1 and 48, p=.006
Risk at Naives - F=58.6, df 1 and 48, p<.0001.

Figure 4.6 Mean values for Risk factor by Group



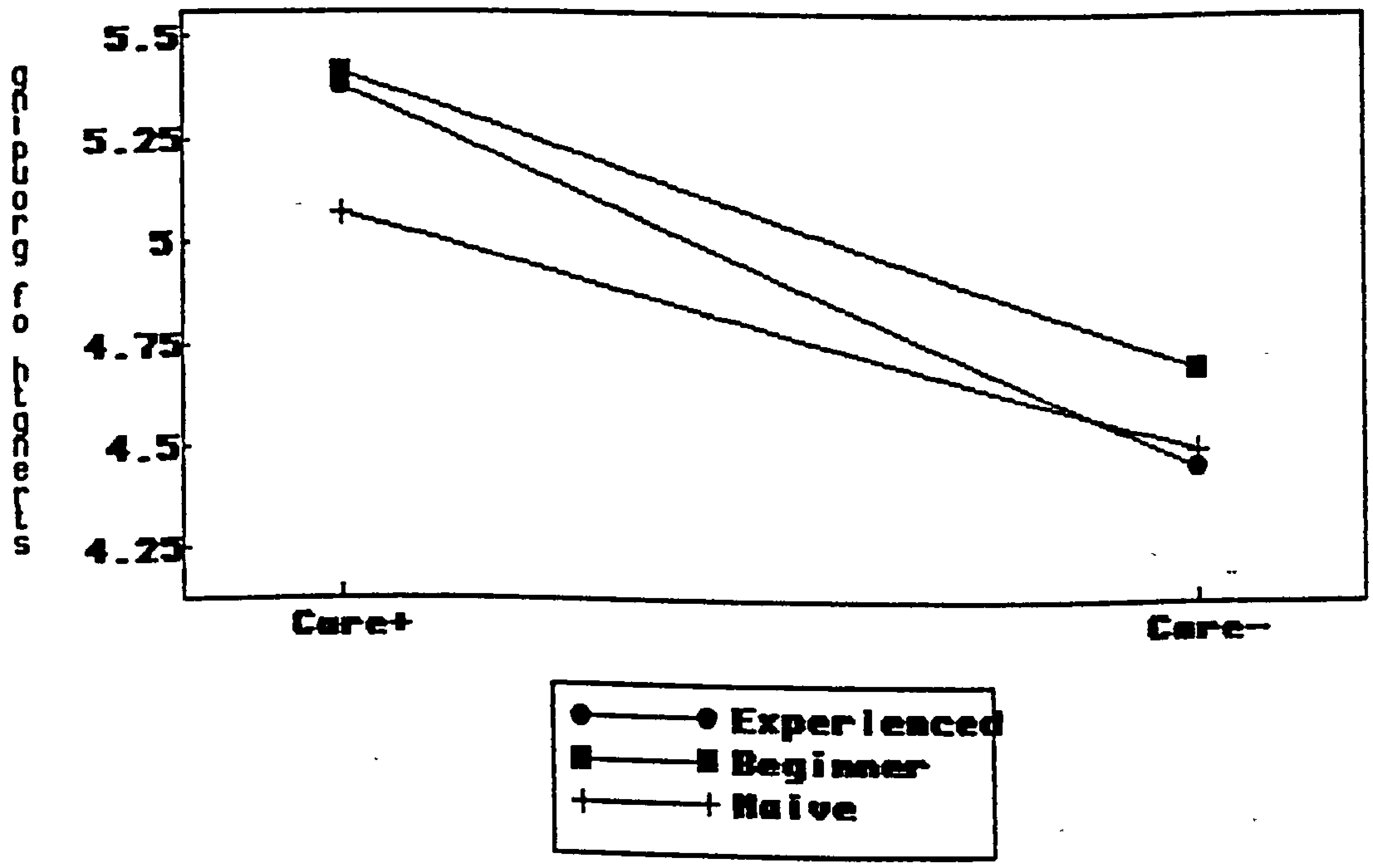
Factor 4 - Care The mean values for each level of this factor by group were as follows:

subjects	Care+	Care-
Experienced	5.37	4.45
Beginners	5.40	4.69
Naives	5.06	4.49
overall mean	5.279	4.544

The main effect for Care was highly significant ($F=17.54$, df1 and 48, $p<.001$). The main Experience:Care interaction, shown graphically in Figure 4.7, was clearly not significant ($F=.35$, df2 and 48, $p=.70$). Simple effects analysis of this interaction showed that the effect of Experience at both Care+ and at Care- was not significant. There was, however, a differential effect of Care at each level of Experience, as the following summary results indicate:

Care at Experienced - $F=9.28$, df 1 and 48, $p=.004$
Care at Beginners - $F=5.50$, df 1 and 48, $p=.023$
Care at Naives - $F=3.37$, df 1 and 48, not significant.

Figure 4.7 Mean values for Care factor by Group



One final overall result which should be noted is that mean ratings given by each group for each factor were without exception different in the same direction for all groups of subjects. Hence,

Co-occurrence+ pairs has a higher mean value then Co-occurrence- pairs,
Text+ pairs has a higher mean value then Text- pairs,
Risk+ pairs has a higher mean value then Risk- pairs, and
Care+ pairs has a higher mean value then Care- pairs.

Discussion

Of chief interest in the analysis is the expansion of interactions between subject groups and each factor into simple effects. There appears to be no sound requirement for looking at the higher order interactions, however before beginning the discussion of the results of interest there remains one other effect - the between subjects main effect. That the Experience effect was not significant is of some interest. One one hand it could be argued that this result challenges the predictions made for this experiment. However, more thoughtful consideration of these predictions suggests that there is no real basis for the ratings of one group being 'generally' higher or lower than the ratings given by another group or groups. If, for example, the Experienced nurses felt more confidence when grouping strengths of some 'care implicating' pairs, then the high ratings here would only come to be offset by the low ratings given to non-care implicating pairs.

Of the four factors, the first point to note is the overall success of their power to explain variance of ratings - each factor was reliably significant. These overall effects are, however, too 'coarse' since it can be seen that only rarely did all groups of subjects provide uniform endorsement of the factors. These four factors can now be considered in turn.

The Co-occurrence factor refers to the extent to which a pair of items 'go together in the world', and as such can be taken as representing a prototypical basis to representation. Frequently co-occuring items (Co-occurrence+) might be grouped more closely than rarely co-occuring items, although in terms of the main hypothesis there would no 'functional' or care-implicating basis for chunking items on the basis of co-occurrence. The analysis of simple effects has supported both of these predictions in that both Beginner nurses and (in particular) Naive subjects were strongly influenced by co-occurrence. The variability of the ratings given by Experienced nurses, on the other hand, were not significantly explained by this factor. This finding suggests that a representation based on superficial features may exist for non-expert but not for expert nurses.

The Risk factor refers to the level of problem represented by a pair of items. Items in a pair which stood out as difficulties for the patient (Risk+) might be closely chunked particularly by the nursing subjects since there would be functional reasons for doing so. Moreover, Risk+ items may affect the representation of subjects due to their relative salience. These predictions were supported in that Risk+ pairs were rated significantly higher than Risk- pairs by both Experienced and Beginner nurses. However, the Naive group indicated the most dramatic endorsement of this factor with a large discrepancy between Risk+ and Risk- ratings.

If the items were of an obscure medical nature such that Naive subjects could not know if something was or was not a problem then this result would clearly indicate an experimental artefact or similar difficulty. However, the fact that the items in this patient description were clearly understandable to the lay person indicates rather that Risk salience may reasonably taken as a universal basis for chunking - it therefore becomes more interesting that the Experienced nurses were comparitively weak in their endorsement of this factor ($p<.05$ compared to $p<.0001$ for Naive subjects). It suggests that one or more of the other factors must have been more powerful in explaining the basis to Experienced nurses ratings.

The Text factor refers to the proximity of a pair's items within the patient description. It was predicted that the recency with which the description was read would result in items which were close neighbours (Text-) being most strongly rated, in particular by non-Experienced nurses who lack the more important functional basis for memory organisation and who may consequently be more influenced by superficial features. The results indicated that textual proximity was indeed important and that it was important almost uniformly across all groups. This finding, it is suggested, does not weaken the main conclusions about functional bases for mental representation since it should be borne in mind that the mode of passing the patient information to the subjects was artificial - in reality nurses would more commonly receive patient information via various senses, over a greater spread of time, and with more intervening information.

The Care factor, lastly, is the factor which most directly tests the principal hypothesis of the experiment - that the extent to which a pair of items implicate care will reflect the strength of chunking of that pair by the Experienced nurses in particular. Results supported the prediction in that it was only the nursing subjects who rated the Care+ pairs significantly higher. Moreover, of the two nursing groups, the Experienced nurses were more positive in their endorsement ($p<.01$) than the Beginner nurses ($p<.05$).

Of the groups of subjects, then, it becomes possible to achieve an overview of how each group has indicated the bases for organisation of their representations. It is also possible to identify the most important factor on the basis of variance explained. The Naive group, firstly, seem most influenced by risk salience. For the Beginner group, it is textual distance, and, for the Experienced group, the most important factor was shown to be care implication. This overview is instructive and requires only an additional comment related to the Beginners favouring of textual distance. Perhaps, as a suggestion, this finding is influenced by the fact that Beginners were currently in a classroom setting where rote memorisation could have been a commonly adopted style of learning

Relating these findings to those set out in the introduction will be undertaken following presentation of results from Experiment 2.

Experiment 2 - Unexpected Recall Task

The aim of this experiment, as explained in the introduction, was to further investigate possible expert/novice differences in relation to the mechanisms of memory organisation. The assumption is that the order in which items form the patient description are recalled will reflect active memory organisation effects, and that these structures can be recovered using hierarchical cluster analysis. A secondary aim was to analyse the content of memory by looking at patterns of which items were recalled successfully and which were not recalled, however this aim was abandoned when it became clear that the stimuli contained insufficient items for clear differences to emerge (see results below).

The general hypothesis is that the order in which items are recalled will reflect either a superficial or deeper level representation. Specifically, an analysis of the order in which items were recalled was planned which might demonstrate that the proximities between items on recall had become different from the proximities of items within the text. In particular, it is predicted that there will be a functional (ie care-implicating) basis to the active organisation of items in the memory of experienced subjects.

Method

The recall task was contained in the last page of the materials booklet. Hence all subjects completed the task after having first rated the item pairs in Experiment 1. An additional intervening task between reading of patient description and the recall task was undertaken by subjects in order to further ensure that the recall task was unexpected. This task was a simple series of questions relating to experience and qualifications (see booklet in Appendix 6). The recall task contained instructions for subjects to write down as many as possible of the items contained in the patient description which they could remember.

Preliminary analysis Initial analyis established the raw scores simply of recall of all items by group. The maximum which could be recalled =11, comprising 5 'high problem' items, 3 'low problem' items, and 3 'care' items. Successful recall was judged on the basis of presence or absence of a given item. Since organisation rather than content of memory was of principal interest, an item was scored as recalled if the essential features of that item were present. Hence, for example, additional or absent prepositions were permitted. In the event, there were no serious inaccuracies which resulted in an item being scored as absent. Whether this was due to the repetition of items during the rating task or whether due to the small number and straightforward nature of the items is not clear. Descriptive data of mean number of items recalled by Group are set out in Table 4.12.

Table 4.12 Descriptive data of number of items recalled by Group.

group	mean items recalled	sd
Experienced	9.94	.83
Beginners	9.59	1.03
Naives	6.65	1.78

Between group testing with independent samples t test (one tailed) showed that the Naive subjects recalled significantly less items than both the Beginners ($t=5.7, df 16, p<.001$) and the Experienced group ($t=6.7, df 16, p<.001$). Since there were no differences between the two nursing groups and since the number of items recalled by these groups was near perfect, it was decided to abandon comparative testing of which items were recalled versus which were forgotten.

The main analysis performed on the data, then, tested for memory organisation effects. The 8 'description' items as well as the 3 'care' items were analysed in order to test for such effects.

It was predicted that care items would serve to organise representation of patient information, therefore the care items would be recalled 'amongst' rather than separately from the description items.

It was firstly necessary to compute an index of the distance between items when recalled by subjects in each group. This was achieved by the SPSSx PROXIMITIES procedure where the position of recall (range=1 to 11) for each item was analysed across the 17 subjects in each group. On the relatively few occasions when an Experienced or a Beginner subject failed to recall an item then the number entered was the median position of recall for that item by that group. The Naive subjects, however, were much more frequent in failing to recall items. Since substituting missing values with a group figure would have demanded acceptance of potentially weak assumptions, the analysis of memory organisation effects was restricted to the two groups of interest - the nursing groups - and not performed for the Naive group.

PROXIMITIES produced a matrix of squared Euclidean distance coefficients for all possible pairs of the 11 items. A matrix was prepared for each of the Experienced and the Beginner groups. The requirement from these matrices was a procedure for identifying subgroups of items (or clusters) based on the proximity measures. To achieve this, each matrix was input to the SPSSx CLUSTER procedure using the hierarchical agglomeration algorithm based on average linkage between groups. The algorithm operates by initially considering each item as an individual cluster. From these 11 clusters, at step 1 the two 'closest' items are combined into a single cluster - hence forming 10 clusters. At each subsequent step an additional cluster is formed either by joining an item to an already existing cluster; two separate cases into a single cluster; or two multi-item clusters until all 11 items are merged into a single cluster.

Results - Cluster analysis

Results are presented for the Experienced and the Beginner groups in the form of Agglomeration Schedules and Dendrograms. The Agglomeration Schedule contains the number of items or clusters being combined at each step. In Table 4.13 overleaf, for example, the first line indicates under 'Clusters Combined' that items 9 and 10 (sheepskin and position) at this stage. The squared Euclidean distance between these two clusters (items at this point) is displayed under in the column 'Coefficient'. 'Items merged' refers to which items or clusters of items were joined at each stage. A ' + ' (in bold typeface) denotes the items or clusters which are being joined.

A Dendrogram visually represents the steps in the hierarchical clustering solution. The clusters as they are combined are shown along with the values of the coefficients at each step. Produced by the SPSSx CLUSTER procedure, the dendrogram does not plot the actual proximity coefficients of each agglomeration step, rather the coefficients are rescaled to numbers between 0 and 25. The ratio of the distances between steps is, however, preserved.

Figure 4.8 overleaf depicts the results of the cluster analysis, in the form of a Dendrogram, for the Beginner subjects. Taken in conjunction with the Agglomeration Schedule (Table 4.13), it can be seen that there were 3 relatively close pairs of items. These pairs can be listed along with their respective squared Euclidean distance coefficient (rather than rescaled distance):

PAIR	Sq. Euclidean Dist.
sheepskin + position	6.63
circulation + BP	8.48
bowels + urine	8.83

The point at which the 3 'care' items were merged with the 'description' items (excepting age) was not until step 9 where the squared Euclidean distance was relatively large (=17.41).

Table 4.13 Agglomeration Schedule for cluster analysis on position of item recall for Beginner subjects

	Clusters Combined			
Step	cluster 1	cluster 2	Coefficient	items merged...
1	9	10	6.63	9 + 10
2	5	6	8.48	5 + 6
3	7	8	8.83	7 + 8
4	2	7	11.31	7&8 + 2
5	9	11	11.31	9&10 + 11
6	3	5	13.41	5&6 + 3
7	2	4	13.68	7&8&2 + 4
8	2	3	14.47	7&8&2&4 + 5&6&3
9	2	9	17.41	7&8&2&4&5&6&3 + 9&10&11
10	1	2	22.37	7&8&2&4&5&6&3&9&10&11 + 1

Figure 4.8 Dendrogram for cluster analysis on position of item recall for Beginner subjects

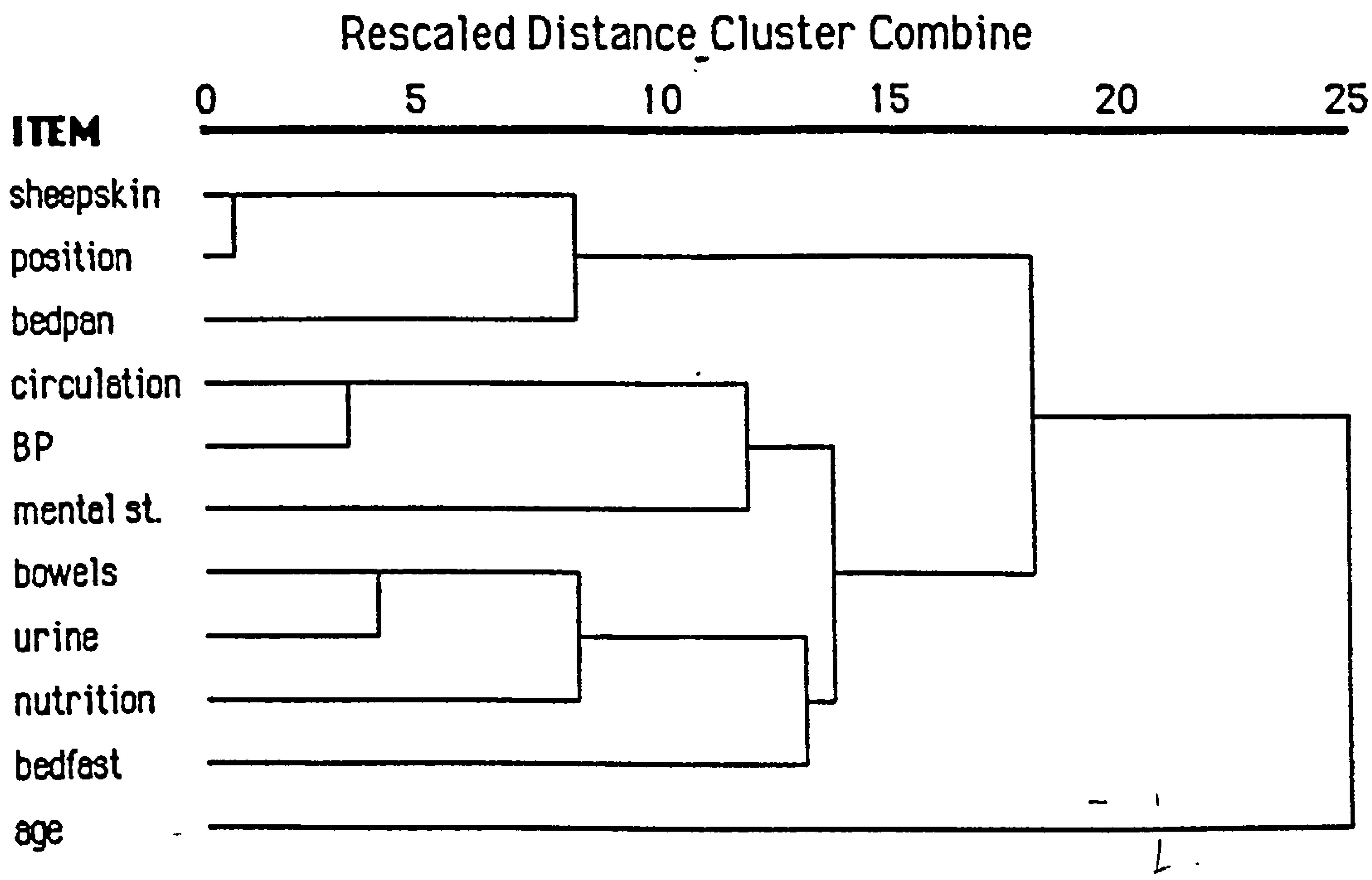


Figure 4.9 depicts the results of the cluster analysis, in the form of a Dendrogram, for the Experienced subjects. Taken in conjunction with the Agglomeration Schedule (Table 4.14 overleaf), it can be seen that there were 2 relatively close pairs of items. These pairs can be listed along with their respective distance coefficients:

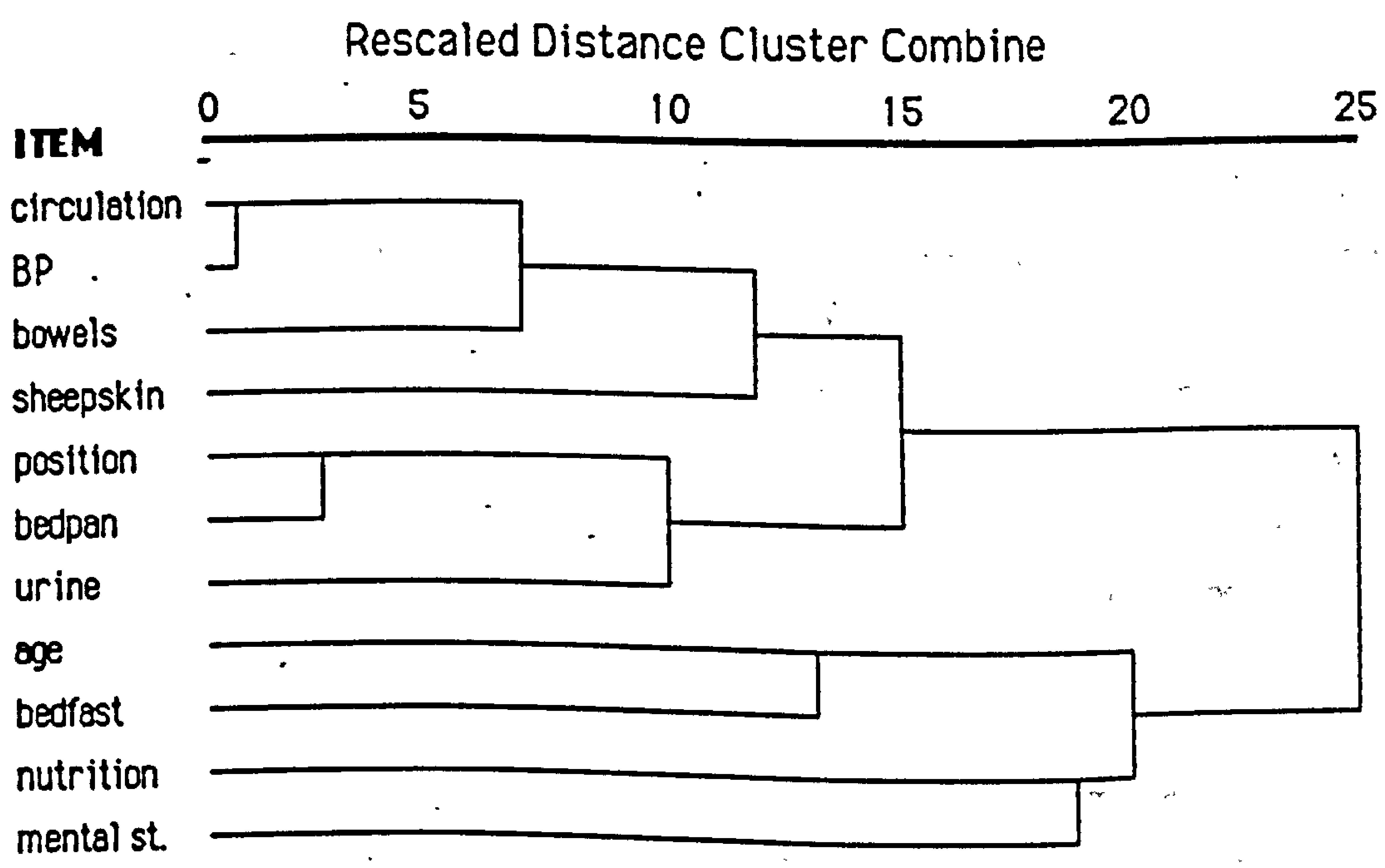
PAIR	Sq. Euclidean Dist.
circulation + BP	6.40
position + bedpan	7.74

The point at which care items begin to be merged with description items is step 4 (coefficient=11.84).

Table 4.14 Agglomeration Schedule for cluster analysis on position of item recall for Experienced subjects

	Clusters Combined			
Step	cluster 1	cluster 2	Coefficient	items merged....
1	5	6	6.40	5 + 6
2	10	11	7.74	10 + 11
3	5	7	9.75	5&6 + 7
4	8	10	11.84	10&11 + 8
5	5	9	12.93	5&6&7 + 9
6	1	4	13.30	1 + 4
7	5	8	14.08	5&6&7&9 + 10&11&8
8	2	3	16.46	2 + 3
9	1	2	17.09	2&3 + 4&2&1
10	1	5	20.09	5&6&7&9&10&11&8 + 2&3&4&2&1

Figure 4.10 Dendrogram for cluster analysis on position of item recall for Experienced subjects



One further aspect of interest from inspection of the two Dendrograms is the extent to which 'problem' items are clustered with 'non problem' items. For the Experienced group, the 3 'non problem' items (circulation, bowels, BP) formed into a cluster at a relatively early point (step 3) and did not merge with a cluster containing 'problem' items (urine) until step 6. The picture for the Beginner group differed in that a non problem item had combined with a problem item at step 3 (bowels + urine) followed by more general merging at steps 4 and 6.

Discussion

At the gross level it has been shown that the two groups of nurses are more successful in recall of patient details than the non-nursing subjects. Although this effect was seen as applying to both high problem items and to low problem items, it is nevertheless unremarkable given the relative unfamiliarity of the terms to non nurses.

The cluster analysis provided evidence to back the finding of Experiment 1 related to memory organisation effects. Given that the original text separated patient description items from care items, it might be predicted that the intervening task of rating only the description pairs would further separate these two classes of items. The main hypothesis, however, predicts that experienced nurses will mentally represent the patient they were presented with by organising the description items around the care these descriptions imply.

The cluster analysis, based on the proximity of item positions on recall, strongly suggested that the organising effect was sufficiently robust in the Experienced nurses to overcome any separation occasioned by the textual and intervening task. Hence, in the original text the care items were presented in a homogeneous group at the end of the description. However, care items seemed to be recalled by these nurses among description items. This picture, however, did not emerge for the Beginner nurses - evidence that the care items were recalled in a rather homogeneous group (once more at the 'end') is found when the cluster containing these items was only seen to combine with description items at the penultimate (ninth) step.

This finding is supported by an artefactual problem in the experimental design in that the intervening self-rating task did not utilise the 3 care items. It might be expected, therefore, that the separation of 'description' from 'care' items brought about by text would be further augmented by the rating task. Despite this, it was found that experienced subjects showed amalgamation of care and description items, suggesting a robust effect.

Of even greater interest is the nature of the observed clustering between care and description items. For example, the twin facts that the patient was incontinent of urine and that she required a bedpan as necessary were clustered by experienced nurses relatively early at step 4. Since these two facts are directly related in terms of cue-and-nursing response, there is consequently a strong suggestion of a functional basis to memory organisation. This conclusion is, however, only suggested by this finding and the general amalgamation of care and description items. Other predicted linkages do not bear out the hypothesis – for example, 'bedfast' and 'turn 2 hourly' do not closely cluster in the experienced nurses' dendrogram.

It would be misleading to concentrate simply on apparent evidence for deeper level representation. Superficial features, especially textual proximity, seem to affect clustering, particularly for the Beginner subjects. Hence each of the three pairs of items which are most closely clustered by Beginners can be seen to be 'neighbours' in the text (sheepskin and position, circulation and blood pressure, urinary and bowel continence). However, it not possible to draw safe conclusions from this finding since each of these pairs could also be taken as functionally corresponding in terms of deeper conceptual knowledge. To take as this conceptual model the Activities of Living framework, it can be seen that circulation and blood pressure belong to 'Breathing' while urinary and bowel continence each implicate 'Elimination'.

Further observations could be made about the apparent separation of 'problem' items from 'non-problem' items, particularly for the experienced subjects. Hence the 3 non-problem items (circulation, blood pressure, and bowels) are clustered together in the top half of the experienced nurses' dendrogram and do not join with the 5 problem items until the last linkage. Once again, however; safe conclusions cannot be drawn in support of findings from Experiment 1. Clearly these possibilities lead to a suggestion for further experimental work – different versions of the same text where textual order was manipulated would be a possibility.

<u>General Discussion and Conclusions</u>

The first general finding is that both Experiment 1 and 2 have demonstrated active memory organisation effects for patient information, a finding which was strongly predicted from the review of literature. The more specific finding has been that a distinction has been found with respect to experienced and beginner nurses in terms of the influence of superficial and deeper level effects on knowledge representation. Hence in Experiment 1 beginner nurses (but not experienced nurses) were strongly influenced by co-occurrence (prototypical) relations between items. Conversely, the care implications of related items was found to influence grouping of items by experts but not by beginners. Each Experiment supported these findings. Given that both risk and textual factors were found to influence all subjects, the general conclusion which can be made is that support has been demonstrated for the dual representation model outlined in the introduction.

It is interesting to note the processing implications of this dual representation model. A further finding of Chi et al (1981) was that an activated schema was used by experts to gather further information necessary to construct their functional representations - suggesting an interaction between both backward and forward reasoning strategies. A similar attention to construction of an elaborated representation was found by Bhaskar and Simon (1977) in their demonstration that experts tend to work forward by using cues to make inferences about information needed to solve the problem.

There are signs also that medical diagnostic studies are adopting this model. Hence Feltovitch, Johnson, Moller and Swanson (1984) show that experts' disease knowledge contains procedures for differentiating between diagnoses ('Logical Competitor Sets') and that expert behaviour was characterised by an interaction between lower and higher levels of representation which was not purely top-down or bottom-up. Patel and Groen (1986) have gone further by reporting findings which support the finding that the medical diagnostic process can be fully explained by forward reasoning from a knowledge based representation.

The focus of nursing authors, however, seems to be firmly in the tradition of hypothesis testing of incoming data to stored prototypical patient representations. However, unlike the model offered by Chi et al (1981), nurses have tended to stress that this 'patterm matching' should be accomplished by <u>conscious</u> processing. One reason for this prescription is the few findings that novice nurses often fail to recognise cues or to group them meaningfully (del Bueno 1983). Another explanation could be the strong political interest in nationally-defined nursing diagnoses.

For whatever reason, noted authorities such as Carnevali (1983) and Gordon (1987) have responded with models which are essentially based on inference and conscious pattern matching. Hence incoming patient cues are to be *consciously* sorted and clustered according theoretical concepts. When a diagnostic cue is noticed the nurse is to activate the appropriate initial hypotheses from memory and deliberately seek confirmatory evidence in order to match this patient with the groups of necessary and sufficient cues which have set down nationally as definitive of each diagnosis.

Medical decision making studies are cited in support of these models. As discussed in the introduction, however, this evidence itself is not conclusive. Moreover, strong reasons were offered for doubting the applicability of medical models since the nursing situation typically involves a greater number of cues and a many more than one final diagnosis. No account taken of the overwhelming memory load which this nursing assessment implies - in one recent study Corcoran (1986) asked nurses to verbalise their diagnostic plan after having read patient case histories. The information was in 20 categories relating to pain problems, however this represents only a fragment of what would normally be available. Nevertheless, the nurses generated up to 35 alternative courses of action before making decisions which often disregarded critical data.

The point to be made concerning Corcoran's (1986) almost unique work is that some of the more expert nurses showed great accuracy with their eventual decisions. Her research was essentially descriptive. However if a prescription for nursing decision making could be based on process analysis studies of the cognition of expert nurses then by implication the resulting decision models would have demonstrable validity. This credo has driven the present project. As Benner (1984) has shown, expert nurses have developed perfectly adequate cognition - an understanding of the mental representations and processing mechanisms of these nurses would provide theoretical models which were not potentially castles made of the sands of supposition.

It is interesting that Abraham (1988) has suggested a shift in theoretical formulations of nursing knowledge structures by proposing the notion of 'nursing diagnostic structure' which includes not only knowledge of prototypical diagnoses but also knowledge of the interventions which are associated with each of them. The findings from the present study have made a beginning in this quest for a research-validated understanding.

Lastly, it was mentioned above that areas of further work have been suggested. The experimental design of such studies could incorporate testing of the validity of a conceptual model of pressure sore aetiology which was discussed and to an extent supported by the analyses in Chapter 3. This model conceptualised four dimensions to the deeper knowledge which an experienced nurse holds of pressure sore risk assessment and care planning. The results reported above, as far as they go, can be seen to support firstly the suggestion of deeper level representation and secondly the implication of planning care. The specific nature of this deep knowledge, however, must await confirmation from further research.

CONCLUSIONS & RECOMMENDATIONS

Conclusions and recommendations arise from various parts of the this thesis. These can be presented under the following broad headings:

'Traditional' CAL in nursing It is concluded that there are several grounds for tempering claims which have been made regarding the present state of the innovation. Aside from the excessive nature of these claims, the stongest point supporting this conclusion is the notable absence of adequate evaluation. It is therefore recommended that well-designed evaluation studies are undertaken which have as a starting point the requirement to properly specify the goal of the innovation.

It is also felt that the methodology of traditional CAL would not provide a suitable vehicle for the simulation of cognitive processing rather than the more straightforward presentation of 'factual' knowledge. This feeling can be translated into an empirical question; it is therefore recommended that comparitive studies are undertaken to clarify the limitations and applicability of traditional CAL methodology.

Intelligent CAL in nursing The conclusion which can be drawn from the general field of ICAL is that this innovation offers greater potential for simulating processing aspects of nursing cognition. Nevertheless, such non-nursing applications which have been reported to date have taken seriously the goal of emulation. That is, there has been a lack of adequate study of human cognition prior to construction of computer programs. Ahead of further work in this field, therefore, this conclusion must be offered as tentative.

The present study, it is contended, goes some way toward augmenting the potential of ICAL with respect to nursing. It is borne in mind, however, that the knowledge domain of pressure sore risk assessment may possess characteristics which limit optimism for the achievement of ICAL. The degree to which knowledge can be specified is crucial for the use of the approach adopted in this study. It could be that knowledge of, for example, the factors predicting imminent aggression in a patient is of a considerably more 'nebulous' nature. It is therefore recommended that further work be undertaken to assess the success of modeling other knowledge domains such as the aggression example cited above.

Methodology The complexity of the task of undertaking adequate study of human cognition cannot be underestimated. In the present project, study of a relatively circumscribed knowledge domain involved considerable work given the declared intention of striving at all times to incorporate reliable data from valid methodology. A further problem arose from the attempt to build a model based on several 'experts', rather than on one individual. If a contribution to methodology has been made by this project, then it might firstly concern the stepwise approach to knowledge acquistion, and secondly the juxtaposition of behavioral data from the Process Tracing experiment with qualitative analysis of verbalisations made by subjects.

Alternative approaches to modeling, such as problem state specification and Protocol Analysis, were considered prior to the adoption of the approach used. Given the uncertainty surrounding a choice of appropriate method, it can be recommended that this question is deserving of further study.

Nursing cognition Two findings from the present project can be seen to challenge positions taken elsewhere. Firstly, empirical differences in terms of measured cognition have been demonstrated between nurses deemed as expert and those deemed as proficient. Secondly, there is evidence to suggest that flaws exist in the prevailing view that nurses employ a 'diagnostic hypothesis' style of patient assessment which relies on superficial descriptors of the patient. This project has argued that a 'deeper' representation involving action schema seems to more adequately fit the evidence.

Given, however, the strong North American nursing focus on diagnosis as an apparently discrete goal in patient assessment, it becomes imperative that further experimental (rather than theoretical) work is carried out in order to illuminate the nature of expert nursing cognition. Given the exploratory nature of the present study, it is appropriate that the method and conclusions are subjected to both replication and extension to cover additional knowledge domains.

The management of further research deserves consideration given that skills and resources necessary cannot be located solely within a single discipline. To an extent, this is true of all applied psychology ventures. However, in this case it can be seen that there is a requirement for resources from nursing practice and theory, from experimental and empirical psychology, and from Informatics/Artificial Intelligence. An appropriate umbrella term would be Human Computer Interaction (HCI). By implication, a collaboration can be recommended between centres where these resources can be located.

Given the imminent investment worldwide in Nursing Information Systems, there would be considerable gain from achieving an understanding of nursing cognition.

APPENDIX 1 Listing task - pressure sore risk attributes

Thank you for helping with this research project which is sponsored by the Scottish Home and Health Department.

The purpose of the project is to prepare teaching packages which are based on ward-based expert nurses' knowledge. At this early stage I am trying to find which are the important factors used by nurses when they are assessing 'risk' in their patients. In the two examples here you will be asked about the factors which predict risk of a patient developing pressure sores and, secondly, risk of a patient behaving aggressively toward nurses.

Please make a list of the factors which, in your experience, are generally the factors which predict that a patient will develop pressure sores. Write down as many or as few as you think fit, but please don't confer.

Please turn over for the other example.

APPENDIX 2 Instrument for collection of patients comprising database 1

Western Infirmary Glasgow

Scottish Home and Health Department

Gartnavel General Hospital

PRESSURE SORE SURVEY

A number of senior clinical nurses in your hospital have been cooperating in a research project aimed at producing a teaching aid which will help prepare nurse learners to acquire the assessment skills necessary for effective patient care planning. This part of the survey is concentrating on assessment of risk of developing pressure sores.

What we would like now is detailed information on how the pressure sore risk factors identified by your collegues apply to a large sample of typical patients in your hospital.

Thank you for agreeing to help.

INSTRUCTIONS FOR COMPLETING THE QUESTIONNAIRE

1. There are 10 identical patient assessment forms inside this booklet - one for each patient we would like you to assess with regards to pressure sore risk factors.

2. Select the patients as follows.............
   - if you have a patient (or patients) with a pressure sore then assess him/her/them first of all.
   - other patients for inclusion should be simply every second name in your Kardex or bedstate list.

3. Try to fully complete each assessment - always select the description which BEST FITS the particular patient you are assessing, even if a description may not be exact.

Patient's initials..... Sex.....male / female

Age..... Diagnosis........................

For each of the following Activities of Living (AL s), circle numbers beside descriptions which BEST FIT this patient.

AL		POSSIBLE NURSING ASSESSMENTS
MOBILISING - present functioning	1.	fully ambulant, restrictions few if any
	2.	bed or chairfast with short assisted walks only
	3.	bed or chairfast
	4.	bedfast with free movement in bed
	5.	bedfast and virtually immobile in bed
- dependency	1.	patient lifts and turns self, no nursing input required
	2.	1 nurse required to assist only
	3.	1 - 2 nurses required, patient contributes some effort
	4.	2 nurses required, patient can contribute little or no effort
EATING AND DRINKING		relative to his/her height, this patient is...
- build	1.	significantly underweight, diet supplement could be appropriate
	2.	slightly underweight
	3.	within average limits of build
	4.	slightly overweight
	5.	significantly overweight; reduction diet could be appropriate
- nutritional state	1.	current state of nutrition seems adequate
	2.	signs of protein and/or vitamin deficiency
BREATHING		patient is in hypotensive or shock state....yes / no
		shows signs of poor peripheral circulation..yes / no

P.T.O.

APPENDIX 3 Instructions to subjects undertaking Process Tracing Experiment

(screen 1)

Welcome to the experiment, would you please type in your first name (and press Return):-

(screen 2)

As you will know the experiment concerns the risk of patients developing pressure sores. Firstly, though some elementary tasks with the 'mouse' will familiarise you with its use. You will see that moving the 'mouse' moves the the arrow on the screen. When deciding about how at-risk each patient is and when assessing each patient, you will need to move the wee arrow into a button' and click the 'mouse'.

Now try it on this button when you are ready ....

(button reads "next screen please")

(screen 3)

I'm sorry if this seems trivial, it's to ensure that you know what to do. Could you please click this much smaller button

□

(screen 3)

Now click this one to receive instructions about your task

(button reads "instructions button")

(screen 4)

Basically you are to gather information about some patients, then you will be asked to assess them. The 'patients' come in two different forms ....

Type 1 - you will see the names of the factors about a patient. If you want to know more about any factor in order to assess that patient, then click the wee button beside that factor.

Type 2 - you will see on the screen all the details about that factor. You dont need to click any buttons beside factors.

(button reads "next screen")

(screen 5)

All we ask you to do is to 'think aloud' while assessing each patient's risk of developing a pressure sore, and when you feel you have enough information to decide whether a patient is high, medium or low risk, then simply click the appropriate button at the bottom of the screen

(button reads "next screen")

(screen 6)

How many buttons do you click for each patient? You should gather only as much information as you need in order to make an at-risk decision. Request information as you require it to make your decision for a patient.

It may well be that you feel that in any particular trial, not all of the information potentially available is necessary for you to make your decision. In this situation you may leave some factor names covered. Conversely, you should not attempt to unduly restrict the number of factors on which you base your decision.

(button reads "next screen")

(screen 7)

Remember, there are no 'correct' answers. You are the expert, and we are trying to find out about this expertise. We're interested, primarily, in how you as a skilled practitioner go about making decisions on pressure sore risk, and only secondly in the decision you come to.

Now click this button to get some practice at the task. Please remember to 'think aloud' throughout each trial. You may ask any questions now, or again after the practice trials.

(button reads "practice trials")

APPENDIX 4 Instrument for collection of patients comprising database2

Ward ....

Patient's initials ....

SEX
☐ male
☐ female

AGE
☐ 49 or below
☐ 50 - 69
☐ 70 - 89
☐ 90 or over

primary diagnosis ....

MOBILITY
☐ bedfast & virtually immobile in bed
☐ bedfast with free movement in bed
☐ bed or chairfast
☐ bed or chairfast with short assisted walks only
☐ fully ambulant. Restrictions few if any

PERIPHERAL CIRCULATION
☐ ok
☐ poor

MENTAL STATE
☐ heavily sedated or unconscious
☐ marked disorientation with restlessness
☐ mild disorientation or semi-conscious
☐ alert and orientated

FAECAL CONTINENCE
☐ patient has full bowel control
☐ occasional faecal incontinence
☐ patient suffers diarrhoea

NUTRITIONAL STATE
☐ evidence of protein and/or vitamin deficiency
☐ current state of nutrition seems adequate

BLOOD PRESSURE
☐ patient is hypotensive
☐ not hypotensive

BUILD
☐ significantly overweight
☐ slightly overweight
☐ within average limits of build
☐ slightly underweight
☐ significantly underweight

SKIN TYPE
☐ shiny and transparent areas - 'tissue paper'
☐ fine & delicate - 'papery'
☐ rather dry & thin
☐ normal & healthy for age

URINARY CONTINENCE
☐ frequently incontinent / bypassing catheter
☐ occasional incontinence (eg at night)
☐ catheterised/urodome - not bypassing
☐ continent with nurses' help (eg bottle or commode)
☐ fully continent & self-caring

LIFTING & TURNING
☐ patient lifts and turns self - no nursing input required
☐ 1 nurse required to assist only
☐ 1 - 2 nurses required - patient contributes some effort
☐ 2 nurses required - patient can contribute little or no effort

PTO

Based on the description of the patient which you have just completed, how at risk is this patient currently from developing pressure sores (or further sores) ? So under this definition someone could be high risk even although they are receiving excellent preventative care.

LOW RISK ☐ MEDIUM RISK ☐ HIGH RISK ☐

Below are some categories of nursing care which aim to reduce risk of pressure sore development. Could you please choose the most appropriate for this patient.

POSITIONAL PRESSURE RELIEF (whether carried out by nurse or patient)
☐ position should be changed 2 hourly or less.
☐ position should be changed about 4 hourly
☐ need for positional relief not really applicable

'MOISTURE MANAGEMENT'
☐ carry out catheter / urodome care
☐ toilet 2 - 4 hourly
☐ ensure bottle / bedpan / commode supplied as necessary
☐ need for moisture management not really applicable

PRESSURE RELIEF AIDS
☐ 'full' aids required (eg Spenco mattress)
☐ some aids required
☐ need for pressure relief aids not really applicable

LIMB EXERCISES
☐ passive exercises appropriate
☐ encourage patient to exercise limbs
☐ need for limb exercises not really applicable

CONDITIONS REQUIRING ALLEVIATION (choose more than one if applicable)
☐ skin condition
☐ high protein diet appropriate
☐ high calorie diet appropriate

APPENDIX 5 Information selection routes to decision point for cognitive model

note 1 - after the first route, each numbered route begins with an attribute value. To find the full statement of route x, add all attribute values beginning 'leftwards' of route x. For example, route 2 should fully read:

2 MOBILITY - bedfast & virtually immobile in bed
MENTAL STATE - heavily sedated or unconscious (catheter assumed)
BUILD - within average limits / slightly underweight
NUTRITIONAL STATE -seems adequate :

note 2 - decision points are represented by colon :

note 3 - where applicable, assumtions made by the model about unseen attribute values are placed in brackets.

note 4 - the model considers only 'legal' values. This means that, for example, no route can operate for patients who are both 'unconscious' and 'fully continent of urine'.

1 MOBILITY - bedfast & virtually immobile in bed
MENTAL STATE - heavily sedated or unconscious (catheter assumed)
BUILD - significantly overweight / slightly overweight (nutrition ok assumed):

2 BUILD - within average limits / slightly underweight :

3 BUILD - significantly underweight (poor nutrition assumed)
CIRCULATION - ok :

4 CIRCULATION - poor (poor skin assumed) :

5 MOBILITY - bedfast & virtually immobile in bed
MENTAL STATE - marked disorientation with restlessness
SKINTYPE - all values
URINARY CONTINENCE - frequently / occasional incontinent :

6 URINARY CONTINENCE - continent with nurses' help :

7 URINARY CONTINENCE - catheterised
FAECAL CONTINENCE - all values :

8 MOBILITY - bedfast & virtually immobile in bed
MENTAL STATE - mild disorientation / alert and orientated
URINARY CONTINENCE - frequently / occasional incontinent
SKINTYPE - tissue / papery / dry & thin
CIRCULATION - ok :

9 CIRCULATION - poor (poor skin assumed) :

10 MOBILITY - bedfast & virtually immobile in bed
MENTAL STATE - mild disorientation / alert and orientated
URINARY CONTINENCE - frequently / occasional incontinent
SKINTYPE - normal & healthy for age :

APPENDIX 5 (cont.)

11 MOBILITY – bedfast & virtually immobile in bed
MENTAL STATE – mild disorientation / alert and orientated
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – full control of bowels
NUTRITIONAL STATE – evidence protein or vitamin deficiency
(poor build assumed)
SKINTYPE – all values :

12 MOBILITY – bedfast & virtually immobile in bed
MENTAL STATE – mild disorientation / alert and orientated
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – full control of bowels
NUTRITIONAL STATE – seems adequate :

13 MOBILITY – bedfast & virtually immobile in bed
MENTAL STATE – mild disorientation / alert and orientated
URINARY CONTINENCE – continent with nurses' help
NUTRITIONAL STATE – seems adequate :

14 NUTRITIONAL STATE – evidence protein or vitamin deficiency
(poor build assumed)
SKINTYPE – all values :

(the 14 'bedfast & immobile' information selection routes represent a total number of 70 possible paths to decision point if each attribute value is taken separately. To take the final route, for example, 'CIRCULATION – poor/ok' is strictly speaking 2 separate paths)

15 MOBILITY – bedfast / bed or chairfast / chair with assisted walks
URINARY CONTINENCE – frequently / occasional incontinent
SKINTYPE – all values other than normal & healthy for age
CIRCULATION – both values
MENTAL STATE – all values :

16 MOBILITY – bedfast / bed or chairfast / chair with assisted walks
URINARY CONTINENCE – frequently / occasional incontinent
SKINTYPE – normal & healthy for age
MENTAL STATE – all values :

17 MOBILITY – bedfast / bed or chairfast / chair with assisted walks
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – patient has full bowel control
MENTAL STATE – marked disorientation with restlessness
SKINTYPE – all values :

18 MOBILITY – bedfast / bed or chairfast / chair with assisted walks
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – patient has full bowel control
MENTAL STATE – mild disorientation / alert & orientated
NUTRITIONAL STATE – seems adequate :

APPENDIX 5 (cont.)

19 MOBILITY – bedfast / bed or chairfast / chair with assisted walks
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – patient has full bowel control
MENTAL STATE – mild disorientation / alert & orientated
NUTRITIONAL STATE – evidence protein or vitamin deficiency (poor build assumed)
SKINTYPE – all values :

20 MOBILITY – bedfast / bed or chairfast / chair with assisted walks
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – occasional incontinence / diarrhoea
SKINTYPE – all values
MENTAL STATE – all values :

21 MOBILITY – bedfast / bed or chairfast / chair with assisted walks
URINARY CONTINENCE – continent with nurses' help
MENTAL STATE – marked disorientation with restlessness
SKINTYPE – all values :

22 MOBILITY – bedfast / bed or chairfast / chair with assisted walks
URINARY CONTINENCE – continent with nurses' help
MENTAL STATE – mild disorientation / alert & orientated
NUTRITIONAL STATE – seems adequate :

23 NUTRITIONAL STATE – evidence protein or vitamin deficiency (poor build assumed)
SKINTYPE – all values :

(the 7 routes under MOBILITY – bedfast / bed or chairfast / chair with assisted walks represent some 450 separate paths through the patient assessment. This high number is largely accounted for by the fact that all routes are in effect for 3 values of the first attribute – Mobility.)

24 MOBILITY – fully ambulant. Restrictions few if any
MENTAL STATE – marked disorientation with restlessness
SKINTYPE – all values
URINARY CONTINENCE – all values other than catheterised :

25 MOBILITY – fully ambulant. Restrictions few if any
MENTAL STATE – marked disorientation with restlessness
SKINTYPE – all values
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – all values :

26 MOBILITY – fully ambulant. Restrictions few if any
MENTAL STATE – mild disorientation / alert & orientated
URINARY CONTINENCE – frequently / occasionally incontinent
SKINTYPE – normal & healthy for age :

27 MOBILITY – fully ambulant. Restrictions few if any
MENTAL STATE – mild disorientation / alert & orientated
URINARY CONTINENCE – frequently / occasionally incontinent
SKINTYPE – poor
CIRCULATION – poor / ok :

28 MOBILITY – fully ambulant. Restrictions few if any
MENTAL STATE – mild disorientation / alert & orientated
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – patient has full bowel control
NUTRITIONAL STATE –seems adequate :

29 NUTRITIONAL STATE – evidence protein or vitamin deficiency
(poor build assumed)
SKINTYPE – all values :

30 MOBILITY – fully ambulant. Restrictions few if any
MENTAL STATE – mild disorientation / alert & orientated
URINARY CONTINENCE – catheterised
FAECAL CONTINENCE – occasional incontinence / diarrhoea
SKINTYPE – all values :

31 MOBILITY – fully ambulant. Restrictions few if any
MENTAL STATE – mild disorientation / alert & orientated
URINARY CONTINENCE – continent with nurses' help / fully continent
NUTRITIONAL STATE –seems adequate :

32 NUTRITIONAL STATE – evidence protein or vitamin deficiency
(poor build assumed)
SKINTYPE – all values :

(92 separate paths for this value of Mobility. In total, the number of separate paths which are possible when assessing a patient = 612)

APPENDIX 6 Booklet completed by subjects undertaking Experiments 1 and 2 (Ch. 4)

Thank you for agreeing to help with this research into the nursing process.

Your task will take about 10 minutes.

So that your time is not wasted and the session is a useful one you must not turn over any of the pages in this booklet until you are asked to.

INSTRUCTIONS

Below is a brief description of a patient who is at risk from developing pressure sores. The description consists of a few details about the patient followed by some points from the plan of care being delivered by the nurses caring for her.

Would you read this description carefully and try to build up a good understanding (or 'picture') of this patient in your mind.

Please begin reading now and dont turn to the next page until you are directed. If you have spare time, use it to re-read the description and plan of care.

A Patient at risk of developing pressure sores

Mrs Ritchie (aged 73) has a protein deficient nutritional state. She is mildly disorientated and bedfast and immobile in bed. Her circulation is good, blood pressure is normal, and she has good control of her bowels. However she is occasionally incontinent of urine.

The nurses caring for Mrs Ritchie, who is lying on a sheepskin to relief pressure, are ensuring that roughly every two hours her position is changed and she is encouraged to use a bedpan.

Remember, don't turn over until you are asked to - use any spare time to re-read the description and plan of care.

(From this point on please don't turn back to read the description again)

In the list below, each of the facts from the description of the patient on the previous page is arranged with one other fact, so that all possible 'pairs' appear at some point. There are in fact 28 pairs in the list.

What we want you to do (when you are told) is to indicate how "grouped" or "linked" are the two items of each pair **within your own memory**.

Now this task is surprisingly DIFFICULT. Perhaps you could now read the main reasons for this before you go on ....

WHAT WE **DO** WANT TO KNOW - an assessment from you of how closely in your mind the two items are linked after having just read about this patient.

WHAT WE **DONT** WANT TO KNOW -
1. how you think they should be grouped (ie what the books say or whatever), or..
2. how you think they could be grouped.(ie they seem like things that go together in the world)

So remember, what we want is how closely the items of each pair are **currently** grouped or linked in your own memory.

After reading each pair of facts, indicate with an "X" on an appropriate point of the adjacent line how 'grouped in your mind' you think this pair currently is. The list continues over the next two pages. Please begin:

EXAMPLES FROM THE LIST OVERLEAF

	not grouped at all	strongly grouped
aged 73 / good circulation	X------------------------------------	|
mildly disorientated / bedfast and immobile in bed	|------------------------------------	|

APPENDIX 6 (cont.)

not grouped at all ——— strongly grouped

aged 73 /
good circulation |-----------------|

mildly disorientated /
bedfast and immobile in bed |-----------------|

good control of bowels /
aged 73 |-----------------|

good circulation /
occasionally incontinent |-----------------|

mildly disorientated /
normal blood pressure |-----------------|

bedfast and immobile in bed /
protein deficient nutritional state |-----------------|

normal blood pressure /
protein deficient nutritional state |-----------------|

good control of bowels /
normal blood pressure |-----------------|

aged 73 /
normal blood pressure |-----------------|

protein deficient nutritional state /
mildly disorientated |-----------------|

good circulation /
bedfast and immobile in bed |-----------------|

occasionally incontinent /
good control of bowels |-----------------|

protein deficient nutritional state /
aged 73 |-----------------|

normal blood pressure /
good circulation |-----------------|

protein deficient nutritional state /
good control of bowels |-----------------|

APPENDIX 6 (cont.)

	not grouped at all ——— strongly grouped
mildly disorientated / occasionally incontinent	|-----------------------------------|
bedfast and immobile in bed / good control of bowels	|-----------------------------------|
aged 73 / occasionally incontinent	|-----------------------------------|
good circulation / protein deficient nutritional state	|-----------------------------------|
occasionally incontinent / protein deficient nutritional state	|-----------------------------------|
mildly disorientated / good control of bowels	|-----------------------------------|
aged 73 / mildly disorientated	|-----------------------------------|
bedfast and immobile in bed / normal blood pressure	|-----------------------------------|
good circulation / good control of bowels	|-----------------------------------|
normal blood pressure / occasionally incontinent	|-----------------------------------|
bedfast and immobile in bed / aged 73	|-----------------------------------|
mildly disorientated / good circulation	|-----------------------------------|
occasionally incontinent / bedfast and immobile in bed	|-----------------------------------|

THANK YOU – we realise this task is far from easy.

Could you just wait a few moments (until we say) before turning the page.

It would be helpful for us to know one or two things about you with respect to this task.

1. Put an *x* in the box which best describes your experience with a patient of this type....

- ☐ no experience whatsoever
- ☐ a little experience (ie up to several weeks)
- ☐ some experience (ie up to several months)
- ☐ considerable experience (more than a year).

2. (This question will not apply to some of you). In your current (or very recent) ward, would you have come across this type of patient? Disregard the sex of the patient and choose one box to put an *x* in...

☐ not really ☐ sometimes ☐ routinely

1

3. (This question will not apply to some of you). Do you have any nursing certificates? Put an *x* in any appropriate boxes...

☐ Registration ☐ Enrollment ☐ Post-basic course(s)

PLEASE DO NOT TURN OVER THE PAGE UNTIL WE ASK YOU TO.

APPENDIX 6 (cont.)

Now could you try to recall all you can about this patient.

We would like you to write facts about the patient and the nursing care given - write down the things **in the order they come back to you**.

Put each item of information you remember on a separate line. Take your time, it's not an exam and there's no rush.

No keeking at any of the previous pages!

APPENDIX 7 ANOVA summary table - Self-rating of item grouping (Experiment 1)

Source of Variation	df	Sum of Squares	Mean Square	F	p
E	2	10.066	5.033	.257	.7746
Error	48	940.831	19.601		
O	1	155.314	155.314	34.546	.0000
EO	2	45.885	22.942	5.103	.0098
Error	48	215.801	4.496		
T	1	97.456	97.456	25.781	.0000
ET	2	1.596	.798	.211	.8105
Error	48	181.449	3.780		
OT	1	48.044	48.044	23.390	.0000
EOT	2	151.110	75.555	36.783	.0000
Error	48	98.596	2.054		
R	1	173.255	173.255	54.604	.0000
ER	2	55.444	27.722	8.737	.0006
Error	48	152.301	3.173		
OR	1	192.176	192.176	75.424	.0000
EOR	2	28.772	14.386	5.646	.0063
Error	48	122.301	2.548		
TR	1	370.711	370.711	89.388	.0000
ETR	2	100.973	50.487	12.174	.0001
Error	48	199.066	4.147		
OTR	1	294.240	294.240	102.252	.0000
EOTR	2	2.135	1.067	.371	.6921
Error	48	138.125	2.878		
C	1	110.294	110.294	17.545	.0001
EC	2	4.463	2.232	.355	.7030
Error	48	301.743	6.286		
OC	1	169.588	169.588	31.368	.0000
EOC	2	72.154	36.077	6.673	.0028
Error	48	259.507	5.406		
TC	1	261.574	261.574	72.944	.0000
ETC	2	16.551	8.276	2.308	.1104
Error	48	172.125	3.586		
OTC	1	190.240	190.240	87.797	.0000
EOTC	2	35.252	17.626	8.135	.0009
Error	48	104.007	2.167		
RC	1	1.588	1.588	.386	.5374
ERC	2	135.066	67.533	16.405	.0000
Error	48	197.596	4.117		
ORC	1	176.961	176.961	35.086	.0000
EORC	2	24.444	12.222	2.423	.0994
Error	48	242.096	5.044		
TRC	1	44.240	44.240	10.229	.0024
ETRC	2	61.664	30.832	7.129	.0019
Error	48	207.596	4.325		
OTRC	1	1.770	1.770	.328	.5694
EOTRC	2	136.620	68.310	12.667	.0000
Error	48	258.860	5.393		

REFERENCES

Abraham, I.L. (1988) Issues in the application of artificially intelligent decision support technology to nursing diagnosis. in K.J. Hannah, M. Reimer, W.C. Mills, and S. Letourneau (eds) Clinical Judgement and Decision Making: The Future with Nursing Diagnosis. Proceedings of the International Nursing Conference, 1987, Calgary. New York: Wiley & Sons

Adams, J.B. (1976) A probability model of medical reasoning and the MYCIN model. Mathematical Biosciences 32: 177-186

Adelson, B. (1981) Problem solving and the development of abstract categories in programming language. Memory and Cognition 9(4): 422-433

Albert, D.A. (1978) Decision theory in medicine. Health and Society 56,3: 362-401

Altschul, A.T. (1978) A systems approach to the nursing process. J. Advanced Nursing 3: 333-340

Anderson, J.R. (1983) The architecture of cognition. Cambridge, Mass.: Harvard

Anderson, J.R., Boyle, D.F. and Reiser, B.J. (1985) Intelligent Tutoring Systems. Science 228: 456-462

Anderson, J.R. and Reiser, B.J. (1985) The LISP Tutor. Byte 10 (4)

Armstrong, S.L., Gleitman, L.R. and Glietman, H. (1983) What some concepts might not be. Cognition 13: 263-308

Aspinall, M.J. (1979) Use of a decision tree to improve accuracy of diagnosis. Nursing Research 28(3): 182-185

Attwood, M.E. and Polson, P.G. (1976) A process model for water jug problems. Cognitive Psychology 8: 191-216

Ausubel, D., Novak, J. and Hanesian, H. (1978) Educational Psychology: A Cognitive View. 2nd Edition. New York: Holt, Rinehart and Winston

Ball, M.J. and Hannah, K.J. (1984) Using Computers in Nursing. Reston, Virginia: Reston Publishing Co.

Bandura, A. (1977) Social Learning Theory. New Jersey: Prentice-Hall

Barbenel, J.C., Jordan, M.M., Nicol, S.M. and Clark, M.O. (1977) Incidence of pressure-sores in the Greater Glasgow Health Board. Lancet 2: 548-550

Barratt, E. (1987) Putting risk calculators in their place. Nursing Times February 18: 65-70

Barrow, H.S. and Tamblyn, R.N. (1976) Guide to the development ofof skills in problem solving and (clinical) diagnostic reasoning. Monograph No. 1. Hamilton, Ontario, Canada: McMaster University

Barton, A. and Barton, M. (1981) The management and prevention of pressure sores. London: Faber & Faber

Baumann, A. and Bourbonnais, F. (1982) Nursing decision making in critical care areas. J. Advanced Nursing 7: 435-446

Bendall, E. (1975) So you Passed Nurse? London: Royal College of Nursing

Benner, P. (1984) From Novice to Expert: Power and excellence in Nursing Practice. Palo Alto: Addison-Wesley

Bitzer, M.D. and Boudreaux, M.C. (1969) Using a Computer to Teach Nursing. Nursing Forum 8(3): 234-254

Bitzer, M.D., Boudreaux, M.C. and Avner, R.A. (1973) Computer-Based Instruction of Basic Nursing Utilising Enquiry Approach. Computer-Based Education Research Laboratory, University of Illinois

Block, D. (1975) Evaluation of nursing care in terms of Process and Outcome. Nursing Research 24(4): 256-263

Brachman, R.J. and Levesque, H.J. (1985) Readings in Knowledge Representation. Los Angeles: Morgan Kaufman

Braden, B.J. and Bergstrom, M. (1987) A conceptual schema for the study of the etiology of pressure sores. Rehabilitation Nursing 12(1): 8-12

Bratt, E. and Vockell, E. (1986) Using Computers to Teach Basic Facts in the Nursing Curriculum. J. Nursing Education 25(6): 247-251

Broderick, M.E. and Ammentorp, W. (1979) Information structures: An analysis of nursing performance. Nursing Research 28(2):106-110

Brown, J.S. Tanner, C.A. and Padrick, K.P. (1984) Nursing's search for scientific knowledge. Nursing Research 33(1), 28-32

Bruner, J. (1960) The Process of Education. Cambridge, Mass.: Harvard University Press

Bruner, J.S., Goodnow, J.J., and Austin, J.G. (1956) A Study of Thinking. Wiley

Buchanan, B.G., Barstow, D.R., Bechtel, R., Bennet, W., Clancey, W., Kulikowski, C., Mitchell, T. and Waterman, D.A. (1983) Constructing an expert system. in F. Hayes-Roth, D.A. Waterman and D.B. Lenat (eds) Building expert systems Addison-Wesley

Buchanan, B.G. and Shortliffe, E.H. (1984) Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. New York: Addison-Wesley

Burton, R. (1982) Diagnosing Bugs in a Simple Procedural Skill. in D. Sleeman & R. Brown (eds), Intelligent Tutoring Systems. New York: Academic Press

Cantor, N. and Mischel, W. (1977) Traits as prototypes: effects on recognition memory. J. Personality and Social Psychology 35: 38-48

Carbonell, J.R. (1970) AI in CAI: an Artificial Intelligence Approach to Computer-Assisted Instruction. IEEE Transactions on Man-Machine Systems 11: 190-202

Carnevali, D.L. (1983) Nursing Care Planning: Diagnosis and Management. (3rd edition) Philadelphia: J.B. Lippincott

Carr, B. and Goldstein, I. (1977) Overlays: A theory of modelling for computer aided instruction. MIT AI Memo 406

Chase, W.G. and Simon, H.A. (1973) Perception in chess. Cognitive Psychology 4: 55-81

Chi, M.T.H., Feltovich, P.J. and Glaser, R. (1981) Categorisation and representation of physics problems by experts and novices. Cognitive Science 5: 121-152

Clancey, W.J. (1979) Tutuoring Rules for Guiding a Case Method Dialogue. Int. J. Man-Machine Studies 11: 25-49

Clancey, W.J. (1983) The epistemology of a rule-based expert system: A framework for explanation. Artificial Intelligence 20: 215-251

Cohen, G. (1986) Everyday memory. in Memory: A cognitive approach. G. Cohen, M.W. Eysenck and M.E. Le Voi. Milton Keynes: Open University Press

Cohen, J. (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20: 37-46

Cohen, P.R. and Gruber, T.R. (1984) Reasoning about uncertainty: A knowledge representation perspective. in Expert Systems. State of the Art Report 12.7 Maidenhead: Pergammon Infotech

Collart, M.E. (1973) Computer- Assisted Instruction and the Teaching-Learning Process. Nursing Outlook 21(8): 527-532

Collins, A. and Quillian, M.R. (1969) Retrieval Time from Semantic Memory. J. of Verbal Learning and Verbal Behaviour 8: 240-247

Conklin, D. (1981) A Study of Computer-Assisted Instruction in Nursing Education. Unpublished Masters Thesis, University of Calgary

Corcoran, S.A. (1986) Task complexity and nursing expertise as factors in decision making. Nursing Research 35(2): 107-112

Corso, J.F. (1967) The experimental psychology of sensory behaviour. New York: Holt, Rinehart and Winston

Cousins, D. (1986) Computer Aided Learning, Computer Literacy and the Nurse Tutor. Proceedings of First National Conference on the Use of Computers in Health Care Education and Training, September 1986. Bristol: NHS Training Authority

Cronbach, L.J. (1980) Toward reform of program evaluation: aims, methods, and institutional arrangements. London: Josel-Bass

Dawes, R.M. (1979) The robust beauty of improper linear models in decision making. American Psychologist 34(7): 571-582

David, J., Chapman, J. and Lockett, B. (1983) Pressure Sores. Report of the Northwick Park Nursing Practice Research Unit, Harrow, Middlesex

Davis, R. (1982) Expert systems: Where are we? and where do we go from here? AI Magazine 3(2)

de Groot, A.D. (1965) Thought and choice in chess. The Hague: Mouton

Diem, R.A. (1982) Education and Computer Technology : Some Unresolved Issues. Educational Technology June 1976: 19-21

Donabedian, A. (1976) Some basic issues in evaluating the quality of health care. In Issues in Evaluation Research pp. 3-28, Am. Nursing Association

Duda, R.O., Hart, P.E., Nilsson, N.J. and Sutherland, G. (1978) Semantic network representations in rule-based inference systems. In D.A. Waterman and F. Hayes-Roth Pattern directed inference systems. New York: Academic Press

Edgell, S.E. (1978) Configural information processing in two-cue probability learning. Organisational Behaviour and Human Performance 22: 404-416

Einhorn, H.J. and Hogarth, R.M. (1981) .Behavioural decision theory: Processes of judgement and choice. Annual Review Of Psychology 32: 53-88

Ellis, H. (1965) The Transfer of Learning. New York: Macmillan

Elstein, A.S., Shulman, L.S. and Sprafka, S.A. (1978) Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard University Press

Elstein, A.S., Holmes, M., Ravitch, M.M., Rovner, D.R., Holzman, G.B., and Rothert, M.L. (1983) Medical decisions in perspective: Applied research in cognitive psychology. Perspectives in Biology and Medicine 26(3): 486-501

Ericsson, K.A. and Simon, H.A. (1984) Protocol Analysis. Verbal reports as data. MIT Press

Evans, S. (1983) Nursing applications of an expert system. MEDINFO-83. ed. van Bemmel. North-Holland

Feltovich, P.J., Johnson, P.E., Moller, J.H. and Swanson, D.B. (1984) LCS: The role and development of medical diagnositc expertise. In W.J. Clancey and E.H. Shortliffe (eds) Readings in Medical Artificial Intelligence. Reading, MA.: Addison-Wesley

Fink, P.K., Lusth, J.C. and Duran, J.W. (1986) A general expert system design for diagnostic problem solving. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol PAMI-7(7)

Fleiss, J.L. (1971) Measuring nominal scale agreement among many raters. Psychological Bulletin 76: 378-382

Fox, J. (1980) Making decisions under the influence of memory. Psychological Review 87(2): 190-211

Fox, J. (1987) Making decisions under the influence of knowledge. in J. Morris (ed) Modelling Cognition. London: Wiley & Sons

Fox, M.E. (1984) Expert Systems for Education and Training. in Expert Systems. State of the Art Report 12.7 Maidenhead: Pergammon Infotech

Friedman, P.E. (1981) Acquisition of procedural knowledge from domain experts. in IJCAI-81 Proceedings, Vancouver: 856-861

Gagne, R. (1962) The Acquisition of Knowledge. Psychological Review 69: 355-365

Gagne, R. (1977) The Conditions of Learning. 3rd. Edition. New York: Holt, Rinehart and Winston

Gammack, J.G. and Young, R.M. (1984) Psychological techniques for eliciting expert knowledge. in M.A. Bramer (ed) Research and Development in Expert Systems. Cambridge: Cambridge University Press

Gaschnig, J., Klarh, P., Pole, H., Shortliffe, E., and Terry, A. (1983) Evaluation of expert systems: Issues and cas studies. in Building Expert Systems ed. F. Hayes-Roth, D.A. Waterman, and D.B. Lenat, New York: Addison-Wesley

Gibson, E. (1969) Principles of perceptual learning and development Appleton-Century-Crofts

Glaser, R. (1976) The Processes of Intelligence and Education. in L.B. Resnick (ed) The Nature of Intelligence. Hillsdale NJ: Erlbaum

Goldstone, L.A. and Goldstone, J. (1982) The Norton score: An early warning of pressure sores? J. Advanced Nursing 7: 419-426

Goodman, G. (1980) Picture memory: How the action schema affects retention. Cognitive Psychology 12: 473-495

Gordon, D. (1984) Identifying the use and misuse of formal models in nursing practice. in P. Benner From Novice to Expert: Power and excellence in Nursing Practice Palo Alto: Addison-Wesley

Gordon, M. (1973) Information processing strategies in nursing diagnosis. in Ninth Nursing Research Conference, San Antonio. Kansas City: Am. Nurses' Association

Gordon, M. (1980) Predictive strategies in diagnostic tasks. Nursing Research 29(1): 39-45

Gordon, M. (1987) Nursing Diagnosis. Process and Application (2nd edition) New York: McGraw-Hill

Gotts, N.M. (1984) Knowledge Acquisition for Medical Expert Systems - A Review. Artificial Intelligence in Medicine Group University of Sussex. Publication 5

Gould, D. (1986) Pressure prevention and treatment: an example of nurses' failure to implement research findings. J. Advanced Nursing 11: 389-394

Grier, M.R. (1976) Decision making about patient care. Nursing Research 25(2): 105-110

Grier, M.R. (1981) The need for data in making nursing decisions. Ch. 2 in H.H. Werley and M.R. Grier (eds) Information Systems for Nursing. New York: Springer

Grobe, S.J. (1983) Protocols for Software Selection, Development and Evaluation for Nursing Education. in The Impact of Computers on Nursing. Proceedings of the IFIA-IMIA Workshop. September, 1982. M. Scholes, Y. Bryant, B. Barber, (eds) Amsterdam: North-Holland

Grobe, S.J. (1984) Conquering Computer Cowardice. J. Nursing Education 23(3): 232-239

Hammond, K.R. (1966) Clinical inference in nursing: A psychologist's viewpoint. Nursing Research 15:27-38

Hammond, K.R. (1980) Human Judgement and Decision Making. New York: Praeger

Hammond, K.R., Kelly K.J., Castellan, N.J., Schneider, R.J. and Vancini, M. (1966). Clinical inference in nursing: Use of information seeking strategies. Nursing Research 15, 330-336

Hannah, K.J. (1983) Computer Assisted Learning in Nursing Education : A Macroscopic Analysis. in The Impact of Computers on Nursing. Proceedings of the IFIA-IMIA Workshop. September, 1982. M. Scholes, Y. Bryant, B. Barber, (eds) Amsterdam: North-Holland

Hart, P.E. (1982) Direction for AI in the eighties. SIGART Newsletter (79)

Hartley, J.R. (1981) An Appraisal of Computer Assisted Learning in the United Kingdom. Ch. 2 in Selected Readings in Computer Based Learning. ed. N.J. Rushby, London: Kogan Page

Hartley, J.R. (1981b) How expert should an expert system be? IJCAI-81 Proceedings, Vancouver

Harris, A.L. and Hampson, S.E. (1980) Processing information within implicit personality theory. B.J. Social and Clinical Psychology 19: 232-242

Hassett, M. (1984) Computers and Nursing Education in the 1980s. Nursing Outlook 32(1): 34-36

Hawkins, C.A. (1978) C.B.L. Why and Where is it Alive and Well ? Computers and Education. Vol 2: 187-196

Hawkins, D. (1986) In R. Johnstone, System building - Breaking the rules. Expert System User 2(2): 23-25

Hayward, J. (ed) (1986) Report of the Nursing Process Evaluation Working Group. Nursing Education Research Unit, Kings College, London: DHSS

Hoffer, E.P., Mathewson, H.O. and Loughrey, A. (1975) Use of Computer-Aided Instruction in Graduate Nursing Education : A Controlled Trial. J. Emergency Nursing 1(2): 27-29

Hoy, D. (1985) An Evaluation of a Computer Assisted Learning Software Development Programme for Nursing Education in Scotland. Unpublished MSc Thesis, University of Edinburgh

Hoy, R. (1983) Nurse Education and Computers : Time for Change. in The Impact of Computers on Nursing. Proceedings of the IFIA-IMIA Workshop. September, 1982. M. Scholes, Y. Bryant, B. Barber, (eds) Amsterdam: North-Holland

House, E.R. (1980) Evaluating with validity. London: Sage Pub.

Huckaby, L.M.D., Anderson, N., Holm, D.M. and Lee, J. (1979) Cognitive, Affective, and Transfer of Transfer of Learning Consequences of Computer-Assisted Instruction. Nursing Research 28(4): 228-233

Hunt, E., Marin, J. and Stone, P.J. (1966) Experiments in induction. New York: Academic Press

Hyslop, A. (1986) Putting AI into CAI. British Computer Society Nursing Specialist Group Newsletter. March 1986.

Hyslop, A. and Jones, B.T. (1986) Realising the potential of expert systems in nursing education by tackling the problem of knowledge acquisition: An illustrated example of a research-based approach. Proceedings of the First National Conference on the Use of Computers in Healthcare Education and Training. NHS Training Authority, Bristol.

Hyslop, A., Jones, B.T., and Ritchie, I. (1987) The Glasgow Expert Systems in Nursing Project: from Process via Structure to Outcome Evaluation. Current Perspectives in Healthcare Computing, 1987.

Hyslop, A. (1987) Are Programs Intelligent? Nursing Times (Computers in Nursing News) February 25th, 1987.

Itano, J.K. (1988) The clinical judgement process in registered nurses and nursing students. in K.J. Hannah, M. Reimer, W.C. Mills, and S. Letourneau (eds) Clinical Judgement and Decision Making: The Future with Nursing Diagnosis. Proceedings of the International Nursing Conference, 1987, Calgary. New York: Wiley & Sons

Jaccard, J. and Sheng, D. (1984) A comparison of six methods for assessing the importance of perceived consequences in behavioural decisions: Applications from attitude research. J. Experimental Social Psychology 20: 1-28

Jackson, P.R. (1983) An easy to use BASIC program for agreement among many raters. B. J. Clinical Psychology 22: 145-146

Jacoby, J. (1977) The emerging behavioural process technology in consumer decision-making research. in W.D. Perreault (ed) Advances in consumer research Vol.4. Chicago: University of Illinois

Johnson-Laird, P.N. (1983) Mental Models Cambridge: Cambridge University Press

Jonassen, D.H. (1985) The Electronic Notebook : Integrating Learning Strategies in Courseware to Raise the Levels of Processing. in Aspects of Educational Technology XV11. eds. B.S. Alloway and G.M. Mills (eds) London: Kogan Page

Jones, B.T. and Hyslop, A. (1986) The potential of expert systems in (nursing) education: Basic concepts and a way forward. Proceedings of the First National Conference on the Use of Computers in Healthcare Education and Training. NHS Training Authority, Bristol.

Jones, B.T. (1987) Claims made about automated rule induction. Internal Technical Report, Dept. Psychology, University of Glasgow

Jones, J. (1986) An investigation of the diagnostic skills of nurses on an acute medical unit relating to the identification of risk of pressure sore development in patients. Nursing Practice 1: 257-267

Kazdin, A.E. (1982) Single-case research designs. New York: Oxford University Press

Kemmis, S., Atkin, R. and Wright, E. (1977) How Do Students Learn? Working Papers on Computer Assisted Learning. Occasional Paper No. 5, Centre for Applied Research in Education, University of East Anglia

Kerr, J.C., Stinson, S.M., Bay, K., Thurston, N.E. and Leatt, P. (1980) Pressure sores: Nurses' knowledge, attitudes and clinical judgement - A preliminary investigation. Faculty of Nursing, University of Alberta, Edmonton

Kim, M.J., McFarland, G.K. and McLane, A.M. (1984) Classification of nursing diagnoses: Proceedings of the fifth national conference. St. Louis: McGraw-Hill

Kirchoff, K.T. and Holzemer, W.L. (1979) Student Learning and a Computer-Assisted Instructional Program. J. Nursing Education 18(3): 22-30

Koch, B. and Rankin, J. (1985) Educational Computing. Nursing Times, Oct 30, 1985: 45-47

Koch, B. and Rankin, J. (1987) Simulation and Evaluation. in Computers and their Application in Nursing. B. Koch, J. Rankin (eds) London: Harper & Row

Kochar, A.K. and McLean, J. (1985) The Design and Development of Computer-Aided Learning Systems for Industrial Applications. in Aspects of Educational Technology XV11. B.S. Alloway and G.M. Mills (eds) London: Kogan Page

Kolodner, J.L. (1982) The role of experience in development of expertise. AAAI-82: 273-277

Kritek, P.B. (1988) Risks and Realities. in K.J. Hannah, M. Reimer, W.C. Mills, and S. Letourneau (eds) Clinical Judgement and Decision Making: The Future with Nursing Diagnosis. Proceedings of the International Nursing Conference, 1987, Calgary. New York: Wiley & Sons

Lachman, J.L., Lachman, R. and Thronesberry, C. (1981) Metamemory through the adult life span. Developmental Psychology 15: 543-51

Lane, G.H., Cronin, K.M. and Peirce, A.G. (1983) Flow charts - clinical decision making in nursing. Philadelphia: Lippincott

Larsen, D.E. (1983) The Use of Computer-Assisted Instruction to Teach Calculation and Regulation of Intravenous Flow Rates to Baccalaureate Nursing Students. Unpublished Doctoral Dissertation, University of Michigan(cited in Hannah 1983)

Leinhardt, G. and Greeno, J.G. (1986) The Cognitive Skill of Teaching. J.Educational Psychology 78(2): 75-95

Lesh, R.A. (1976) The Influence of an Advance Organiser on two Types of Instruction Units on Finite Geometry. J. for Research in Mathematics Education 7: 82-86

Lewis, C.M. and Hammer, J.M. (1986) Significance testing of rules in rule-based models of human problem solving. IEEE Transactions on Systems, Man and Cybernetics, SMC-16(1), 154-158

Liebowitz, J. (1986) Useful approach for evaluating expert systems. Expert Systems 3(2): 86-96

Lichtenstein, S. (1982) Commentary on issues in protocol analysis. in G. Ungson and D. Braunstein (eds) Decision making: An interdisciplinary inquiry. Boston: Kent

Lindsay, P.H. and Norman, D.A. (1977) Human Information Processing. 2nd Edition. New York: Academic Press

Manpower Services Commission (1986) Matching Training to the Requirements of Adult Learners. Report on Phases 1 and 2 by the MSC Industrial Training Research Unit, Cambridge

Marascuilo, L.A. and McSweeny, M. (1977). Non-parametric and distributuion-free methods for the social sciences. Brooks-Cole: Belmont.

Maslow, A. (1971) The Farther Reaches of Human Nature. Harmondsworth: Penguin

Mates, S. and Sidel, V.W. (1981) Quality assesment by process and outcome methods. American J. of Public Health 71: 687-693

Mayer, R.E. (1978) Advance Organisers that Compensate for the Organisation of Text. J. Educational Psychology 70: 880-886

MacDonald, B., Jenkins, D., Kemmis, S. and Tawney, D. (1975) The Programme at Two: An UNCAL Evaluation Report on the National Development Programme in Computer Assisted Learning. chs. 2-5, appendix A. Centre for Applied Research in Education, University of East Anglia

McClelland, J.L., Rumlehart, D.E. and the PDP Research Group (1986) Parallel Distributed Processing, Volume 2: Psychological and Biological Models. MIT Press

McClelland, L. (1978) Equal versus differential weighting for multiattribute decisions: There are no free lunches. Center for Research on Judgement and Policy Report 207. Institute of Behavioural Science, University of Colorado

McLaren, R. (1985) Knowledge acquisition by rule induction. R and D Management 15(2): 159-166

McKeithen, K.B and Reitman, J.S. (1981) Knowledge organisation and skill differences in computer programmers. Cognitive Psychology 13: 307-325

Meadows, L.S. (1977) Nursing Education in Crisis : A Computer Alternative. J. Nursing Education 16(5): 13-21

Meddis, R. (1984) Statistics using ranks. Oxford: Basil Blackwell

Meehl, P.E. (1954) Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis: Uni. of Minnesota Press

Mervis, C.B. and Rosch, E. (1981) Categorisation of Natural Objects. Annual Review of Psychology 32: 89-115

Michie, D. (1984) Automating the synthesis of expert knowledge. ASLIB Proceedings 36,7/8: 337-343

Miller, G.A. (1956) The magical number seven plus or minus two. Psychological Review 63: 81-97

Mills, G.M. (1985) Categories of Educational Microcomputer Programs: Theories of Learning and Implications for Future Research. Aspects of Educational Technology XVIII. B.S. Alloway and G.M. Mills (eds) London: Kogan Page

Mirrin, S. (1983) Making the Most of the Microcomputer in Nursing Education. in The Impact of Computers on Nursing. Proceedings of the IFIA-IMIA Workshop. September, 1982. M. Scholes, Y. Bryant, B. Barber, (eds) Amsterdam: North-Holland

Mowforth, P. (1986) Some applications with inductive expert system shells. TIOP-86-002, Turing Institute, Glasgow

Newell, A. and Simon, H.A. (1972) Human Problem Solving. New York: Prentice-Hall

Nightingale, F. (1894) Training of Nurses. in Quain (ed) Dictionary of Medicine. London: Longman's

Nisbett, R.E. and Wilson, T.D. (1977) Telling more than we know: Verbal reports on mental processes. Psychological Review 84: 231-259

Norman, D.A. and Rumelhart, D.E. (1975) Memory and Knowledge. in Explorations in Cognition D.A. Norman (ed). San Francisco: W.H. Freeman

Norman, S.E. (1983) Computer Assisted Instruction in Nurse Education - An Approach. in The Impact of Computers on Nursing. Proceedings of the IFIA-IMIA Workshop. September, 1982. M. Scholes, Y. Bryant, B. Barber, (eds) Amsterdam: North-Holland

Norman, S.E. (1983 b) Computer-Assisted Learning in Nurse Education in North America. London: Royal College of Nursing Publication

Norman, S.E. and Townsend, E. (1982) Computers in Nurse Education - A Focused Bibliography. Sheffield City Polytechnic

Norton, D., McLaren, R. and Exton-Smith, A.N. (1962) An investigation of geriatric nursing problems in hospital. Edinburgh: Churchill Livingstone

Ohlsson, S. (1986) Some Princilples of Intelligent Tutoring. Instructional Science 14: 293-325

Osborne, S. (1987) Pressure sores: A quality circle investigation. Nursing Times February 18: 73-76

O'Shea, T. and Self, J. (1983) Learning and Teaching with Computers. Harvester Press: Brighton

Osiobe, S.A. (1985) Sources of information in biomedical decision-making. Methods of Information in Medicine 24(4), 225-229

Ozbolt, J.G., Schultz, S., Swain, M.A., and Abraham, I.L. (1985) A proposed expert system for nursing practice. J. of Medical Systems 9,1/2: 57-68

Padrick, K.P (1988) Teaching clinical decision making. in K.J. Hannah, M. Reimer, W.C. Mills, and S. Letourneau (eds) Clinical Judgement and Decision Making: The Future with Nursing Diagnosis. Proceedings of the International Nursing Conference, 1987, Calgary. New York: Wiley & Sons

Patel, V.L. and Groen, G.J. (1986) Knowledge based solution strategies in medical reasoning. Cognitive Science 10: 91-116

Papert, S. (1980) Mindstorms: Children, Computers and Powerful Ideas. New York: Basic Books

Parsonage, J. (1986) Reviewing Quality in Computer Based Training. Proceedings of First National Conference on the Use of Computers in Health Care Education and Training, September 1986. Bristol: NHS Training Authority

Payne, J.W. (1976) Task complexity and contingent processing in decision making: An information search and protocol analysis. Organisational Behaviour and Human Performance 16: 366-387

Payne, J.W., Braunstein, M.L. and Carroll, J.S. (1978) Exploring predecisional behaviour: An alternative to decision research. Organisational Behaviour and Human Performance 22: 17-44

Piaget, J. (1970) The Science of Education and the Psychology of the Child. New York: Orion Press

Pleasance, P. (1984) Computer assisted learning for nurses. Unpublished paper presented at Library Association - Mental Health & Welfare Group Annual Study Conference, July 8th, 1984.

Pople, H.E. (1973) On the mechanisation of abductive logic. Proceedings of the Third International Joint Conference on Artificial Intelligence: 147-152

Pople, H.E. (1977) The formulation of composite hypothesis in diagnostic problem solving: An exercise in synthetic reasoning. Proceedings of the Fifth International Joint Conference on Artificial Intelligence: 1030-1037

Quinn, F.M. (1980) The Principles and Practice of Nurse Education. London: Crook Helm

Quinlan, J.R. (1979) Discovering rules from large collections of examples: A case study. In D. Michie (ed) Expert systems for the micro-electronic age. Edinburgh: Edinburgh University Press

Quinlan, J.R. (1983) Learning efficient classification procedures and their application to chess end games. In R.S. Michalski, J.G. Carbonell and T.M. Mitchell Machine Learning: An artificial intelligence approach. Palo Alto: Tioga Pub.

Ramsey, C.L, Reggia, J.A., Nau, D.S. and Ferrentino, A. (1986) A comparative analysis of methods for expert systems. Int. J. Man-Machine Studies 24: 475-499

Rankin, J., Koch, B., and McGuire, M. (1986) Attitudes of student nurses to computer assisted learning. Unpublished research report, Forth Valley College of Nursing and Midwivery, Falkirk, Scotland

Reiser, A.J., Anderson, J.R. and Farrell, R.G. (1985) Dynamic Student Modelling in an Intelligent Tutot for LISP Programming. Proceedings of the Ninth International Conference on Artificial Intelligence

Richards, B., Al-Basri, L., Campbell, A., Morris, J. and Minshull, J. (1986) Computer Aided Learning in the NHS. Proceedings of First National Conference on the Use of Computers in Health Care Education and Training, September 1986. Bristol: NHS Training Authority

Richer, M.H. (1986) An evaluation of expert system development tools. Expert Systems 3(3): 166-181

Risner, P.B. (1986) Diagnosis: Analysis and synthesis of data. Ch 7 in J.W. Griffith-Kenney and P.J. Christensen (eds) Nursing Process (2nd edition) St.Louis: Mosby

Ritchie, I. (1987) Inside a prototypical rule inducer. Internal Technical Report, Dept. Psychology, University of Glasgow

Roper, N., Logan, W.W. & Tierney, A.J. (1985) The Elements of Nursing (2nd edition). Edinburgh: Churchill Livingston

Rogers, C. (1969) Freedom to Learn. Ohio: Charles Merrill

Rubin, A.D. (1975) Hypothesis formulation and evaluation in medical diagnosis. Technical Report No. AI-TR-316. Cambridge, MA: Artificial Intelligence Laboritory, MIT

Rumlehart, D.E. and Norman, D.A. (1983) Representation of Knowledge. reprinted in Issues in Cognitive Modelling. A.M. Aitkenhead and J.M. Slack (eds) London: Lawrence Erlbaum

Rushby, N.J. (1979) An Introduction to Educational Computing. London: Crook Helm

Russo, J.E. (1978) Eye fixations can save the world: A critical evaluation and a comparison between eye fixations and other information processing methodologies. in J.K. Hunt (ed) Advances in Consumer Research Vol 5. Chicago: University of Illinois

Schenk, R.C. (1980) Language and memory. Cognitive Science 4: 243-284

Schneider, D.J. and Blankmeyer, B.L. (1983) Prototype salience and implicit personality theories. J. Personality and Social Psychology 44(4): 712-722

Schneider, W. and Shiffrin, R.M. (1977) Controlled and automatic human information processing II. Perceptual learning, automatic attending, and a general theory. Psychological Review 84: 127-190

Scholes, M., Bryant, Y. and Barber, B. (eds.) (1983) The Impact of Computers on Nursing. Proceedings of the IFIA-IMIA Workshop. September, 1982. M. Scholes, Y. Bryant, B. Barber, (eds) Amsterdam: North-Holland

Sheehan, J. (1986) Curriculum Models: Product versus Process. J. Advanced Nursing 11: 671-678

Shortliffe, E.H. (1976) Computer-based medical consultations: MYCIN. New York: Elsevier

Shortliffe, E.H., Buchanan, B.G., and Feigenbaum, E.A. (1979) Knowledge engineering for medical decision making: a review of computer based clinical decision aids. in W.J. Clancey and E.H. Shortliffe (eds) Readings in Medical Artificial Intelligence. Reading, Mass.: Addison-Wesley

Simon, H.A. (1978) Rationality as a process and as a product of thought. Am. Economic Review 68: 1-16

Skinner, B.F. (1968) The Technology of Teaching. New York: Appleton-Century-Crofts

Sleeman, D. (1982) Assessing Aspects of Competence in Basic Algebra. in D. Sleeman and R. Brown (eds), Intelligent Tutoring Systems. London: Academic Press

Slovic, P., Fischoff, B. and Lichtenstein, S. (1977) Behavioural decision theory. Annual Review of Psychology 28:1-39

SNB (1986) Introduction to ECG. CAL program written Foresterhill College of Nursing & Midwifery CAL Group and programed by the Scottish Microelectronics Development Programme under the sponsorship of the National Board for Nursing, Midwifery and Health Visiting in Scotland. (cited in Koch & Rankin 1987)

Stainton, M.C. (1988) Teaching clinical decision making. in K.J. Hannah, M. Reimer, W.C. Mills, and S. Letourneau (eds) Clinical Judgement and Decision Making: The Future with Nursing Diagnosis. Proceedings of the International Nursing Conference, 1987, Calgary. New York: Wiley & Sons

Sweeney, M.A. (1983) Learning Needs and Computers. in The Impact of Computers on Nursing. Proceedings of the IFIA-IMIA Workshop. September, 1982. M. Scholes, Y. Bryant, B. Barber, (eds) Amsterdam: North-Holland

Sweeney, M.A., O'Malley, M. and Freeman, E. (1982) Development of a Computer Simulation to Evaluate Performance of Nursing Students. J. Nursing Education 21(9): 28-38

Szolovits, P. and Long, W.J. (1982) The development of clinical expertise in the computer. in P. Szolovits (ed) Artificial Intelligence in Medicine. Boulder, Colorado: Westview Press

Tait, K., Hartley, J.R. and Anderson, R.C. (1973) Feedback Procedures in Computer-Assisted Arithmetic Instruction. British J. Educational Psychology 43: 161-171

Tanner, C.A. (1983) Research on clinical judgement. in W.M. Holzemer (ed) Review of research in nursing education. Thorofare, New Jersey: Slack

Tanner, C.A. (1988) Theoretical perspectives for research on clinical judgement. in K.J. Hannah, M. Reimer, W.C. Mills, and S. Letourneau (eds) Clinical Judgement and Decision Making: The Future with Nursing Diagnosis. Proceedings of the International Nursing Conference, 1987, Calgary. New York: Wiley & Sons

Thiele, J.E., Baldwin, J.H., Hyde, R.S., Sloan, B. and Strandquist, G.A. (1986) An investigation of decision theory: What are the effects of teaching cue recognition? J. Nursing Education 25(8): 319-324

Townsend, I.J. (1980) Rediscovering the Wheel. Feedback No. 3. May/June. NHS Learning Resources Unit, Sheffield

Townsend, I.J. (1983) The Second Coming - Resurrection or Reservation? in The Impact of Computers on Nursing. Proceedings of the IFIA-IMIA Workshop. September, 1982. M. Scholes, Y. Bryant, B. Barber (eds) Amsterdam: North-Holland

Tversky, A. and Kahneman, D. (1974) Judgement under uncertainty: Heuristics and biases. Science 185: 1124-31

Valish, A.U. (1975) The role of computer assisted instruction in continuing education of registered nurses: an experiemental study. J. Continuing Education in Nursing 6 (1): 13-32

Wellbank, M. (1983) A review of knowledge acquisition techniques for expert systems. Martlesham Consultancy Services, British Telecom Research Laboratories.

Westfall, U.E., Tanner, C.A., Putzier, D.J. and Padrick, K.P. (1986) Activating clinical inferences: A component of diagnostic reasoning in nuring. Research in Nursing and Health 9: 269-277

Whiting, J. (1985) The Nature of Computer-Assisted Learning. in Aspects of Educational Technology XVII. B.S. Alloway and G.M. Mills (eds) London: Kogan Page

Williams, A. (1972) A study of factors contributing to skin breakdown. Nursing Research 21(3): 238-243

Wilson, K.J.W. (1975) A study of biological sciences in relation to nursing. Edinburgh: Churchill Livingstone

Yazdani, M. (1986) Intelligent Tutoring Systems: An Overview. Expert Systems 3(3): 154-163

ADDITIONAL REFERENCES

Bhaskar, R. and Simon, H.A. (1977) Problem solving in semantically rich domains: an example of engineering themodynamics. Cognitive Science 1: 193-215

Ferrari, D. (1986) Considerations on the insularity of performance evaluation. IEEE Transactions on Software Engineering, Vol SE-12, (6)

Jones, B.T. and Ritchie, I. (1987) The origins of Automated Rule Induction. Internal Technical Report, Dept. Psychology, University of Glasgow

Kassirer, J.P. and Gorry, G.A. (1978) Clinical Problem Solving: A Behavioural Analysis. Annals of Internal Medicine 89: 245-255

Timpke, J. and Janney, C.P. (1981) Teaching drug doses by computer. Nursing Outlook, June 1981: 376-377

UKCC (1986) UKCC Project 2000 - A New Preparation for Practice. United Kingdom Central Council for Nursing, Midwifery & Health Visiting, London