# Expert Problem Solving in Mammogram Interpretation: A Visual Cognitive Task


Roger Azevedo

Department of Educational and Counselling Psychology
and
Applied Cognitive Science Research Group

McGill University, Montreal


Thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Educational Psychology
Major in Applied Cognitive Science

October 1997

0-612-44353-1

# DEDICATION

**This dissertation is dedicated to the loving memory of**

**my mother,**

**Henriqueta**

## ACKNOWLEDGMENTS

I would like to thank my supervisor, Prof. Susanne Lajoie for her support and guidance during the conceptualization and realization of my dissertation. Her academic excellence has contributed to my intellectual development and growth throughout my doctoral studies. She epitomizes the "perfect" faculty member and supervisor - one who encourages students to excel, provides ample academic opportunities and research experiences, respects and values students, and above all knows when to "model", "coach", "scaffold", "fade", "articulate", and "reflect".

I would like to express great admiration and appreciation to my committee members and several other faculty members who have played instrumental roles during my studies. They include Profs. Robert Bracewell, Alain Breuleux, Jim Greer, Alenoush Saroyan, and Carl Frederiksen.

I would like thank Dr. David Fleiszer for providing continued support. A very special thank you to Dr. Monique Desaulniers whose invaluable contributions made this study a success. She spent countless hours sharing her expertise and "teaching" me about breast disease and mammography, and combing through pages and pages of verbal protocols. Dr. Patrice Bret facilitated the data collection process. Also, a special thanks to the staff radiologists and radiology residents who participated in this study.

I would also like to thank several of my peers for contributing to this study including Helene Fournier for transcribing the protocols, Litsa Papathanasopoulou and Marguerite Roy for their assistance on issues related to research methodology and statistical analyses, Andre Renaud for his computer/technical assistance, Xiaoyan Zhao for the development of the RadTutor, and Lorraine Meilleur for translating the abstract. I would also like to thank Tom Patrick, Svein-Ivar Lillehaug, Gloria Berdugo, Arshad Ahmad and Janet Blatter for numerous stimulating conversations on several cognitive science topics.

A thanks to several granting agencies for their financial support. This research was supported by doctoral fellowships from the (1) Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR), (2) Social Sciences and Humanities Research Council of Canada (SSHRC), (3) Ministère de l'Éducation du Quebéc (MEQ) and Ministère de l'Industrie, Commerce, Science et Technologie (MISCT), and (4) a combined McGill University and Social Sciences Research Sciences and Humanities Research Council of Canada fund. Funding for the development of the RadTutor has been provided by Dr. David M. Fleiszer and McGill University's Medical Informatics Committee.

Finally, the most special thanks goes to my partner Sonia. She has enriched my life with happiness, friendship, kindness, joy, love and above all has given me the strength to pursue my goals. She has been instrumental during various phases of my doctoral studies and in particular during the completion of this study by sharing ideas, editing earlier versions of this document, and spending countless hours discussing issues related to cognitive science, computer-based instruction, medical education and breast disease.

## ABSTRACT

The problem solving strategies used by staff radiologists and radiology residents during the process of interpreting difficult breast diseases depicted on mammograms were examined. Staff radiologists ($n$ = 10) and radiology residents ($n$ = 10) diagnosed 10 cases under two experimental conditions (5 authentic and 5 augmented). In the authentic condition, standard unmarked mammograms were used. Mammographic findings were highlighted on a second set of the same cases for the augmented condition. The experimental sessions were recorded and the verbal protocols were analyzed based on a coding scheme derived from an analysis of mammography interpretation. The results were used to refine this analysis and develop a model of mammography interpretation. Repeated measures ANOVAs revealed (a) that staff radiologists scanned the cases significantly faster than residents with no significant main effect for condition and no interaction, and (b) no differences between groups in reading time across experimental conditions. No group differences were found in the number of radiological findings, radiological observations, and number of diagnoses across experimental conditions. Frequency analyses revealed that both groups regardless of experimental condition (a) used the same types of operators, control processes, diagnostic plans and goals, (b) committed the same number of errors, and (c) committed case-dependent errors. Analyses revealed that mammography interpretation was characterized by a predominant use of data-driven or mixed strategies depending on case typicality and clinical experience. The fact that few differences were found between the groups on the various measures may be due to the fact that mammogram interpretation is a well-constrained task. The theoretical implications of the study include the need for further research for the purposes of building a more detailed, comprehensive model of the perceptual and cognitive processes underlying mammogram interpretation. Finally, the results have been applied to develop a basic conceptual framework for the development of the RadTutor, a computer-based tutor for training radiology residents to interpret mammograms.

# RÉSUMÉ

Les stratégies de résolution de problèmes du personnel de la radiologie ainsi que les résidents de la radiologie ont été évaluées pendant un processus d'interprétation des cas complexes concernant les maladies du sein représenté sur radiographie. Le personnel de la radiologie ($n$ = 10) et les résidents en radiologies ($n$ = 10) ont chacun posé un diagnositic sur 10 cas sous deux conditions expérimentales (5 authentiques et 5 augmenté). La condition authentique comprenait des mammographies standard non identifiées. La condition augmentée démontrait les mêmes mammographies que la condition précédente sauf que les symptômes étaitent mise en évidence. Les sessions expérimentales étaient enregistrées et les protocols verbaux ont été analysés en suivant une méthode d'analyse basée sur un modèle préliminaire d'une interprétation d'une mammographie. Ces résultats ont contribué au perfectionnement d'un modèle préliminaire d'interprétation de mammographies. Les analyses de variance avec mesure répétées ont démontrées (a) le personnel en radiologie visualisaient les cas significamment plus rapidement que les résidents sans effects principaux significatifs pour l'effet condition et sans interactions, et (b) aucune différence entre les groupes au niveau du temps de lecture pour les deux conditions expérimentales. Aucune différence entre les conditions expérimentales en ce qui concernent le nombre de symptômes radiologiques, observations radiologiques, et le nombre de diagnostics dans les conditions expérimentales. Des analyses de fréquence ont démontrées que les deux groupes peut importe la condition expérimentale (a) utilisent les mêmes types d'opérateurs, processus de contrôle, plans diagnostiques, et buts, (b) ont produits le même nombre d'erreurs, et (c) ont produits des erreurs spécifiques aux cas. Les analyses ont mise en évidence que l'interprétation de mammographie est caractérisée par une utilisation prédominante d'une stratégie de raisonnement par chaînement avant ou une stratégie mixe dépendamment du typicalité du cas et de l'expérience clinique de l'individu. Il se peut que le manque de différences entre les groupes selon plusieurs mesures est du au

fait que l'interprétation de mammographies est une tâche très contrainte. Les implications théoriques de cette étude comprennent la nécessité de poursuivre d'autres études dans le but de contruire un modèle encore plus détaillé et compréhensif des processus perceptuels et cognitifs pertinentes a l'interprétation d'une mammographie. De plus, ces résultats ont contribués a l'élaboration d'un cadre conceptuel pour le développement du tuteur RadTutor, un tuteur informatisé pour la formation des résidents en radiologie pour l'interprétation des mammographies.

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF APPENDICES

# CHAPTER 1

# INTRODUCTION

Breast cancer is the leading form of cancer diagnosed in Canadian women (excluding non-melanoma skin cancer), accounting for about 30% of all new cases (Gaudette, Silberberger, Altmayer & Gao, 1996). After the age 30, incidence rates begin to rise, and the highest rates are among women aged 60 and over. Canadian incidence rates have increased slowly and steadily since 1969, rising most rapidly among women aged 50 and over. Canadian incidence rates are among the highest of any country in the world, second only to the United States. Mammographic screening has become an accepted means of substantially reducing breast cancer mortality. Nevertheless, 11% to 25% of cancers are overlooked by radiologists on initial screening mammograms (Goergen, Evans, Cohen & MacMillan, 1997). The high incidence rates together with the rate of misdiagnoses make this an alarming problem which is associated with societal, ethical, and additional medical concerns. Given the scope and seriousness of the problem, it is evident that any promising means for alleviating it should be investigated. Aside from the health and medical sciences, other disciplines such as cognitive science can contribute to this understanding by identifying the cognitive components that constitute proficiency in mammogram interpretation and consequently lead to improved training of future radiologists.

Cognitive science is an interdisciplinary field that is used to build an understanding of “thinking.” A basic assumption is that the mind is a computational system that constructs, manipulates, and represents symbols (Newell & Simon, 1972; Simon, 1979). Contributing disciplines such as cognitive psychology, philosophy, computer science, neuroscience, anthropology, and linguistics provide cognitive science with different ways of investigating the nature of "thinking", including numerous epistemological frameworks, research methodologies, and analytical approaches. Cognitive science research methods for studying thinking range from computer simulations to naturalistic observations. Other methods include: task analysis, recording electrical brain impulses, eye movement recordings, content analysis, meta-analysis, memory for task information, linguistic analyses, accuracy of responses, error analysis, reaction times and protocol analysis (retrospective and concurrent) (for an extensive review refer to Posner, 1989; Ericsson, & Oliver; 1988; Ericsson & Simon, 1993; Ericsson & Smith, 1991b; Olson & Biolsi, 1991). Further, information processing theories, research on expert-novice differences, and the widespread use of cognitive task analysis have contributed to our understanding of learning and instruction in various educational and professional domains such as reading, writing, science, math, history, and medicine.

Cognitive science offers a foundation for the study of mammogram interpretation and application of the results to the improvement of medical training. This study, based on a cognitive science approach, investigates the underlying nature of radiological expertise in both staff radiologists (M.D.s with extensive post-residency training) and radiology residents (M.D.s completing their residency training) by focusing on their problem solving strategies. The results provide a better understanding and a performance model of mammogram interpretation which will lead to identifying better training methods.

Radiological diagnosis is complex, involving several years of acquiring formalized medical knowledge as well as many years of clinical experience. The ability to diagnose accurately necessitates the integration of several bodies of knowledge with separate organizing principles, including physiology, anatomy, pathophysiology, and projective geometry of radiography. Various theoretical frameworks postulate that the attainment of accurate visual diagnostic reasoning abilities involves the interaction between cognitive and perceptual factors. In order to adequately understand the diagnostic process, a more detailed investigation is required.

Numerous researchers, employing disparate theoretical and empirical paradigms, have investigated radiological expertise. Three basic "paradigms" that have been investigated widely are: (a) search studies which investigate eye movement patterns, (b) signal-detection studies which investigate the ability to detect normal and abnormal film findings, and (c) cognitive research aimed at studying the underlying cognitive and perceptual factors involved in radiological expertise. Relatively few studies (Lesgold, Feltovich, Glaser & Wang, 1981; Lesgold, Rubinson, Feltovich, Glaser, Klopfer & Wang, 1988; Rogers, 1992; 1996) have actually investigated the underlying cognitive and perceptual factors. These studies have focused specifically on the interpretation of chest x-rays. As a result, a fundamental understanding of the constitution and acquisition of expertise in other radiological sub-specialties, such as mammography interpretation, has yet to be achieved. This will require the use of appropriate cognitive science methodologies such as protocol analysis and the study of the problem solving strategies of radiology professionals with varying levels of expertise in the interpretation of mammograms.

The present study investigated the problem solving strategies used by staff radiologists and radiology residents during the process of diagnosing difficult breast diseases depicted on mammograms. It was designed to investigate the diagnostic problem solving strategies used by staff radiologists and radiology residents during the interpretation of difficult breast disease cases depicted on mammograms. The results of this study address existing problems associated with radiology residency training and are used as the basis for the design of a computer-based learning environment to train radiology professionals in the interpretation of mammograms.

## Research Objectives

The specific research objectives addressed in this study included:

1) Identify a cognitive model of diagnostic problem solving in mammography interpretation.
2) Identify the problem solving strategies, operators, and control processes used by staff radiologists and radiology residents used during mammography interpretation.
3) Conduct in-depth analyses of protocols from several breast disease cases to exemplify typical staff radiologists' and radiology residents' problem solving strategies.

4) Analyze the frequency and types of errors committed by both groups while diagnosing the breast disease cases, and provide in-depth analyses of protocols from these cases.
5) Study the effects of two experimental conditions (authentic and augmented) on several aspects of the groups' performance (number of mammogram findings, observations and diagnoses, scanning time and reading time, and accuracy).

The following chapter presents a review of the literature and research on problem solving, expertise, medical cognition and radiological diagnosis.

# CHAPTER 2

# REVIEW OF THE LITERATURE

A review of the literature and research is presented in this chapter. The literature that served as the foundation for this study is presented in four sections. It provides overviews of the following literature: (a) problem solving, (b) expertise, (c) overview of medical cognition including medical problem solving and expertise, and (d) research in the area of radiological diagnosis. The rationale and problem statement for the study are introduced in the last section. The purpose of this chapter is to briefly introduce the area of cognitive science and information processing theory as the theoretical foundation, and verbal protocol analysis as the analytical method used in this study of the problem solving processes of staff radiologists and radiology residents diagnosing breast disease cases.

## The Study of Problem Solving

> "Problem solving by recognition, by heuristic search, and by pattern recognition and extrapolation are examples of rational adaptation to complex task environments that take appropriate account of computational limitations - of bounded rationality. They are not optimizing techniques, but methods for arriving at satisfactory solutions with modest amounts of computation" (Simon, 1990, p. 11).

Nearly all human activity can be viewed as problem solving. Historically, problem solving has evolved from the study of knowledge-lean laboratory tasks (e.g., Tower of Hanoi) to semantically-rich, real-world complex tasks (e.g., electronics troubleshooting). This section presents a brief historical overview of the problem solving literature. Gestalt psychologists such as Kohler (1925) and Duncker (1945) conducted the earliest experimental problem solving research. They concentrated on insight problems in which solutions follow rapidly once the appropriate steps have been taken. The challenge in solving these problems was recognizing the relations among the various aspects of a situation. This type of research study was followed by research on knowledge-lean tasks. These tasks require no background knowledge to complete and everything a participant needs to perform the task is included in the instructions. An example of such a task is the Tower of Hanoi. The study of knowledge-lean tasks led to the formulation of Newell and Simon's (1972) book on human problem solving. Their research led to the development of numerous domain specific performance models based on verbal protocols, latencies between steps, and eye movement data. The conception of problem solving was based on Newell and Simon's (1972) seminal research in artificial intelligence and computer simulation of human thought.

In the late 1970s, attention shifted from knowledge-lean to knowledge-rich tasks such as algebra, chess, bridge, computer programming, and medical diagnosis. The focus on knowledge-rich tasks involved research aimed at investigating the differences between novices and experts. Incorporating participants with varying levels of expertise while holding task domain constant allowed researchers to investigate the underlying effects of expertise on problem solving behavior. The typical study gave the same set of problems to experts and novices and used protocol analysis to examine differences in the performance of both groups. Results revealed that experts were able to recognize the type of problem, retrieve a solution from memory, and immediately generate a solution. In contrast, novices may have lacked domain knowledge and as a result searched for a solution. In sum, expertise allows one to substitute recognition for search (VanLehn, 1989).

**Theoretical Constructs in Problem Solving**

The IPT is based on the Physical System Hypothesis (Newell & Simon, 1976) which states that a system will be capable of intelligent behavior if and only if it is a physical system. A physical system is a system capable of inputting, outputting, storing, and modifying symbol structures, and carrying out some of these actions in response to the symbols themselves. Symbols are any kinds of patterns on which these operations can be performed, where some of the patterns denote actions (Newell, 1990). Due to the limitations in the computing speeds and power of physical systems such as the human brain, intelligent systems must use approximate methods to handle most tasks. This led to the principle of bounded rationality (Simon, 1989). According to Simon (1989), this principle states that since humans can seldom solve problems exactly, the optimization strategy suggested by rational analysis is seldom available. Thus, humans must find techniques for solving problems approximately and arrive at different solutions depending on what approximations are reached. Therefore, to describe, predict and explain human behavior of a system of bounded rationality, we must construct a theory of the system's processes and describe the environments to which it is adapting (Simon, 1990). This leads to the computational feasibility within the limits of bounded rationality which is to store in LTM knowledge and strategies that reduce the computational requirements of the tasks.

According to the principle of bounded rationality, the mechanisms used during problem solving (to cope with the real-life complexity) include *recognition processes*, *heuristic search*, and *serial pattern recognition*. Experts make extensive use of *recognition processes* based on stored knowledge to handle everyday tasks. This recognition capability is based on an approximately 50,000 or more stored cues and associated knowledge allowing them to rapidly solve problems (Chi, Glaser & Farr, 1988). Similarly, computer simulations such as EPAM (Feigenbaum & Simon, 1984) provide explanatory mechanisms for recognition-based expertise. *Heuristic search* deals with problems whose solutions are not provided by immediate recognition but which require further analysis. The application of heuristics may be based on the (a) ability to draw upon structural information to guide search to the goal (based on the application of powerful task specific heuristics in a highly structured task domain), and (b) application of weak methods such as satisficing and means-ends analysis when there is practically no task domain structure. *Serial pattern recognition* is the ability to find patterns in sequences of letters and geometric shapes, etc. (Kotovsky & Simon, 1973).

**The Nature of Problem Solving**

Problem solving is high level cognitive processing directed at transforming a problem situation from its current state into a goal state where a solution is not immediately obvious (Mayer, 1989; 1990; 1992). This definition includes elements that need further elaboration. Mayer (1994) has provided an elaboration of important concepts related to problem solving. First, problem solving is *cognitive* since it occurs within the problem solver's information-processing system. Therefore, problem solving can only be inferred from changes in the problem solver's behavior. Secondly, problem solving is *computational*, in that it involves manipulating or performing mental operations on information in the problem solver's memory. Thirdly, problem solving is *directed* since the problem solver engages in problem solving in order to achieve a particular goal. Lastly, problem solving is *personal* since what constitutes a problem for one person may not be problem for another.

Based on information-processing theories (Newell & Simon, 1972; Anderson, 1993b; Newell, 1990), problem solving involves the construction of a problem space and the implementation of a search strategy. A *problem space* is a representation of the given state, goal state, and all intervening states generated by making legal moves as the problem solver transverses this space. The problem space also includes the problem solver's strategic knowledge, which may include methods previously acquired through experience in the domain and general problem solving methods (Greeno & Simon, 1988). The actual reference to a *problem solving state* is ambiguous since it could mean some external state of affairs or some internal coding of that state of affairs. A *problem solving operator* is an action that transforms one state into another. A *search strategy* is a procedure for moving from one state to the next in the problem space. Search can be allocated among two collaborating processes, the backup and proceed strategy (VanLehn, 1989). The *backup strategy* maintains the set of old states and chooses one when necessary. The *proceed strategy* chooses an operator to apply to the current state, applies it, and evaluates the resulting states. Problem solving strategies have been divided into weak and strong methods and are discussed in more detail in the next section.

Problem solving can be analyzed according to three major processes. These include representation processes, solution processes, and control processes. *Representing* occurs when a problem solver creates an internal mental representation of the problem including the initial state, goal state and operations. *Solving* occurs when a problem solver establishes and carries out a plan. *Controlling* occurs when a problem solver monitors progress, considers alternative plans, and reviews what has been accomplished.

## Problem Solving Methods

Problem solving methods refer to the principles for selecting operators (Anderson, 1993a). These methods can vary from blind search to executing an algorithm that is guaranteed to find a minimum-step solution. Traditionally, problem solving methods have been divided into two categories: weak-methods and strong methods. Weak methods are general problem solving heuristics that can be applied to a wide variety of tasks. For example, means-ends analysis can be used in solving cryptoarithmatic problems and chess games. Strong methods are domain specific, an example being problem solving methods that apply to solving math word problems. However, cognitive research indicates that humans tend to use simpler methods (Anderson, 1990).

Several weak method problem solving strategies have been identified as characteristic of novice problem solving methods in knowledge-lean tasks. These methods include, forward chaining, backward chaining, random search, trial-and-error, operator subgoaling, random search, hill climbing, and means-ends analysis (Anderson, 1987; Laird, Newell & Rosenbloom, 1987; Lesgold, 1988; Newell, 1980; Newell & Simon, 1972). In *forward chaining* search starts with the initial state. Heuristics are then used to select an operator from those applicable to the current state, and the operator is applied and the strategy repeats. *Backward chaining* can be used only when a solution state is specific and the operators are invertible. This strategy builds a solution path from the final state to the initial state. In *random search* all possible solution paths from the initial state to the goal are considered in an unsystematic (random) manner. In *trial-and-error* all possible solution paths from the initial state to the goal are searched systematically. In *operator subgoaling* one chooses an operator without considering whether the operator can be applied to the current state. In *random search*, the problem solver randomly selects a move to one of the possible next states. In *hill climbing*, the problem solver always selects the move that will get the problem closer to the goal. *Means-ends analysis* involves two features observed in human problem solving, *difference reduction* and *subgoaling*. Difference reduction refers to the tendency of problem solvers to select operators that produce states that are more similar to the goal state. Subgoaling involves the creation of a stack of subgoals. For example, in using means-ends analysis a problem solver might set the goal of transforming the current state into a goal state and create a subgoal and the appropriate operator to eliminate the difference. In means-ends analysis the problem solver compares the current state with the goal of the problem and or a subgoal that the problem solver is trying to achieve, and an operator is selected that can reduce the differences between the current state and the goal.

Research on transfer and expertise highlights the role of domain-specific problem solving and the role of strong methods (e.g., Chi, Glaser & Farr, 1988). Strong problem solving methods are domain-specific and require a great deal of domain knowledge. Smith (1991, p. i) has summarized this line of research by noting that "recent research in medicine and certain other domains has strongly emphasized the context specificity of the problem solving process."

**Types of Problems**

A distinction has been made between well-defined and ill-defined problems. A well-defined problem has a clearly specified given state, goal state and operators that may be applied to problem states. For example, an algebra example such as X+2=6(X-4) is considered a well-defined problem since the initial state (i.e., the equation), the goal state (i.e., solve for X), and the legal operators (i.e., adding and subtracting known and unknowns on each side of the equation) are clearly stated. In contrast, an ill-defined problem lacks a clear specification of one (or more) of the given state, the goal state, or the operators. For example, solving a case of a hemodynamically unstable patient with multiple injuries following a motor vehicle accident might be considered ill-structured since the initial state, operators, and goal state are not clearly specified.

Other types of problems include, *routine*, *nonroutine*, *adversary*, *nonadversary*, *transformation*, *arrangement*, *induction*, *deduction*, and *divergent* problems. A *routine* problem is one which a problem solver has learned to solve previously. A *nonroutine* problem is unlike any other problem that the problem solver has encountered before, so that a solution must be original. *Adversary* problems involve two or more solvers who compete against one another in a game such as chess. *Nonadversary* problems do not involve competition against opposing players (e.g., Tower of Hanoi). *Transformation* problems provide a problem solver with the initial state, but, he/she must determine the proper sequence of operators to apply in order to transform the given state through a series of intervening states to the goal state (e.g., algebra equations). In *arrangement* problems, all of the elements of the problem are given and the problem solver must determine how to recognize the givens in order to satisfy the goal (e.g., cryptarithmetic problems). *Induction* problems consist of providing a series of instances from which the problem solver must induce a rule or pattern that describes the structure of the problem (e.g., analogy problems). In *deduction* problems, premises are given and the problem solver must apply the appropriate rules to draw a conclusion (e.g., syllogisms). Finally, in *divergent* problems, the problem solver is given a situation and asked to generate as many solutions as possible (e.g., list all the possible uses for a brick).

## Schema-Driven Problem Solving

Extensive research has been conducted in numerous knowledge-rich domains such as word problems in physics and mathematics in order to develop an understanding of schema-based problem solving. Routine problem solving (also referred to as schema-driven problem solving) is based on the assumption that an expert develops a collection of schemata through repeated experience with solving particular types of problems. Routine problem solving occurs when an individual repeatedly solves the same type of problems. With repeated experience the solution process is routinized and therefore the need for attention to understanding the problem and search for a solution are circumvented. For example, if the expert is given an easy, routinely encountered problem, then he/she will solve it without any search. According to VanLehn (1989; 1996) the expert selects and executes a problem solution that is appropriate to the problem. In this case, the understanding process involves deciding what class of problem it is, and the search process consists of executing the solution process associated with that class of problems.

In routine problem solving, it is assumed that experts have a large variety of problem schemata. A problem schema consists of information about the class of problem it applies to and the information about a solution. In other words, an expert encodes his/her knowledge in a schema which is used in recognizing and solving a particular class of problems. According to VanLehn (1989) routine problem solving consists of three processes: (a) selecting a schema, (b) adapting (instantiating) it to the problem, and (c) executing its solution process. Schema selection is based on a triggering process that is not well understood whereby a particular schema "pops into mind." The triggering of schema guides the interpretation of the rest of a problem. Studies of radiology diagnosis (e.g., Lesgold et al., 1981; 1988) have provided evidence that this schema triggering process occurs rapidly within the perceptual phase of diagnosis while a participant is scanning an image. Similarly, researchers studying experts reading physics problems (Chi, Feltovich & Glaser, 1981) found similar evidence of schema triggering. The second process involves the instantiation of a schema to a given problem. Instantiation involves adapting the schema to the given problem. Instantiation involves noting which information is given in the task domain and what is missing. This includes adding the parts of the problem (i.e., the fillers) into the slots of a schema. In routine problem solving, instantiating a schema means filling in its slots. Schema instantiation is viewed as an elaborative process since it enriches the description of the problem rather than changing the problem state. The third process involves the production of a solution to the given problem. This is an important step in the problem solving process since it has been shown to be difficult to demonstrate experimentally what exactly the problem solver's solution procedures are.

There are many cases where schema-driven problem solving can be nonroutine. An example is the case where more that one schema is applicable to a given situation. Hypothetically, in this case, schema selection would entail the application of an operator that would produce a new state (Larkin, 1983). Thus, it may be useful to view schema selection as search. Another situation occurs when no schema will cover the whole problem but parts of other schemata do. Larkin (1983) has provided evidence that experts combine schemata when solving word physics problems. A third type of difficulty occurs when a solution process is halted during the execution of a solution procedure. This state is called an *impasse* (Brown & VanLehn, 1980; VanLehn, 1982) and is based on extensive research originally used to explain children's behavior in executing arithmetic procedures. Larkin (1983) and VanLehn (1982) demonstrated that participants respond to an impasse by *repair* which entails rejecting the adopted schema and selecting another. This section has reviewed some important processes involved in routine and nonroutine schema-driven problem solving including schema selection and instantiation, impasses and repairs.

**Characteristics of Human Problem Solvers**

A recent review by Mayer (1989) has summarized some important characteristics of human problem solvers. The five characteristics are as follows:

*(1) Humans systematically distort problems to be consistent with prior knowledge.*

Problem solvers use prior knowledge when interpreting an initial problem representation, and this can result in distortions that make a problem consistent with prior knowledge. This type of distortion may result in a problem solver working on problem that is different than the one that was originally presented.

(2) *Humans focus on inappropriate aspects of problems.* The initial limited representation that a problem solver creates can limit the subsequent search process. For example, when presented with a word problem a problem solver may represent the given goals and operators in a way that is consistent with the problem statement but which limits the process of problem solution. Thus, the initial representation is a crucial step that can affect the solution process.

(3) *Humans change the problem representations during problem solving.* As the search process progresses from the initial problem representation to the goal state the problem solver may reformulate the original representation of the problem.

(4) *Humans apply procedures rigidly and inappropriately.* Rigidity in problem solving occurs when the problem solver applies well-learned but inappropriate procedures. Automaticity of problem solving procedures can lead to rigidity in the problem solution process.

(5) *Humans let their beliefs guide their approach to problem solving.* The belief system of the problem solver affects the control of problem solving including reasonable ways of solving a problem, when to terminate the problem solving process, problem solver's persistence, and selection of solution procedures.

**Summary**

In sum, this section has introduced an overview of the theoretical problem solving constructs, a brief history of problem solving, a description of the problem solving process, factors affecting problem solving, principle constructs used to describe problem solving, problem solving methods, types of problems, schema-driven problem solving, and the characteristics of human problem solvers. There are many other issues related to the research on problem solving that has not been discussed including effects of practice, problem isomorphs, transfer and problem solving by analogy just to enumerate a few. An extremely important and related area of problem solving is the study of expert-novice differences which is discussed in the following section.

## The Study of Expertise

Important early work on expertise includes studies by deGroot (1946) and Chase and Simon (1973). Their pioneering work on chess propagated a great deal of subsequent research (Chi, Glaser & Farr, 1988; Ericsson & Smith, 1991b; Ericsson, 1996; Feltovich, Ford & Hoffman, 1997). This research is based on the IPT framework which has attempted to explain exceptional performance in terms of knowledge and skills acquired through experience. According to the IPT, elementary information processes and basic capacities remain intact during skill acquisition and incremental increases in knowledge and skill are due to the extended effect of experience (Ericsson & Charness, 1997). The focus of this section, however, is on the IPT approach to studying expertise and thus includes the following (a) the study of expert performance, (b) research results on expert-novice differences, and (c) criticisms of contemporary expertise research.

### The Study of Expert Performance

Expert performance has been defined as consistently superior performance on a specified set of representative tasks in a domain (Ericsson & Smith, 1991a). The study of expert performance involves the use of some *metric* to identify expert performance. Depending on the domain, expert performance is determined by judges, results of competitive tournaments (e.g., sports), certification (e.g., medicine), official standings (e.g., chess) or nomination (e.g., academia). The identification of expert performance is based on a standard methodology in cognitive psychology. This involves analyzing the methods participants use to generate the correct response to a specific task given their knowledge about the domain.

Charness, Krampe and Mayr (1996) have recently outlined a taxonomy of five factors important to expertise and skill acquisition. The factors are: (a) external social factors, (b) internal motivational/personality factors, (c) external informational factors, (d) practice, (e) and the cognitive system. The *external social factors* include parent support, coaches, role models, cultural support, financial support, and competing demands. The *internal motivational/personality factors* include the degree of introversion and extroversion, attention span, compulsivity, and competitiveness. *The external information factors* include organizational aspects of the domain (e.g., clubs, national structure, rating system), and dissemination channels (e.g., journal, newsletters, books, databases, magazines, etc.). There are several important aspects of practice, such as the intensity (deliberate or casual), duration, and content of a practice session. The *cognitive system* focuses on the knowledge base (including chunk size, retrieval structures), problem solving processes (including representations, and search mechanisms), working memory capacity, speed of processing, learning rate, and forgetting rate. This taxonomy can serve as a framework for researchers studying expertise in various domains. For example, someone studying sports expertise is probably more likely to focus on the external social factors and practice. In contrast, a researcher studying medical cognition might focus their investigations on the role of the cognitive system and practice.

The typical approach to the study of expertise has consisted of comparing the performance of participants with varying levels of experience (two or more levels) as they perform particular tasks. This consists of comparing the performance of a group of experts in a particular domain to that of a group of novices. However, it is not sufficient to compare experts and novices only on task outcomes. The fact that experts perform better than novices does not sufficiently explain the underlying properties of human information processing and the problem solving processes involved in expert behaviour. A finer-grained level of analysis is required to reveal the underlying reasons that account for expert-novice differences in task outcomes.

The typical research study includes a task analysis of the domain followed by the collection of verbal protocols of participants with varying levels of expertise in a particular domain as they solve a problem. The verbal protocols are segmented, coded and analyzed with the objective of investigating differences between groups based on the problem solving strategies and/or knowledge structures as they appear in the protocol.

Three general types of experimental paradigms have been employed in expertise research (Ericsson & Smith, 1991a; Gilhooly, 1990; Olson & Biolsi, 1991): (a) verbal or "think-aloud" protocols collected during problem solving, eliciting a problem representation through retrospective reports, or eliciting explanations of certain problem aspects, (b) knowledge structure probing tasks like concept sorting, similarity rating, intermittent probing of the problem representation, free associations and interview studies, and (c) recall or recognition studies, in which the recall or recognition tasks are either announced before the study task is administered (intentional memory) or not (incidental memory). A large number of studies has been conducted employing each of these major paradigms in various educational and professional domains.

An enormous body of research on expertise has accumulated as cognitive scientists have investigated the underlying properties of human information processing and problem solving in various domains. These include: chess (Chase & Simon, 1973; Ericsson & Smith, 1991a; deGroot, 1978), algebra (Novick, 1988), geometry (Koedinger & Anderson, 1990), physics (Chi, Feltovich, & Glaser, 1981; Clement, 1991; Elio & Scharf, 1990; Larkin, McDermott, Simon & Simon, 1980; Simon & Simon, 1978), biology (Chi, de Leeuw, Chiu & LaVancher, 1994), bridge (Charness, 1979; 1989), computer programming (Adelson, 1981; Anderson, Pirolli & Farrell, 1988; Corbett, Anderson & Patterson, 1990; Ehrlich & Soloway, 1984; Greer & McCalla, 1994; Soloway, Adelson & Elrich, 1988), military strategic thinking (Forsythe & Barber, 1992), electronics trouble-shooting (Lajoie & Lesgold, 1989; Lesgold & Lajoie, 1991), technical writing (Breuleux, 1991), surgery (Kushniruk, Patel & Fleiszer, 1995), psychiatry (Cantor, Smith, French, & Mezzich, 1980; Nurcombe & Fitzhenry-Coor, 1982), cardiology (Feltovich, Johnson, Moller & Swanson, 1984; Hassebrock & Prietula, 1992; Johnson, Duran, Hassebrock, Moller, Prietula, Feltovich, & Swanson, 1981), neurology (Gale & Marsden, 1983;

Wortman, 1966), radiology (Lesgold et al., 1981; 1988; Rogers, 1992; 1996), writing (Scardamalia & Bereiter, 1991), auditing (Tubbs, 1992; Vaatstra, Boshuizen, & Schmidt, 1993), music (Ericsson, Krampe, & Tesch-Romer, 1993; Sloboda, 1991), history (Wineburg, 1991), typewriting (Gentner, 1988), mental calculation (Staszewski, 1988), and judicial decision-making (Lawrence, 1988) (for an extensive review see Chi, Glaser & Farr, 1988; Ericsson, 1996; Ericsson & Smith, 1991b; Feltovich, Ford & Hoffman, 1997). The results of the novice-expert differences are discussed in the following section.

**Research Findings on Expert-Novice Differences**

Experts develop an organization of knowledge that can be understood through the examination of their memory performance, pattern recognition, problem representation, and application of problem solving procedures in their domains. In general, the results can be summarized by stating that experts: (a) excel mainly in their domain, (b) perceive large meaningful patterns in their domain, (c) are faster than novices in solving problems in their domain, (d) quickly solve these problems with little error, (e) have superior short-term and long-term memory, (f) see and represent a problem in their domain at a deeper (more principled) level than novices, (g) spend a great deal of time analyzing problems qualitatively, (h) have strong meta-cognitive skills, (i) develop expertise from knowledge initially acquired by weak methods (e.g., means-ends analysis), (j) solve typical problems in their domain by searching forward during problem solving from given information rather than backwards from goals, (k) performance improves steadily with practice, and (l) perform in a way similar to non-experts outside their domain (Glaser & Chi, 1988; Holyoak, 1991).

A number of general conclusions have been drawn concerning the nature of expertise based on extensive research in the area. Briefly, expertise research results can be summarized as follows.

1) Expert knowledge is better structured and organized than is non-expert knowledge. This superior structuring and organization is conceptualized and accounted for by a large

repository of schemata (Gilhooly, 1990; Glaser, 1986; Rumelhart & Norman, 1978; VanLehn, 1989), by more efficient chunks (Engle & Buckstel, 1978), or by compilation of production rules in a production system (Anderson, 1983b; 1987). As a consequence, experts are able to (a) show pattern recognition behavior or "immediate comprehension" (Ericsson & Smith, 1991a), (b) better remember new information in their domain (Chase & Ericsson, 1981), (c) display inference patterns that can be characterized as forward reasoning (Patel, Evans & Groen, 1989; Patel & Groen, 1986), (d) build a thorough representation of a problem (Chi, Feltovich, & Glaser, 1981; Novick, 1988; Lesgold, 1988), (e) make more inferences from data (Patel & Coughlin, 1985; Patel & Frederiksen, 1984), (f) remember relevant information (Coughlin & Patel, 1987; Norman, Brooks, & Allen, 1989), and (g) display opportunistic behavior in the face of new evidence (Lesgold, 1984; Lesgold & Lajoie, 1991; Waldrop, 1984). The availability of well-tuned, flexible schemata makes it possible for experts to reserve working-memory capacity for additional processing. As a consequence, experts can handle noisy data (Glaser, 1986). Expert performance breaks down, however, if meaningless patterns or random data are presented (Engle & Buckstel, 1978; de Groot, 1978; Norman, Brooks & Allen, 1989). Finally, experts are reported to have superior incidental, but not intentional, memory for information from their domain (Norman, Brooks & Allen, 1989).

2) Although experts also have quantitatively more knowledge than novices, there is no evidence showing differences in basic capacities between participants at different levels of experience (Ericsson & Smith, 1991b). Salthouse (1991) proposed that expertise is the process or processes involved in circumventing the common constraints of human information processing. In fact, on routine tasks, experts are faster and more accurate than novices (Charness, 1979, Norman et al., 1989) as well as less susceptible to time constraints on processing (Schmidt & Boshuizen, 1993).

3) In many cases, expert knowledge is at least partially of a procedural, implicit or tacit nature (Ehrlich & Soloway, 1984; Gick & Holyoak, 1983). For example, findings which indicate that expert physicians use contextual factors in activating diagnostic hypotheses may be interpreted as an indication that they are using tacit or implicit knowledge (Kushniruk, Patel & Fleiszer, 1995).
4) Rule-based problem solving is a feature of novice behaviour (Allen & Brooks, 1991; Dreyfus & Dreyfus, 1986). This is at odds with attempts to implement expert knowledge in expert systems. Further, it is also inconsistent with many psychological approaches to expertise (e.g., Holyoak & Thagard, 1989). However, encouraging novices to learn and apply expert rules (as applied in overlay student models) does not lead to increased performance (Holyoak, 1991). The notion that expertise is not purely rule-based has recently stimulated the development of hybrid models and expert systems based on connectionist networks (Holyoak, 1991; Schyns, 1991).
5) An intermediate effect (inverted-U relationship) between expertise level and a number of dependent measures has often been reported, primarily in the medical domain (Lesgold, 1983; Lesgold et al., 1981; 1988). For example, in Lesgold and colleagues' study, intermediates comprised third and fourth year radiology residents (who were compared to a group of first and second year residents [novices] and a group of staff radiologists [experts]). The dependent measures usually include: (a) recall verbosity (Schmidt, Boshuizen & Hobus, 1988), (b) size of the problem representation (Schmidt, Boshuizen & Hobus, 1988), (c) amount of reasoning during problem solving (Kaufman & Patel, 1988), (d) extensiveness of search (Charness, 1991), (e) number of categories employed in a categorization task (Murphy & Wright, 1984; Norman, Rosenthal, Brooks & Muzzin, 1989), and (f) performance (as in radiology, Lesgold, 1983; Lesgold et al., 1981). Intermediates seem to be especially susceptible to constraints on information processing time, while neither novice nor expert performance is similarly

affected. Also, intermediates tend to be slower than both experts and novices if no time constraints are imposed (Norman, Brooks & Allen, 1989).

The generalizations extending from previous research findings have to be interpreted with caution. For example, the findings indicating that experts display forward reasoning have not always been confirmed (Patel & Groen, 1986). In addition, if information is sequentially presented to a participant (e.g., during interactive tasks), forward reasoning processes may be impeded. Moreover, some studies have demonstrated that this type of reasoning is explicitly associated with novice behavior (Carlson, Khoo, Yaure & Schneider, 1990; Lancaster & Kolodner, 1988). According to Koedinger and Anderson (1990), experts show forward reasoning behavior only on relatively simple tasks. It seems safe to conclude that this type of behaviour is displayed by experienced participants on particular tasks (Gilhooly & Green, 1988). Also, in some domains like historical problem solving, expert behaviour is characterized by deeper, more thorough searching. In general, expertise researchers have used problems that are difficult for novices but routine for experts (Anzai, 1991; Chi, Feltovich & Glaser, 1981; Lesgold, 1981), while only a few have employed either difficult or extremely difficult problems.

**Criticisms of Contemporary Expertise Research**

Several researchers (Hatano, 1996; Holyoak, 1991; Holyoak & Spellman, 1993; Zeitz & Glaser, 1996) have recently challenged the generalizations made by other researchers in terms of novice-expert differences, since the current cognitive theory leaves some research findings on expert performance unexplained. For example Holyoak (1991) enumerates research findings that run counter to the dominant cognitive perspective. Some notable exceptions include:

*(1) Experts sometimes achieve mediocrity.* For example, Camerer and Johnson's (1991) research on expert clinical decision making has indicated that experts are sometimes more accurate than novices but they are rarely better than simple statistical models.

(2) *Experts sometimes feel more pain.* For example, Scardamalia and Bereiter (1991) found that expert writers work harder than nonexperts since they engage in more planning, problem solving, and goal revision.

(3) *Means-ends analysis can impair learning.* According to the cognitive perspective initial skill acquisition through weak methods such as means-ends analysis provides the foundation for expertise development through the mechanism of knowledge compilation. Sweller (1988) found that teaching participants to solve algebra problems by means-ends analysis actually impairs performance on transfer tests. According to Sweller and Chandler (1991) a more effective approach involves free forward search in the absence of an explicit goal.

(4) *Expertise can be decoupled from memory performance.* A standard finding based on early work by deGroot (1965) and Chase and Simon (1973) in the area of chess, has been that experts have superior memory for stimuli in their domains. This effect is stronger when the area of study happens to be memory performance (Ericsson & Staszewski, 1989). However, research in computer programming (Adelson, 1984), medicine (Patel & Groen, 1991) and chess (Charness, 1991) have provided evidence that expertise and performance can be decoupled. For example, Patel and Groen (1991) indicated that memory for clinical cases does not always increase with medical expertise.

(5) *Expert search strategies are extremely varied.* The acquisition of expertise is typically accompanied by a shift from backward search to forward search. However, in computer programming (Anderson, Farrell & Sauers, 1984) and complex medical problem solving (Lajoie, Azevedo, Fleiszer, in press; Patel, Kaufman & Magder, 1996) both novices and experts use backward search from goals. More generally, this finding suggests that expertise in complex, ill-structured tasks is characterized by flexible switching among alternative strategies. Also, it appears that both experts and novices access knowledge in their LTM and attempt to instantiate the current problem

into the instantiated schema which involves both top-down and bottom-up processing (Reimann & Chi, 1989).

*(6) Performance may not show continuous improvement with practice*. Performance on many tasks tends to improve with practice following a power function. Exceptions are noted in complex tasks where acquisition of expertise in not based just on the greater speed and processing efficiency but on the restructuring of the task itself.

*(7) Learning does not always require goals or feedback.* Skill acquisition theories (e.g., ACT-R) emphasize that skill acquisition requires feedback regarding the success or failure in achieving goals. However, recent research on the acquisition of musical expertise is better learned by exposure to music since goal achievement may actually be detrimental. (Sloboda, 1991).

### Summary

In summary, this section presented the (a) study of expert performance, (b) research results on expert-novice differences, and (c) criticisms of contemporary expertise research. Contemporary theories must be elaborated in order to incorporate learning mechanisms other than knowledge compilation and its variants. Other criticisms include the failure to consider sociocultural contexts in the acquisition of expertise (Brown Collins & Duguid, 1989), qualitative changes in the acquisition process (Hatano, 1996), and the fact that there are different courses of expertise development as well different types of experts (Hatano & Inagaki, 1986; Sloboda, 1991).

## The Study of Medical Problem Solving and Expertise

The study of medical expertise has a relatively short history. However, the research in the area of medicine has made significant contributions in diverse areas such as: (a) novice-expert differences, (b) problem solving and diagnostic reasoning processes, (c) the role of medical knowledge, (d) the structure of medical knowledge, (e) medical education and training, and (f) computer-based applications. Empirical studies of medical problem solving[1] originate from two research traditions. The IPT tradition studies cognition in various domains to uncover the cognitive mechanisms underlying performance (Patel & Groen, 1986; Hassebrock & Prietula, 1992). This tradition is also referred to as the "descriptive" theory which considers expert problem solving as the "gold standard." In contrast, the decision making tradition compares human decision making processes with normative models based on expected utility theory (Weinstein & Fineberg, 1980) and probability theory (Tversky & Kahneman, 1973). This theory has been referred to as the "normative" tradition where the normative models are considered the "gold standard." The focus of this section is on the descriptive tradition since it was used as the theoretical framework for this study. In recent years, numerous studies have investigated many aspects of expertise in medicine, including the role of hypothesis generation, reasoning, and a variety of knowledge representation types, and several models of medical cognition have been developed. This section presents a subset of these studies.

**Hypothetico-Deductive Method**

Pioneering research on medical cognition using the descriptive tradition was conducted during the 1970's by Elstein, Shulman, and Sprafka (1978), who applied information-processing theory to the study of medical reasoning. Specifically, they designed a series of research studies to investigate the hypothetico-deductive method of medical problem solving. The hypothetico-deductive method was believed to be a fundamental cognitive mechanism underlying both expert and novice medical problem solving. In hypothetico-deductive methods, the participant generates a small number of hypotheses from the initial information provided in the case, which guides subsequent data

---

[1] The terms "reasoning", "decision making", "judgement", amd problem solving" are encountered in the research literature and are typically associated with disparate theoretical persepctives and corresponding analystical methods. For example, "reasoning" often refers to formal domains such as syllogistic reasoning, "decision making" ofter refers to choices made between therapeutic alternatives, "judgement" often refers to the probabilistic inference, and "problem solving" often refers to tasks that can be represented in terms of problem spaces, operators and problem solving strategies.

collection activities in a deductive manner (from expectations based on hypotheses). Subsequent information then serves to validate or invalidate initial hypotheses, or to suggest additional hypotheses. This method is therefore characterized by the alteration of hypothesis generation and deductive verification of predictions based on hypotheses. This method also emphasizes the process of problem solving as opposed to the content (medical knowledge), a position which was widely favored during that period (Cutler, 1979; Elstein et al., 1978). However, the results of their studies as well as those conducted by various other researchers indicated that:( 1) the hypothetico-deductive method sometimes resulted in errors, (2) expertise often depended on an individual's experience with particular types of cases, and (3) domain knowledge was actually a more important determinant of diagnostic accuracy (Claessen & Boshuizen, 1985; Elstein et al., 1978; Neufeld, Norman, Feightner, & Barrows, 1981; Norman, Jacoby, Feightner & Campbell, 1979).

The results of expertise studies in other domains led Groen and Patel (1985) to question the hypothetico-deductive method in medical diagnosis. The research results from other domains characterized expert performance as pattern-recognition and data-driven. Patel and colleagues (Patel & Arocha, 1988) employed a pathophysiological task to trace the problem solving inferences of medical personnel with varying levels of expertise. Their results (e.g., Patel & Groen, 1988) showed that in comparison to novices and intermediates, experts used (a) data-driven inferencing from symptoms to diagnoses, and (b) "strong" knowledge-based methods instead of goal-driven problem solving (from diagnoses to symptoms). However, this process only characterizes experts solving routine problems in their domains. Interestingly, when solving nonroutine, difficult problems, or problems outside their domains experts resort to backward inferencing where use of these weak methods is associated with diagnostic errors (Patel, Arocha & Groen, 1987).

**Hypothesis Generation**

Numerous studies have subsequently been conducted investigating the role of hypotheses in the reasoning of medical personnel with varying levels of expertise. Several studies by Patel & colleagues (Arocha, Patel, & Patel; 1993; Joseph & Patel, 1990; Patel, Arocha, & Kaufman, 1994) have employed a sequential task presentation method (presenting case information in segments over time) to study how medical personnel with various levels of expertise develop hypotheses. Specifically, the researchers were interested in studying how participants might modify their initial interpretations or hypotheses based on new case information. Their findings included: (a) medical students tended to generate their final diagnosis early (in the first segment) and persist with case representations that were consistent with their final hypotheses even when faced with contradictory information, (b) medical students provided new hypotheses to explain isolated findings (rather than all of the case information), and (c) participants who initially generated more hypotheses tended to generate more of the accurate (or reasonable) diagnoses.

During the course of numerous studies, Patel & colleagues (Braccio, 1988; Patel & Groen, 1986; Patel & Arocha, 1995) have identified several other important findings. One such finding is that experts generate hypotheses earlier in the diagnostic process than novices do and are better able to evaluate them when faced with subsequent data. Secondly, the researchers found that medical students differ in the amount and use of relevant knowledge, with (a) beginners relying on textbook-type descriptions to interpret case information, (b) intermediates developing fragmented problem representations as well as different hypotheses to explain different findings, and (c) advanced students attempting to provide more encompassing diagnostic hypotheses.

Finally, various researchers have argued that training medical personnel to use hypothesis-directed strategies, such as the hypothetico-deductive method, might interfere with the acquisition of the basic schemata required to understand and solve problems in a domain (Boreham, 1995; Custers, 1991).

**Breadth-first Versus Depth-first Search**

Based on the results of several studies, breadth-first search is believed to be a powerful problem solving strategy used by successful problem solvers (Klahr & Dunbar, 1988). This type of search involves an initial generation of several hypotheses which are concurrently evaluated in relation to the available data. This evaluation results in the elimination of the less plausible hypotheses and the remaining ones can be tested subsequently. In contrast, depth-first search involves considering each individual hypothesis in-depth before considering the range of possible hypotheses. This strategy is less powerful, since the problem solver maintains and attempts to evaluate multiple hypotheses during the problem solving process. Various findings suggest that as expertise and domain knowledge increase, there is a shift in strategy use from depth-first to breadth-first. According to Patel, Arocha, and Kaufman (1994) and Patel, Groen, Ramoni and Kaufman (1992) the change in strategy use is a result of increasing levels of knowledge.

**Reasoning Strategies**

Various studies have investigated types of medical reasoning strategies and their use based on levels of expertise (Kuipers & Kassirer, 1984; Norman, 1988; Patel & Arocha, 1995; Patel & Groen, 1986; Patel, Groen, & Norman, 1993; Patel & Ramoni, 1997; Ridderikhoff, 1991). An important distinction which has been widely discussed is that of data-driven versus hypothesis-driven reasoning strategies. Data-driven strategies are guided by the available data (Laird & Newell, 1993; Van Lehn, 1991) and, in medicine, are used by experts working with familiar medical problems. Alternatively, hypothesis-driven reasoning strategies are guided by hypotheses or possible disease types, and portions of the case data may be isolated to support or refute individual hypotheses. These strategies are typically used by novices and may be used by experts diagnosing unfamiliar disease types.

In a cardiology study, Patel and Groen (1986) found that experts had highly organized knowledge of underlying disorders and, as a result, always used data-driven (in their terminology 'forward') reasoning when explaining the main component of their diagnoses. In a later study, Patel, Groen, and Arocha (1990) found that this pattern of expert behaviour was interrupted in cases which contained "loose ends" in the form of inconsistent data or unexplained findings. In these cases, experts reverted to using hypothesis-driven (in their terminology 'backward') reasoning in order to provide coherent explanations about a case. Finally, the researchers also indicate that, while diagnostic accuracy does not require the use of data-driven reasoning strategies, inaccuracy is always associated with an interruption in the use of these strategies.

**Biomedical Knowledge Versus Clinical Knowledge**

Numerous studies have been conducted to investigate the roles of basic science or biomedical knowledge and clinical knowledge in medical reasoning (Boshuizen & Schmidt, 1992; Feltovich & Barrows, 1984; Patel, Groen, & Scott, 1988; Schmidt & Boshuizen, 1992). Basic science or biomedical knowledge consists of knowledge of basic physiological processes and anatomy. In contrast, clinical knowledge refers to knowledge of particular disease types (symptoms and underlying pathologies).

The results from several studies on the role of basic science knowledge are summarized here (Patel & Groen, 1986; Patel, Groen, & Scott, 1988; Patel, Evans, & Kaufman, 1990; Patel, Evans, & Groen, 1989). These results indicate that (a) basic science knowledge may be important in facilitating explanation and coherent communication rather than in problem solving itself, (b) physicians do not generally use scientific explanations but do so when they are not sure about their diagnosis of a case, and (c) basic science knowledge is used in elaboration and provides a way for students to generate more coherent explanations (which are more easily remembered). The results from additional studies indicate that the oversimplified representations of biomedical phenomena used by students do not successfully support clinical reasoning, and that a wide range of misconceptions, errors, and other difficulties were exhibited amongst participants of different expertise levels (Feltovich, Spiro, & Coulson, 1989; Kaufman, Patel, & Magder, 1994; Patel, Kaufman & Magder, 1991).

**Knowledge Representations in Medicine**

Several researchers have investigated the role of various types of knowledge representations in the development of expertise in medicine (Feltovich, Spiro, & Coulson, 1997). In general, these representations are considered to be abstractions or extractions of the essential shared elements in different experiences. In the case of medicine, these largely consist of cognitive structures which represent the critical or essential features of disease types. Their recall from long-term memory is triggered by critical features of a particular situation. In general, developing medical expertise is considered to involve the refinement of these representations, with experts having finely tuned representations that allow rapid and accurate recall and diagnosis.

A knowledge representation structure that has generated a lot of cognitive research is the *schema*, which is an active organization of past reactions or past experiences (Bartlett, 1932). A schema can also be defined as a generic representation featuring the main elements of a concept. The medical expertise literature contains numerous examples of studies aimed at investigating the role of schemata in the development of expertise (Feltovich, Johnson, Moller, & Swanson; 1984; Feltovich, Spiro, & Coulson, 1997; Johnson, Duran, Hassebrock, Moller, Prietula, Feltovich, & Swanson; 1981; Lesgold et al., 1988; Norman, Brooks, Rosenthal, Allen & Muzzin, 1989; Patel & Frederiksen, 1984; Turner, 1988). As an example of findings, a set of studies by Johnson et al. (1981) and Feltovich et al. (1984) determined that the completeness of schemata is a critical factor that influences diagnostic accuracy. As compared to novices, experts were demonstrated as having more complete and detailed schemata for diseases and a more cohesive interconnections between different disease schemata. The researchers concluded that these more refined structures facilitate experts' retrieval of relevant knowledge.

A second type of knowledge representation which has recently been investigated by Patel and colleagues (Arocha & Patel, 1995; Patel, Evans, & Groen, 1989) is *the causal-conditional network.* This represents an application of Kintsch's (1992; 1988; Kintsch & Van Dijk, 1978) construction-integration model of text comprehension. Causal-conditional networks are representations consisting of propositions which are interconnected by causal and conditional rules. Applied to medicine, construction integration theory proposes several component processes: a textbase consisting of propositions (concerning patient complaints), a memory store of disease classification knowledge, and a two-phase construction-integration process (Arocha & Patel, 1995).

Another type of knowledge representation studied in relation to medical problem solving has been the *"illness script"* (Custers, 1995; Feltovich & Barrows, 1984). An illness script is a representation of (a) enabling conditions or factors that influence the probability that a patient has a certain condition or disease, (b) faults or physical 'malfunctions' that are related to a complaint, and (c) consequences (signs and symptoms) of illness (Custers, 1991; Custers, Boshuizen, & Schmidt, 1996). Custers (1991) conducted a series of studies in order to investigate the use of illness scripts and compare them to other types of knowledge representation (e.g. schema). Several results support script theory. Findings included: (a) illness scripts become more integrated and consolidated as experience increases, (b) scrambling case presentation interferes with expert performance, (c) basic science knowledge use decreases as expertise increases, and (d) as compared to novices, experts are able to use enabling conditions to activate an appropriate hypothesis much more effectively.

## Models of Medical Cognition

The extensive research in the area of medical cognition has led to the development of theoretical accounts of medical problem solving. There are two notable examples, Patel and Arocha's (1995) construction integration model already described above and Ramoni and colleagues' (1992; 1993; 1997) select and test model (STModel) to be described below. The STModel is a generic model of medical cognition which described briefly in the next section, since is more useful for the purposes of this study.

### The Select and Test Model (STModel)

The Select and Test Model (STModel) is a 'generic' model of medical diagnostic reasoning (Barosi, Magnani, & Stefanelli, 1993; Ramoni, Stefanelli, Magnani, & Barosi, 1992; Patel & Ramoni, 1997) and is presented in Figure 2.1. This model characterizes medical diagnosis as a two-phase, cyclical process of *hypothesis selection* and *hypothesis*

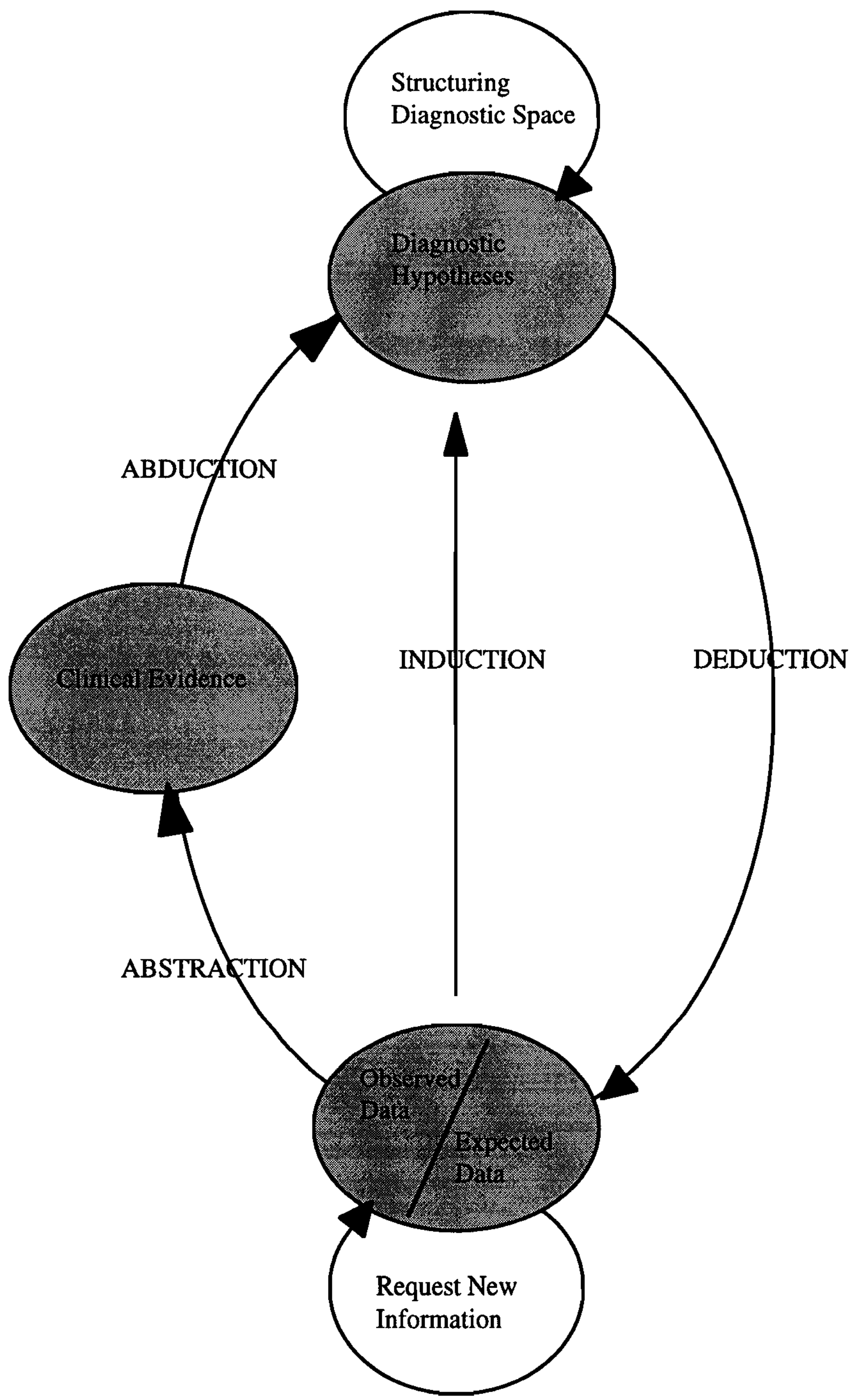



Figure 2.1: The Select and Test Model (Ramoni et al., 1992)

*testing*. It also emphasizes the reasoning processes of *abstraction*, *abduction*, *induction* and *deduction* as key factors in medical diagnosis. Ramoni and colleagues (1992) emphasize that their model reflects the cyclical nature of medical diagnosis. That is, hypotheses about disease are routinely revised based on new information (e.g. test results). These researchers also draw an important distinction between reasoning in the "ideal" sense (i.e. classic logic) and the "limited" or "common sense" way. Medical reasoning is an example of the latter type, in that medical practitioners routinely reason about cases with incomplete information.

For the purposes of this study, an important aspect of the STModel is the four types of medical reasoning it specifies (Ramoni et al., 1992). *Abstraction* can be defined as the process of deriving or isolating important aspects of the information available. In the case of mammography interpretation, abstraction refers mainly to the process of reading the clinical history and examining a set of mammograms for apparently abnormal findings. Once findings have been isolated, *abduction* follows. This type of reasoning is used to hypothesize about potential disease categories and to rank them in terms of probabilities. Thirdly, *deduction* is the reasoning process used to identify what should be expected (i.e. characteristics) of each of the hypotheses identified. Lastly, *induction* results in the identification of the hypothesis which best accounts for what is observed.

In sum, the plethora of studies of medical problem solving and expertise have adopted the IPT as its theoretical framework. The emphasis has been to investigate the problem solving strategies and knowledge representations of medical personnel in various medical specialties in an effort to understand the cognitive mechanisms underlying medical expertise. The widespread proliferation of such studies has resulted in the investigation of issues related of medical cognition including: (a) directionality of reasoning, (b) expert reasoning and domain specificity, (c) hypothesis formation and task difficulty, (d) time-course of hypothesis generation, (e) patterns of reasoning, (f) expert reasoning and task difficulty, (g) nonmonotonicity in the development of expertise, (h) novice reasoning and the coordination of hypotheses and evidence, (i) theories of the development of medical expertise, and (j) implications for classroom teaching and computer-based training.

A review of the research in the area of radiological expertise is presented in the following section by focusing on signal detection, search strategies, and cognitive studies of radiological expertise

## Studies of Radiological Expertise

> "The sick patient is at the very heart of the radiological process, which is initiated by a request for a radiological consultation on a patient who is either known or presumed to be ill. This request reflects the physician's need for further information in order to solve a diagnostic problem" (Abrams, 1981, p. 122).

Radiological expertise has been investigated by numerous authors employing disparate theoretical and empirical paradigms. Three basic "paradigms" that have been applied widely are: (a) signal-detection studies, (b) search studies which investigate eye movement patterns while expert and novices read x-ray films, and (c) cognitive research aimed at eliciting the underlying cognitive and perceptual factors involved in radiological expertise. The vast majority of these studies have focused on the influence of radiographs on the interpreter (i.e., ordinary people, residents and radiologists). Relatively few (Lesgold et al., 1981; 1988; Rogers 1992; 1996; Faremo, 1997) have actually investigated the underlying cognitive and perceptual factors involved in radiological diagnosis. As a result, a fundamental understanding of the constitution and acquisition of radiological expertise has yet to be achieved. The next section will focus on reviewing the literature from these studies.

### Signal Detection Studies

Signal detection theory originated in electrical engineering (by way of psychophysics) and has been a dominant theory in radiological investigations (Norman, Coblentz, Brooks, & Babcook, 1992; Swets, 1996). Signal detection theory (SDT) has often been the basis for studies of image interpretation (Lusted, 1984). The medical applications of SDT focus on two aspects of diagnosis, the capacity to discriminate among diagnostic alternatives (e.g., malignant versus benign), and the decision criterion used to select one alternative over another. Based on this theory, there are at least four possible diagnostic outcomes. A true-positive result occurs when an abnormality is present and the test indicates that it is present. A true-negative is occurs when no abnormality is present and that is exactly what the test indicates. A false-positive occurs when a normal patient is classified as abnormal. Lastly, a false-negative result indicates that a patient is normal when an abnormality is actually present. The diagnostic decisions are made based on conditional probabilities. A conditional probability is the probability of an event occurring given that another has already taken place. In order to do this, STD uses a method termed receiver operating characteristic (ROC). A ROC curve consists of a graph plotting the probability of true-positive against the probability of a false-positive. Further, the abscissa of the ROC curve is the false-positive rate and the ordinate is the true-positive rate. In sum, the radiologist's task can be characterized as a detecting a visual "signal" in a background of visual "noise."

According to Norman and colleagues (1992), the recognition of abnormal shadows on x-rays is viewed as a problem in detecting a signal (radiological abnormality) against a noisy background of body structures. They further state that an explanation for errors of processing can be found in the degree of overlap between the normal (or noise) curve and the abnormal (or signal) curve on some abstracted latent dimension.

These studies have focused primarily on the recognition memory tasks of novices and experts (Berbaum, 1988; Berbaum, Franken, Anderson, Dorfman, Erkonen, Farrar, Geraghty, Gleason, MacNaughton, Phillips, Renfrew, Walker, Whitten & Young, 1993; Berbaum, Franken, Dorfman, Rooholamini, Coffman, Cornell, Cragg, Galvin, Honda, Kao, Kimball, Ryals, Sickels, & Smith, 1991; Dubilet & Hermann, 1981; Kundel, 1985; Swenson, 1980; Swets, 1985). The results suggest that radiological expertise is composed of knowledge of (a) clinically normal examples, and (b) uncharacteristic features signaling pathology. This research paradigm has been identified as a bottom-up approach since it involves observing the influence of film "characteristics" on the observer. The results of these studies generally indicate that radiologists (experts) fail to detect 30% of abnormalities indicated in x-ray films. Furthermore, in 2% of films, they also falsely detect abnormalities where there are none. The causes of these failures include the film's: (a) physical characteristics, and (b) the characteristics of the features. This research has also indicated that the ability to detect is reduced when the participant's view of the film is experimentally confined to the area of abnormality.

Based on the research findings, the advocates of this approach have postulated that the complex process of interpreting radiographs involves eye search patterns, abnormality recognition and diagnostic judgment. Hillard, Myles-Worsley, Johnston and Baxter (1985) proposed that radiologists interpret radiographs by means of schemata, which are internal visual frameworks that develop through experience and training. This process allows for quick differentiation between normal and abnormal findings. Schematic processing functions differently for normal and abnormal radiographic findings. Normal radiographs fit common schemata, are processed automatically, and do not require further analysis.

Abnormal radiographs, in contrast, are memorable because the images differ significantly from the norm. These differences may have important clinical implications. Assessing abnormal features requires a more complex thought process since the diagnostic probabilities are based on the x-ray and the whole patient (Hillard et al., 1985). The initial evaluation is processed by means of a schema that allows one to retrieve past visual experiences (i.e., cases) and knowledge in order to interpret present situations or scenes. Schematic processing has been shown to underlie radiological expertise, and it affects memory as well as other cognitive processing (Hillard et al., 1985).

Further, research findings indicate that the detection of abnormalities in radiographs involves two perceptual components, a fast "Gestalt" phase and a slow search phase (Christensen, Murray & Holland, 1981; Swenson, 1980; Swenson, Hessel & Herman, 1982). The fast component allows detection of abnormalities almost instantaneously by comparing the actual x-ray with the representation (schema) of a normal x-ray stored in memory. The slow component is based on searches involving eye movement, which require considerably more time. Results further indicate that the more experienced observer picks up a greater number of abnormalities in the fast (rapid) perception phase.

To summarize, these findings provide evidence for what some researchers claim to be the role of context in feature detection. The appeal for increased contextual effects in order to account for detection ability suggests that the observer's evolving mental representation of normal and pathological structures does play an important role. Also, these studies have not only focused on the effects of film features on feature detection but have also been experimentally-contrived, lacking ecological validity required to capture the natural problem-solving process of diagnosing radiological evidence.

**Search Studies**

Search studies have investigated the eye movements of inexperienced (radiology residents) and experienced (staff radiologists) participants during the course of radiological diagnosis (e.g., Kundel, Nodine & Carmody, 1978; Swensson, 1980). Search studies have been conducted in two major ways. The first is to experimentally manipulate the areas of the film that can be searched, the time available to search the film, and the information available. The second is to record eye movement data during film reading. A combination of eye movement data and experimental manipulations have also be used. This research paradigm has marked the shift from focusing on errors to focusing on the underlying cognitive processes.

Kundel, Nodine and Carmody (1978 have proposed a five five-step model to describe the radiologist's task. The first stage is orientation to the radiograph as a whole since an experienced radiologist can extract a substantial amount of information from a film after an initial glance (200-300 msec) (Carmody, Nodine & Kundel, 1980; Kundel & Nodine, 1975; Swensson, Hessel & Herman, 1982). In fact, apparent abnormalities tend to be detected almost immediately by comparing patterns held in memory (Chrsitensen et al., 1981). The second stage involves searching the radiograph. The search strategy used depends on whether the lesion is subtle or obvious, and the experience of the reader (Berbaum, Franken & Smith, 1985). The third stage in radiological diagnosis is the recognition of a potential abnormality. A lesion may be recognized but not perceived as a strong enough “signal” to be classified as abnormal. Step four involves making a decision about whether an abnormality is genuine. The fourth stage involves interpretation that leads to a list of diagnostic possibilities (i.e., list of differential diagnoses).

The results from these studies have indicated that there is considerable interpersonal, intrapersonal, and inter-film variation in the overall scanning patterns of radiologists. Search patterns have been characterized as being neither random nor stereotyped. In addition, refixation on the location of suspected abnormality serves successively to build up a mental image (representation) and perhaps its context. Sources of bias influencing eye movement include: (a) perceptions of the film in initial phases of diagnosis, (b) prior knowledge of the patient, (c) prior knowledge of the characteristics of films, (d) the memory and interpretive experiences of the observer, (e) the observer's memory store of interpretative models of anatomy and pathology, and (f) clinical experience. To summarize, findings from search studies tend to indicate that radiological diagnosis is a multi-faceted process involving low-level sensation and perception and higher-level cognitive processes.

## Cognitive Studies on Radiological Expertise

Considering cognitive research, there have been few explicit accounts of the contributions made by the observer during the detection process. The research conducted by Lesgold and colleagues (1981; 1988) aimed to study the observer by modelling the course of cognitive processes rather than by introducing processing constraints into a detection model. The researchers conducted two studies investigating the constitution and acquisition of radiological expertise. Analytical techniques included perceptual probes and in-depth analyses of participants' verbal protocols.

In Lesgold's study, 5 staff radiologists and 23 radiology residents diagnosed ten standard chest x-rays. The residents were in their first, second, third and fourth year of residency. The five staff radiologists had more than 10 years of post-residency training (range of 13 to 27 years of post-residency experience) and estimated that they had analyzed in excess of 250,000 radiographs over the course of their training and medical practice. In contrast, no resident reported having experienced more than 12,500 films. The procedure involved (a) projecting the film for two seconds, (b) asking the participant to report everything seen during the 2 second exposure, (c) prompting the participant to report anything else of interest (e.g., mediastinum), (d) projecting the film again and asking the participant to think aloud, (e) asking the participant for a final report, (f) presenting a small amount of clinical history, (g) allowing participants to view the film again and think aloud, and (h) providing a final report taking into account the clinical data.

The quantitative results were obtained by conducting a three-group one-way ANOVA. The three groups of participants were: (a) staff radiologists, (b) first and second year residents, and (c) third and fourth year residents. The quantitative results indicated that experts significantly outperformed the residents on all quantitative measures. These include: number of findings, number of causes, number of effects, longest reasoning chain, biggest reasoning chain, bigger clusters, and more connectivity between findings.

Based on their in-depth analyses of the participants' protocols (Lesgold et al., 1981; 1988) and the literature on expertise, Lesgold and colleagues (1981; 1988) proposed an explicit description of the behavior of an expert radiologist. This description posited three major phases. First, during the initial phase of building a mental representation, every schema that guides the radiological diagnosis seems to have a set of prerequisites or tests that must be satisfied before it can control the viewing and diagnosis. In the second phase, the expert works efficiently to reach the point where an appropriate general schema is in control. In the third phase, each schema is associated with a set of processes that allow the viewer to reach a diagnosis and confirm it.

In contrast, the researchers describe the less-experienced radiologist's diagnostic performance as characterized by incompleteness in three respects: (a) the confirming or refuting tests are not applied to the invoked schema, (b) a generally appropriate schema is not triggered efficiently enough, and/or (c) the details of the differential diagnosis process are incomplete (Lesgold et al., 1981; 1988).

In general, Lesgold and colleagues (1981; 1988) proposed a multi-step diagnostic process. Initially, a perceptual decision is made, the outcome of which is a differential diagnosis set with associated probabilities. Then, a cognitive process is triggered to resolve ambiguity, either by searching for perceptual features initially missed that might resolve the ambiguity or by taking into account other data sources such as medical history and diagnostic tests. They found that the evolution of diagnostic choice substantially depended on how an individual perceived the features in the x-ray film. These findings indicated that experts use more inferential thinking and as a result, obtain a more precise model of the patient.

More specifically, the research findings indicate that experts build mental representations of patient anatomy, evoke a pertinent schema quickly and exhibit flexibility in tuning their schemata. Secondly, the assignment of x-ray features of normal anatomy schemata determine which features are "left over" and hence show signs of abnormality. Lastly, normal anatomy schemata might contain attached procedures or localization rules for determining where the abnormalities reside. The expert's flexibility in tuning schemata, in the case of a dominant hypothesis and a more remote possibility stemming from inconsistencies presented in the film, depends upon the availability of mental processing capacity. For example, if sub-processes such as localization are not automated and require conscious processing, working memory interference can prevent the construction of an adequately interconnected representation of the patient's anatomy.

To summarize, Lesgold and colleagues (1981; 1988) have characterized expert radiologists as: (a) having the ability to sustain the looking and reasoning cycle even in the face of considerable complexity, (b) being opportunistic planners with very rich recognition and constructive perceptual abilities, (c) being able to ignore irrelevant data, and (d) being more able to take immediate account of relevant data. Lastly, their schema-driven processing was not found to be consistently successful.

A number of general conclusions can be stated. Briefly, cognitive research on radiological expertise (Lesgold et al., 1981; 1988) can be summarized as follows:

1) *Anatomical Representation Ability*: More experienced individuals (a) see more anatomy, (b) build a rich representation of the patient's anatomy, and (c) have greater anatomical specificity in characterizing abnormalities. The issue tends to be one of "robustness of anatomical knowledge", therefore the problem is not being able to see the abnormal part but determining its anatomical location.

2) *Constraint Posting and Global Encoding*: Research has indicated that there are ability-related differences in the extent to which general diagnostic schemata are directly triggered by abnormal features. In general, experts detect patterns of disease that severely constrain the possible interpretation of other abnormalities. In sum, human experts appear to post constraints on their working memory but do defer acting upon them until maximal data are available.

3) *Recursive Diagnostic Process in Radiology*: Protocol analyses have revealed a general diagnostic process comprised of five major components: (a) abnormality location, (b) abnormality feature characterization, (c) anatomical localization, (d) medical explanation, and (e) overall case resolution.

In a recent study on diagnostic radiology in chest radiography interpretation, Rogers (1992; 1996) examined the interaction between perception and problem solving. Her results were used to develop a theory of visual interaction (Rogers, 1995a) and design a computer-based intelligent cooperative assistant (VIA-RAD) (Rogers, 1995b). Verbal protocol data was collected from eight residents and two staff members while they examined seven computer-displayed chest x-rays. Results indicated that accurate perceptual characterization of a finding may still be insufficient to identify a distinct disease category. The level of abstraction used in characterizing findings provided empirical evidence of the transition between the perceptual (e.g., there is a density) and problem solving activities (e.g., in the case where one or more diagnoses are currently active). Bottom-up (data-driven) strategies were supported by use of secondary findings to generate diagnostic hypotheses, use of features to label primary findings, and use of features of primary findings to generate diagnostic hypotheses. Top-down (goal-driven) processes involved (a) confirmation of expectation of secondary findings to support diagnostic hypotheses, (b) use of features of primary findings to rule out competing findings and diagnostic hypotheses, and (c) use of features of primary findings to match or contradict expectations. Three types of errors were identified: (a) detection errors (failure to detect an abnormal features), (b) labelling errors (see an abnormal feature but mislabel it), and (c) integration errors (detecting and correctly labelling an abnormality but failing to use it in the generation of a diagnostic hypothesis).

The VIP model of visual interaction (Rogers, 1995b) is based on the premise that perception and problem solving are distinct processes with different purposes, functions, and structures. However these two distinct processes interact during the process of radiological diagnosis. As such, the VIP model includes bottom-up processes that "deliver" information to the problem solving process, and conversely, depending upon the plans and goals of the organism, the problem solving process has to communicate direction to the perceptual process. This model comprises a mechanism that coordinates these different levels and ensures that plans are executed, modified, or abandoned according to both perceptual information and the current state of the problem solving process. The inputs and outputs of both perceptual processes and problem solving processes require a mediating process to account for the transfer of information. Therefore, the visual interaction process (VIP) is designated the role of maintaining the internal representation depending on the current information from both processes. It also manages the transition from high-level reasoning plans to detailed perceptual plans. For example, suppose there are number of active diagnostic hypotheses and a strategy has been adopted to rule-out some of these hypotheses. A plan to accomplish this may take the following format: (a) find features to rule out competing hypotheses, (b) check these features in the image, and (c) update hypotheses based on what you see. For the perceptual aspect, assume that some abnormality is detected in the image, is labeled, and is delivered to the VIP. The VIP would then examine the internal representation in LTM and retrieve one (or more) hypothesis. In sum, the dynamic nature of the model indicates that the form of information flow varies depending on the direction of processing.

More recently, Faremo (1997) investigated the problem solving processes used by third-year medical students and senior surgical residents in diagnosing breast disease cases. During the experimental sessions, participants were individually asked to identify abnormal mammogram findings for a set of ten breast disease cases, and to provide differential diagnoses and follow-up actions. Verbal protocols were collected as the participants attempted to diagnose the cases. A task analysis was conducted to examine the cognitive processes involved in diagnosing breast disease and the results were used to develop the coding scheme. This coding scheme focused on the use of clinical information, identification of findings on mammograms, differential diagnoses and follow-up actions. The protocols were coded and compared across the two groups. The protocol analyses revealed differences in the problem solving behaviour of students and residents. Groups differed significantly in the accuracy of their responses for findings, diagnoses/differential diagnoses, and follow-up actions based on expert ratings, and the number of differential diagnoses they generated. Students also differed from residents in the number of instances in which they generated multiple diagnoses, in their requests for clinical information, and in the numbers and types of errors they committed (e.g. failing to identify a finding). Detailed analyses were conducted on a subset of the protocols and additional differences both within and between the groups were identified. These differences include the types and frequency of cognitive operators used and the use of hypothesis-driven and data-driven problem solving strategies. Based on the findings from this study, several recommendations were made concerning how the computer-based system should teach breast disease to medical students.

In sum, cognitive research in the area of diagnostic radiology is still in its infancy compared to the corpus of research in other visual domains such as chess and physics. The few studies that have been reviewed have provided an initial characterization of the diagnostic process, the role of schema-driven problem solving, the specification of top-down and bottom-up process involved during the diagnostic reasoning process, and the role of perceptual and problem solving processes. This body of research has focused mainly on the area of chest radiography and in-depth analyses have revealed differences across levels of expertise. However, their utility in terms of developing a process model of mammography interpretation is limited. In order to adequately understand the diagnostic process (in mammography interpretation) a cognitive model characterizing the underlying differences in diagnostic problem solving between radiology professionals with different levels of expertise is needed. Such a model can be developed by studying radiologists' performance during the interpretation of difficult breast disease cases and be elicited using appropriate cognitive science methodologies.

## Statement of the Problem

The purpose of this study was to investigate the problem solving strategies used by staff radiologists and radiology residents during the process of interpreting difficult breast diseases depicted on mammograms. The methodology consisted of an ecologically-valid experiment where the participants diagnosed real cases in their clinical environments without being repeatedly probed (during the diagnostic process) which led to the incorporation of the authentic experimental condition. The second experimental condition (augmented) was based on findings from the expertise research in many domains (e.g., radiology, physics, chess, etc.) that indicate experts are better able to accurately detect relevant task features using "strong" problem solving methods which circumvent understanding and search processes during problem solving (e.g., Glaser & Chi, 1988; Lesgold, 1988; Rogers, 1992; Rogers, 1996; VanLehn, 1989; VanLehn, 1996). As such, this study incorporated two experimental conditions: authentic and augmented. The authentic condition represented the natural problem-solving diagnostic task that radiology professionals perform as part of their job. The task involved reading a type-written clinical history and interpreting the corresponding set of mammograms. In the augmented condition, mammographic findings were highlighted on a second set of films. The incorporation of this experimental condition tested the hypothesis that highlighted mammographic findings (critical to the diagnosis) would facilitate the attainment of a diagnostic schema and thus increase diagnostic accuracy. Third, the study facilitated the characterization of the cognitive processes of radiological professionals by conducting in-depth analyses of their verbal protocols. Lastly, this study aimed at investigating radiological expertise amongst participants with varying levels of expertise (radiology residents and staff radiologists). To summarize, this study empirically investigated the problem solving strategies used by staff radiologists and radiology residents during the process of mammogram interpretation in two experimental conditions (authentic and augmented).

## Research Objectives and Questions

The research objectives and questions addressed in this study included:

1) Identify a cognitive model of diagnostic problem solving in mammography interpretation.
2) Do staff radiologist and radiology residents use different problem solving strategies, operators, and control processes during mammogram interpretation?
3) What are the effects of the authentic and augmented experimental conditions for experts' and novices' on several aspects of the groups' performance, including: (a) number of

mammogram findings, (b) number of observations, (c) number of diagnoses, (d) scanning time to construct an initial mental representation based on the clinical history and set of mammograms, (e) reading time required to solve each breast disease case, (f) accuracy ratings for diagnosis, (g) accuracy ratings for subsequent examination, and (h) accuracy ratings for overall diagnostic accuracy?

4) What are the effects of the authentic and augmented experimental conditions on the frequency and types of errors committed by both groups?

# CHAPTER 3

# METHOD

## Participants

A total of 20 participants, 10 staff radiologists and 10 radiology residents from McGill University's teaching hospital system participated in this study. The ten staff radiologists were practitioners with M.D. degrees and Board Certification in radiology and were affiliated with one of McGill University's teaching hospitals. They included six males and four females. The post-residency training of the staff radiologists ranged from three and a half to 34 years (mean of 20.3 years), including a range of five months to 30 years of mammography training (mean of 13.8 years). Participants' estimates of the number of cases they had analyzed over the course of their medical training varied from 600 to 100,000 mammograms (mean of 30,000 mammograms). They also reported to have "read" (i.e., diagnosed) an average of 66 mammograms per week (range zero to 200 mammograms), and "seen" (i.e., viewed but not diagnosed) an average of 68 mammograms per week (range five to 360 mammograms).

The ten radiology residents had M.D. degrees and were on rotation at one of McGill University's five teaching hospitals. This group was comprised of seven males and three females. They included two third-year residents, one fourth-year resident, and seven fifth-year residents. All of the residents had completed one mammography rotation during their residency training program. They reported to have zero to twelve months of mammography training (mean of six months). Sixty percent of the residents reported to have "read" between 25 and 100 mammograms, while the other forty percent reported to have "read" between 200 and 1,000 mammograms. None of the residents reported that they "read" or "see" mammograms on a weekly basis.

The participants were first contacted by mail. Each received a letter which stated the research objectives, described the methodology, and was signed by McGill University's Radiologist-in-Chief, the consulting mammography expert and the experimenter. Two weeks later, each participant who met the inclusion criterion (residents with at least one mammography rotation and staff radiologists with mammography experience) was contacted by telephone by the experimenter. At this time the nature of the study was further explained and any other questions were answered. Prior to the experimental sessions, each participant signed a consent form (see Appendix A) which guaranteed their anonymity. All of the staff radiologists and radiology residents participated in the study, except for one resident and two staff radiologists. The participants were not remunerated for their participation.

## Materials

The materials consisted of cases, instructions to participants, consent forms, and questionnaires. The equipment used to conduct the study consisted of video and audio equipment. The cases were selected in collaboration with the expert radiologist and subsequently reproduced.

### Materials and Equipment

All participants signed a letter that (a) stated the purpose of the study, (b) gave their informed consent, and (c) explained the provision for their voluntary termination at any point during the course of the study (see Appendix A). The participants also completed a brief questionnaire that indicated the duration and type of their mammography experience (see Appendix B). The questionnaire was used to record each participant's personal data, as well as their previous medical and current radiological experience. The questionnaire data included each participant's: (a) gender, (b) medical specialty, (c) years of medical experience, (d) residency level, (e) estimate of the number of mammograms "seen" and "read" during their entire medical experience, and (f) estimate of the number of mammograms "seen" and "read" per day. Instructions to participants describing the “think aloud” process to be followed during the diagnostic task were type-written and printed on a sheet of paper (see Table 3.1).

Table 3.1

Instructions to Participants

---

**Instructions to Participants**

You will be presented with ten breast disease cases to diagnose. On five cases, the critical findings have already been traced by an experienced staff radiologist.

Each case will be comprised of a brief clinical history and a corresponding set of mammograms. For each case, please read the clinical history out loud, examine and describe the findings as you would normally. Trace the film findings by pointing to them using the permanent marker (either before, after or following your discussion), and provide a diagnosis or differential diagnosis. Suggest further examinations if appropriate.

Please think out loud throughout the **entire** diagnostic process, that is, verbalize all comments and impressions you have as you diagnose each case.

---

Data collection sessions were recorded with a Sony Hi-8 video camera (Model CCD-TR101) mounted on a Vanguard tripod (Model VT-520) and a Marantz high-fidelity audio tape recorder connected to a PZM table microphone. During each experimental session the participant had access to a magnifying glass which could be used to magnify mammographic features and a permanent felt marker which was used to point to the mammographic features. A Sony audio transcriber (Model BM-77) and a Sony Hi-8 VISCA drive (Model CVD-1000) were used to play back the audio and video recordings during subsequent analyses.

**Case Construction**

The ten difficult breast disease cases were selected by the consulting radiologist, in collaboration with the experimenter. The cases were selected so as to include a range of typical and atypical cases with a variety of mammographic manifestations suggestive of benignity and malignancy. Video and audio data were collected as the consulting radiologist diagnosed the cases in her office in front of a view-box. The main objective was to characterize the diagnostic reasoning process as the radiologist integrated the clinical history, previous radiological findings (if available) and mammogram images. For each case, the consulting radiologist was instructed to: (a) construct a clinical history, (b) describe the mammograms in the set (in terms of mammographic view and which breast is depicted), (c) mark the critical radiological findings, (d) discuss the presence of radiological observations, and (e) discuss subsequent radiological examinations, if any, and why they are required (i.e., to either provide additional data and/or constraint the number of differential diagnoses).

Following this process, the cases were prepared for each experimental condition. For cases in the augmented condition, the radiologist highlighted the mammographic findings for each mammogram by using arrows made with a permanent marker. Figure 3.1 illustrates the set of mammograms (Case 1) used in the augmented condition depicting the highlighted mammographic finding with arrows on the mediolateral oblique (MLO) and craniocaudal (CC) views of the right breast. More specifically, there are five arrows on the MLO[2] view and four arrows on the CC[3] view all pointing to the same large partly well-circumscribed mass. For the authentic condition none of the mammographic findings was highlighted.

All mammograms used in this study were reproduced at the Montreal General Hospital's Department of Diagnostic Radiology on a film reproducer. The diagnostic quality of the copies was approved by the consulting staff radiologist (since image resolution is critical in diagnosing mammograms). The original mammograms and the copies were randomly assigned to either the authentic or the augmented experimental condition. In this way, neither of the experimental conditions would be comprised solely of either the original mammograms or the copy films. Therefore, the two experimental conditions consisted of the identical 10 breast disease cases with all of the mammographic findings marked on those belonging to the augmented condition. Each breast disease case

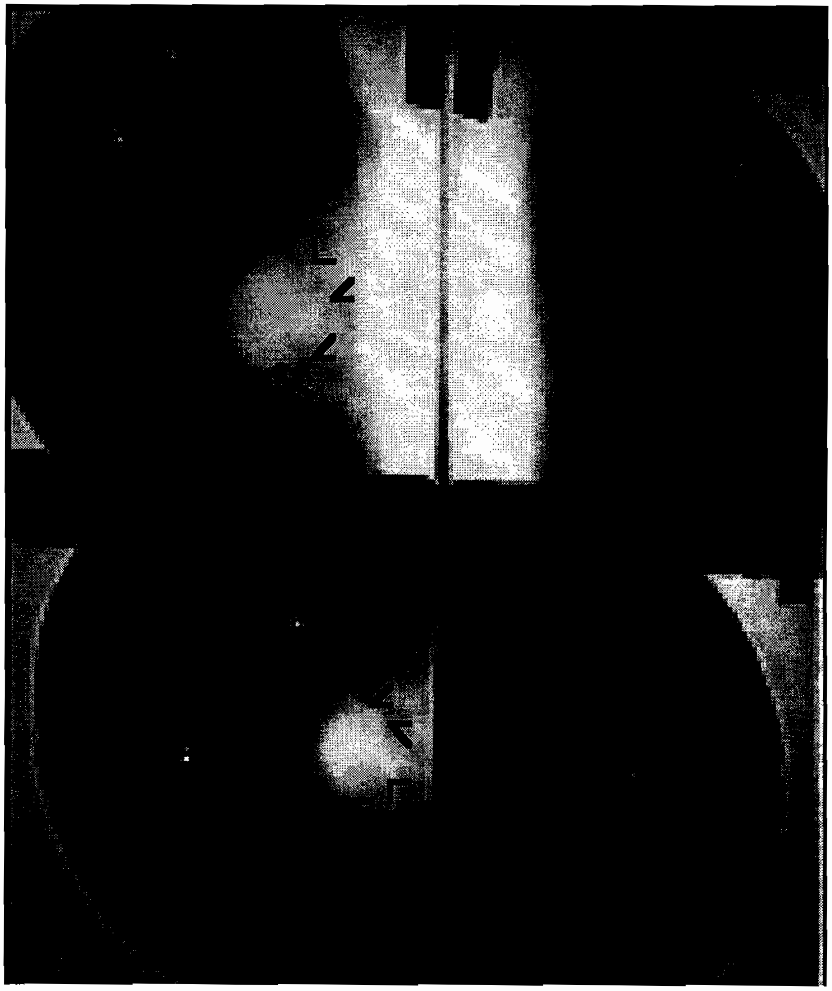
---

[2] In an MLO mammogram the beam enters the breast from the medial side and exposes the film positioned along the lateral surface of the breast.

[3] In a CC mammogram the breast is positioned with the nipple perperndicular to the edge of the recording system and the beam enters from the top of the breast and exits to the detector beneath the breast.

prepared by the consulting radiologist was comprised of a set of mammograms, a clinical history, radiological findings, radiological observations (if any), diagnosis, differential diagnoses (if any), and further investigations (if required)

**Right MLO (Mediolateral)** **Left MLO (Mediolateral)**



**Right CC (Craniocaudal)** **Left CC (Craniocaudal)**

Figure 3.1. An example of a set of mammograms used in the augmented condition depicting the highlighted mammographic finding with arrows on the CC and MLO oblique views of the right breast.

as illustrated in Appendices C to L. Each case that was presented to participants was comprised of a brief clinical history which was taped to a large envelope containing the corresponding set of mammograms.

**Breast Disease Cases**

Ten difficult breast disease cases were used in this study. An additional case was used as the practice case. Cases were selected by the consulting mammography expert, from teaching files which she had not used in the preceding five years. Each case was comprised of a brief clinical history and at least four mammograms including CC and mediolateral MLO views of the left and right breasts. Table 3.2 provides the clinical history, diagnosis, radiological findings (i.e., mammographic features critical in making a diagnosis), and radiological observations (i.e., mammographic features that are not critical in making a diagnosis) for each case used in the study. The cases include three benign (cases 2, 4, and 5) and seven malignant breast disease cases (cases 1, 3, 6, 7, 8, 9, and 10). Detailed case descriptions of the above mentioned information including a brief description of the type of knowledge that may facilitate the diagnosis is provided in Appendix M. The diagnoses for each case were confirmed by pathology reports. For each case, tissue specimens had previously been examined to determine the diagnosis. The breast disease cases included diseases typically encountered in mammography textbooks and clinical research articles, atypical diseases infrequently encountered in daily practice, and diseases with typical mammographic manifestations encountered in daily practice. The mammographic features ranged from fairly obvious to detect to subtle mammographic findings such as pleomorphic microcalcifications that cannot be detected without a magnifying glass.

Table 3.2

The Case Number, Clinical History, Diagnosis, Radiological Findings, and Radiological Observations for the Ten Breast Disease Cases.

| Case Number | Clinical History | Diagnosis | Radiological Findings | Radiological Observations |
|---|---|---|---|---|
| 1 | A 60 year old woman who presented with a mass in the right breast. | Mucin-producing carcinoma (colloid carcinoma or mucinous carcinoma) | (1) large partly well-circumscribed mass (i.e., mass is not completely well-circumscribed) | (1) dense fibroglandular tissue bilaterally<br>(2) large mass lies in the right retroaereolar area |
| 2 | A 54 year old woman who came for a check-up; there was no palpable mass. No significant past history and no previous mammograms. | Left retroaereolar cysts (other smaller nodules not diagnosed but unchanged on follow-up and presumably benign | (1) several small nodules<br>(2) two or three large well-circumscribed nodules | (1) prominent fibroglandular tissue<br>(2) left retroaereolar area is prominent on the left<br>(3) small axillary lymph nodes |
| 3 | A woman in her late forties who presented with a palpable nodule in the lateral portion of left breast. | Infiltrating intraductal carcinoma | (1) an area of retraction with some disorganization in the lateral portion of the left breast<br>(2) poorly defined nodule in the center of the retraction and disorganization | (1) dense fibroglandular tissue |
| 4 | A woman in her sixties presented with no palpable mass and no abnormality on physical examination but she has a family history of carcinoma (her daughter has breast cancer). No previous mammograms. | Benign (calcium in small cysts) | (1) round calcifications on CC view corresponding to linear calcifications on lateral view | (1) several round benign-looking nodules<br>(2) glandular tissue with nodular appearance |

| Case Number | Clinical History | Diagnosis | Radiological Findings | Radiological Observations |
|---|---|---|---|---|
| 5 | This is a young patient in her thirties who presented with a large mass in the lateral portion of the left breast. | Hamartoma or fibroadenolipoma | (1) well-circumscribed mass consisting of fat, and contains encapsulated islands of fibroglandular tissue | (1) dense fibroglandular tissue as usual at this patient's age<br>(2) left breast is larger than the right, this asymmetry is frequent in the normal population |
| 6 | A woman in her fifties who presented with a mass in the upper outer quadrant of the right breast. | Invasive intraductal carcinoma | (1) a density associated with a loss of organization (i.e., architectural distortion and some retraction) | (1) dense fibroglandular tissue bilaterally |
| 7 | An elderly patient who came for a routine examination. There was no abnormality on the physical examination. No previous mammograms. | Carcinoma | (1) a nodule in the center of the right breast<br>(2) microcalcifications of different sizes (i.e., pleomorphic microcalcifications) | • none |
| 8 | A woman in her early fifties who came without a palpable lesion; a few years previously she had breast surgery on the right but the results of surgery are not known. | Carcinoma | (1) irregular (pleomorphic) microcalcifications in a scar | (1) dense fibroglandular tissue bilaterally<br>(2) tissue in the superior region on the left |
| 9 | A 68 year old woman has a palpable abnormality in the lateral portion of the left breast (upper outer quadrant). | Two carcinomas | (1) large spiculated lesion having the appearance of malignancy<br>(2) small retraction may represent a second carcinoma | • none |

| Case Number | Clinical History | Diagnosis | Radiological Findings | Radiological Observations |
|---|---|---|---|---|
| 10 | A 60 year old woman who presented with inflammation of her left breast for two weeks. | Inflammatory carcinoma. Differential diagnoses may include: unilateral edema, lymphoma of the breast, metastasis to the breast, leukemia to the breast, post-radiation, etc. | (1) some asymmetry<br>(2) no discrete mass in the left breast<br>(3) skin thickening around the left breast<br>(4) tissue has a reticulated appearance | • none |

## Procedure

This section includes (a) a description of the experimenter's observation participant activities, (b) an analysis of the domain of mammography and breast diseases, (b) a description of the experimental sessions, and (d) a detailed description of the coding scheme used to analyze the verbal protocols.

### Observation Participant Period

The purpose of the observation participant period (Atkinson & Hammersley, 1994) of this study was primarily to acquaint the experimenter with the methods and behaviors of radiologists in the course of their daily medical practice. The experimenter played the role of complete observer (Bogdan & Bilken, 1992) and therefore did not participate in any of the activities that were observed. These included the diagnoses of breast disease cases in a reading room where both staff and residents spend a considerable amount of time diagnosing various types of cases. The experimenter also spent several weeks observing staff radiologists in their offices during their clinical practice, teaching rounds and mentoring sessions with residents. The observation period provided invaluable information concerning the daily routines of practicing radiologists and radiology residents and an understanding of the interaction between perceptual and cognitive aspects of diagnostic radiology.

The information collected during this phase of the study was subsequently used to (a) construct the instructions to participants, (b) characterize the diagnostic task, (c) design the experimental conditions, (d) collect relevant data, (e) develop a coding scheme at an appropriate level of granularity, and (f) analyze the participants' verbal protocols. Lastly, a comprehensive content analysis of the domain of mammography and breast disease was conducted by consulting various medical sources (Cooper, 1992; Donegan & Spratt, 1995; Homer, 1991; Jackson, 1995; Powell & Stelling, 1994; Tortorici & Apfed, 1995).

**The Domain of Breast Diseases and Mammography**

Certain types of basic knowledge concerning breast disease are required in order to interpret the discussions to follow. For this reason, breast diseases, mammography and mammographic features of breast diseases are discussed in this section.

**Breast Diseases**. Breast cancer is a disease that results from uncontrolled growth and reproduction of abnormal cells in the breast (Donegan & Spratt, 1995). Normal cells divide and multiply also, but not in the random, unrestricted manner characteristic of cancer cells. It is the cancer cell's ability to multiply continuously and spread to other parts of the body that identifies it as "malignant" and potentially life-threatening. Breast cancer is dangerous due to its ability to spread to the lymph nodes under the arms or to other parts of the body. There are other conditions that occur in the breast, such as cysts, non-malignant masses and microcalcifications associated with fibrocystic changes, which are not cancerous, but which also require further evaluation. These conditions are usually referred to as "benign".

Establishing the diagnosis of a breast abnormality involves several important steps (Powell & Stelling, 1994). The first step is usually a physical examination. Most irregularities that form in the breast are not cancer, and most of the diagnoses that are made identify relatively normal conditions such as benign tumors, cysts or calcifications that usually require no treatment. When a lump, also called a mass or lesion, becomes large enough to feel, or palpate, it is said to be palpable. Conversely, a suspicious area that is too small to feel is said to be non-palpable. If one discovers something that may be a change in breast anatomy, the next step is to have the breasts examined by a gynecologist or internist. A physician may refer the patient to a surgeon who specializes in breast cancer and has the necessary expertise to make a reliable diagnosis of any unusual condition. The surgeon may conclude that a condition constitutes one of several natural changes the breast undergoes over time and may recommend only follow-up.

The physician may suspect that an abnormal area is nothing more than a liquid-filled lump or mass called a cyst and attempt to aspirate it by inserting a very thin needle into the abnormal area (Peters, Voegeli & Scanlan, 1989). If it is indeed a cyst, aspiration will remove the fluid and the cyst will collapse. The vast majority of cysts are benign and they are investigated to confirm their diagnosis. If a cyst does not collapse entirely after aspiration, or if the fluid is bloody, the fluid is examined microscopically by a cytopathologist to assess the nature of the cells present. Likewise, if the abnormal area turns out to be solid, the aspirate may contain cells that can also be evaluated microscopically. Only if a surgeon concludes that a condition is suspicious and requires further investigation, will he/she then recommend one or more diagnostic procedures, including a mammogram (Tortorici & Apfel, 1995).

**Mammography.** A mammogram is an x-ray picture of the breast. A radiology technologist performs the mammography by compressing the breasts between two plates attached to a specially designed x-ray machine. The breast is then "photographed" from two separate angles (typically, CC and MLO views of the left and right breasts) and the mammograms are later examined by a radiologist. Mammograms will show the normal features of the breast and may reveal suspicious areas that require further investigation (Homer, 1991). Even if the results of the mammogram are not suspicious, a surgeon may recommend further investigation based solely on the physical examination, as a small percentage of cancers are undetected by mammography. Occasionally women who undergo mammography require magnification or compression views. These magnification views enable a radiologist to better view tiny calcium deposits called microcalcifications or small masses that are undetectable during a clinical breast examination. Often, a magnified or compression view of a suspicious area eliminates it as an area of concern and the radiologist recommends only follow-up. Sometimes, they may recommend a follow-up mammogram in several months to make sure that the area is not changing. Other radiological examinations are sometimes used (Jackson, 1995). For example, an ultrasound of the breast is a procedure commonly used to distinguish a liquid-filled cyst (fluid-filled mass) from a solid mass in the breast. If the mass turns out to be solid, a biopsy might be necessary. A biopsy is a procedure that removes a sample of tissue or cells from the breast so that it can be examined under a microscope by a pathologist for analysis and diagnosis. If a pathologist determines that the tissue or cells are benign (i.e., a benign tumor, cyst or calcification) regular follow-up may be all that is recommended.

## Content Analysis

A comprehensive content analysis (Borg & Gall, 1996) of the domain of mammography and breast diseases was conducted as part of the procedure. The content analysis incorporates the results of the participant observation period and informal meetings with staff radiologists and radiology residents, as well as information on breast disease and mammography from textbooks and clinical literature. The results of the content analysis are discussed in four sections: (a) characterization of mammographic features of breast diseases, (b) classification of benign and malignant breast diseases, (c) further investigations typically performed in cases exhibiting suspicious findings or malignancy, and (d) mammographic features of malignant breast diseases.

The mammographic features of breast diseases are classified as: (a) dominant masses, (b) calcifications, (c) abnormal densities, and (d) associated findings (see Figure 3.2). Each of these can then be further subdivided based on their characteristics (such as margins, shape, size, density, distribution, number, description, and symmetry). These characterisitcs are described below in detail.

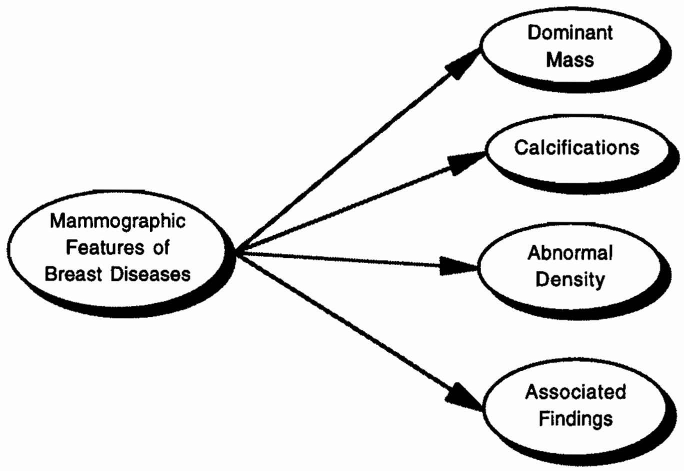



Figure 3.2. Mammographic Features of Breast Diseases.

The first category of mammographic feature is the dominant mass. Dominant masses are typically subdivided according to four characteristics, including (a) margins, (b) size, (c) shape, and (d) density (see Figure 3.3). The margins of a mass can be well-circumscribed (at least 75% of its borders), ill-defined/indistinct, macrolobulated, microlobulated, obscured by tissue, or spiculated. Mass size can be specified as smaller or greater than two centimeters. The shape of a mass can be described as round, oval, or irregular. Mass density can be described as fat-containing (including homogeneous or inhomogeneous), isodense, or high density.

The second category of mammographic features is calcifications. Calcifications are typically classified in terms of three characteristics (a) distribution, (b) number, and (c) description (see Figure 3.4). The distribution of calcifications can be classified as diffuse/scattered, regional, segmental, clustered/grouped, or consisting of multiple groups. The number of calcifications is usually specified as either less than five, five to ten, or more than ten. The description of calcifications is either milk-of-calcium, rim/eggshell, skin, vascular, spherical/lucent-centered, suture, coarse/"popcorn-like", large rod-like, round/oval, dystrophic, punctate (smaller than 0.5 mm), indistinct/amorphous (round or flake), pleomorphic/heterogeneous (granular), or fine (and/or branching and/or linear).

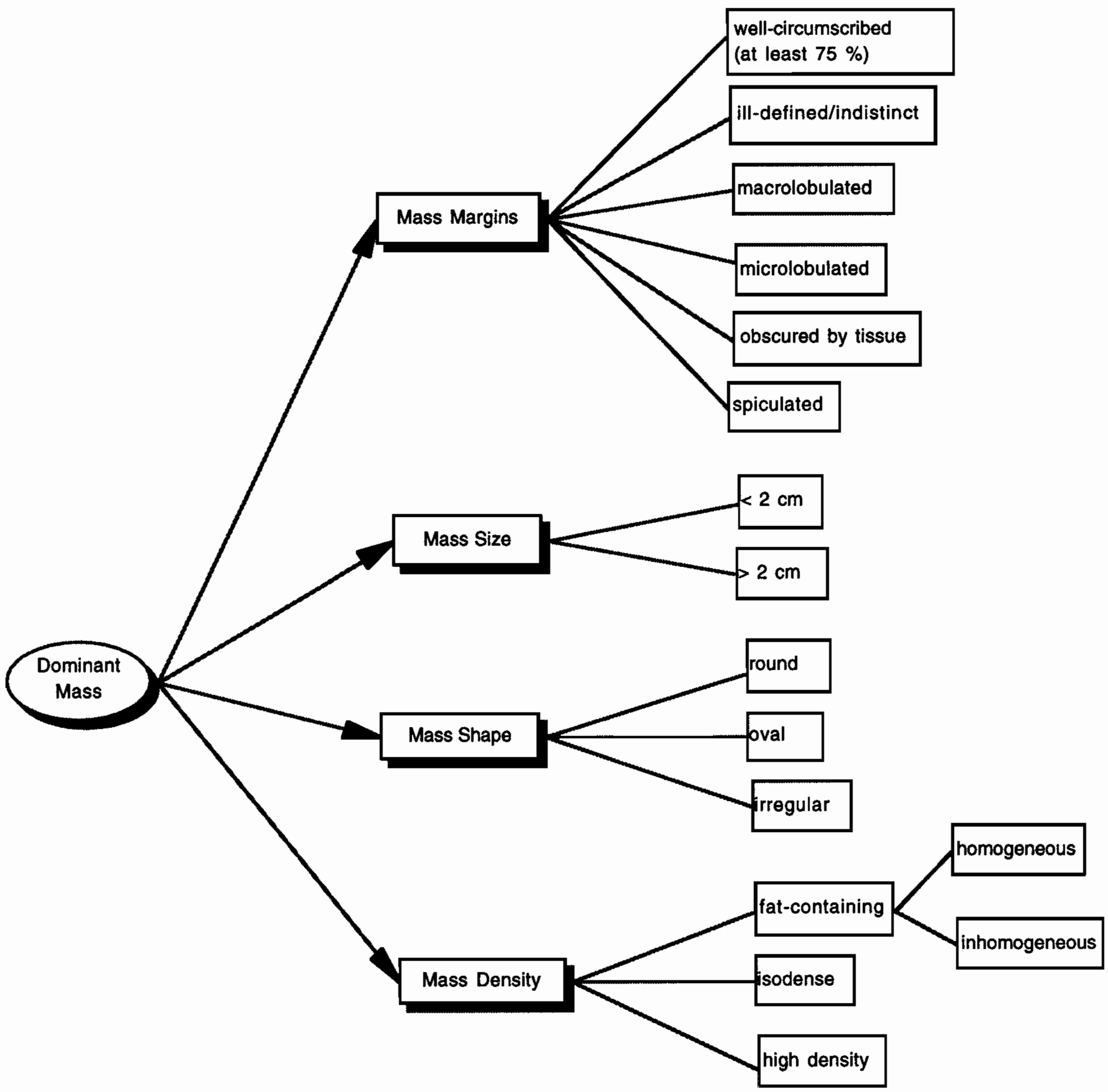



Figure 3.3. Mammographic Features of a Dominant Breast Mass.

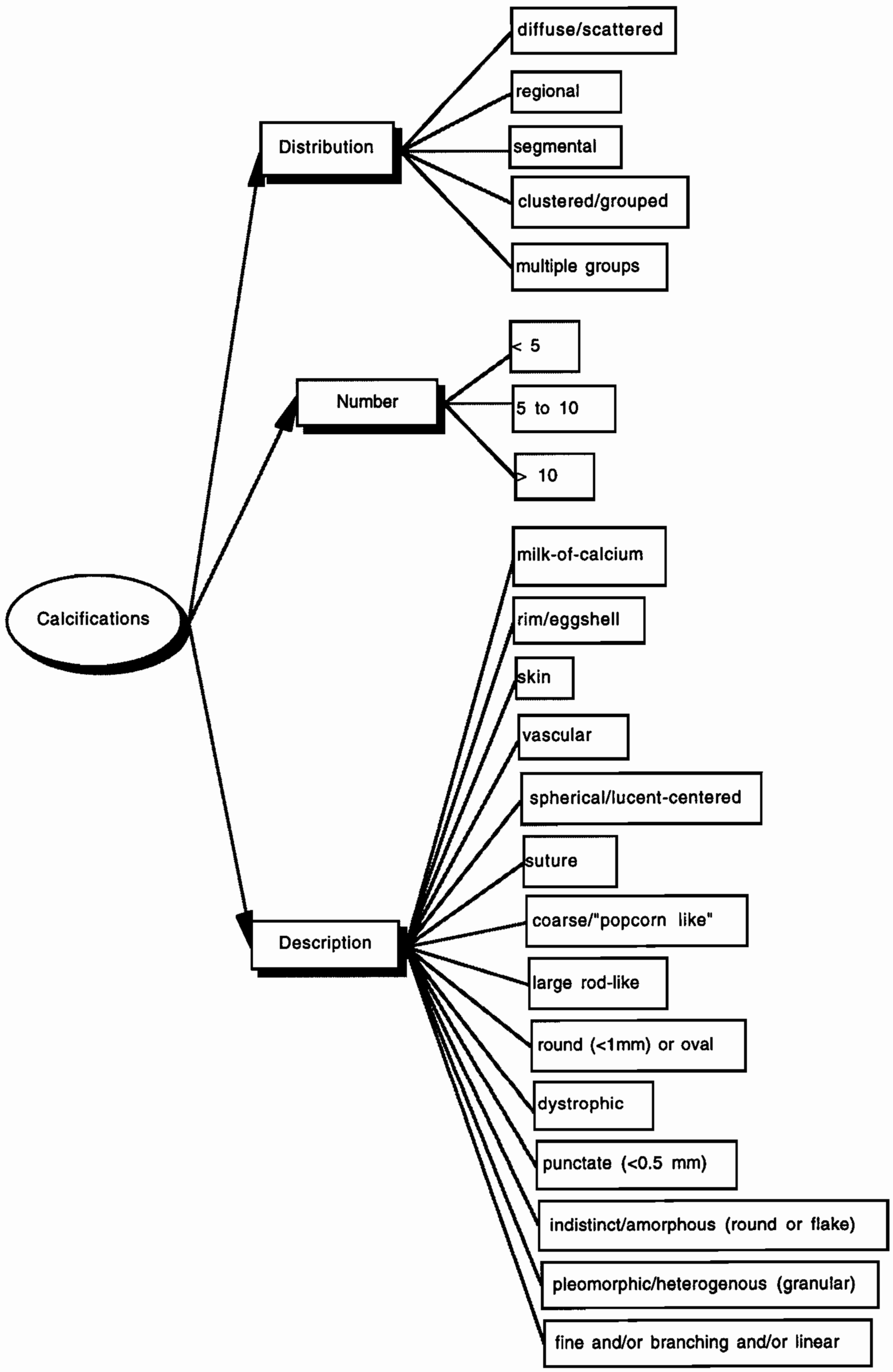



Figure 3.4. Mammographic Features of Calcifications.

The third category of mammographic features is the abnormal density. Abnormal densities are subdivided according to five characteristics (see Figure 3.5). They may be asymmetric and therefore only visible on one mammographic view, visible on more than one view, and/or associated with a palpable mass. They may also be considered to be a developing density. Densities may be associated with architectural distortion or retraction if seen on at least two mammographic views. It may also represent diffuse increased density or multiple masses.

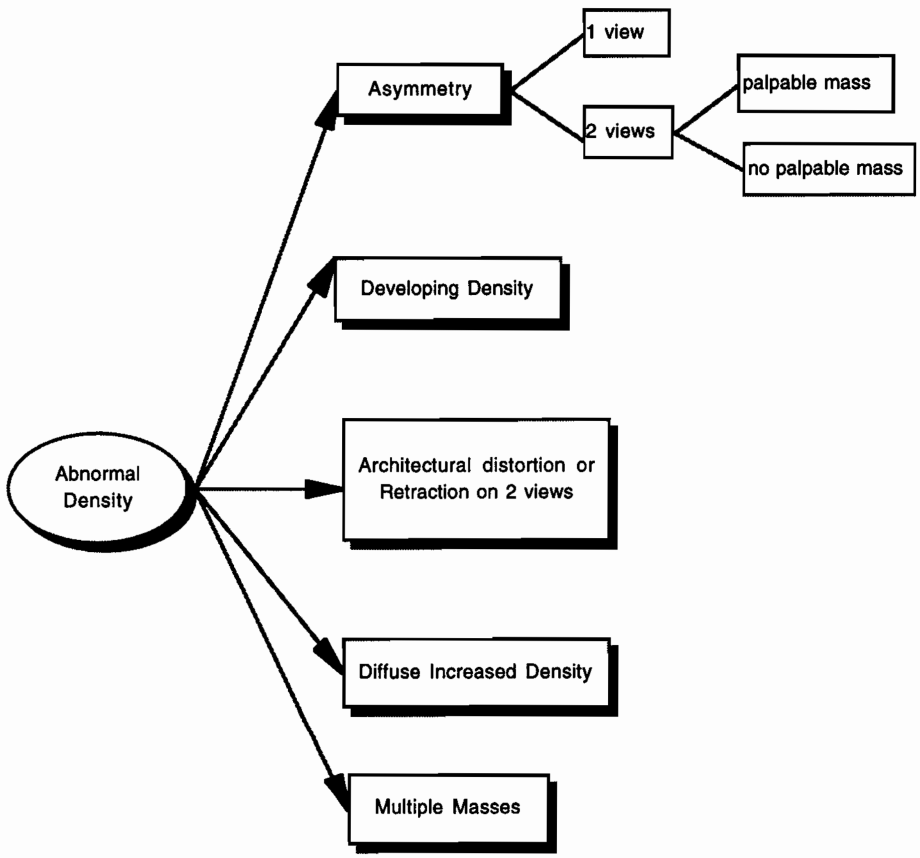



Figure 3.5. Mammographic Features of an Abnormal Density.

The fourth category of mammographic features is termed associated findings (see Figure 3.6). These may include the following: (a) skin lesion, (b) skin thickening (either diffuse or localized), (c) skin retraction, (d) nipple retraction, (e) trabecular thickening,

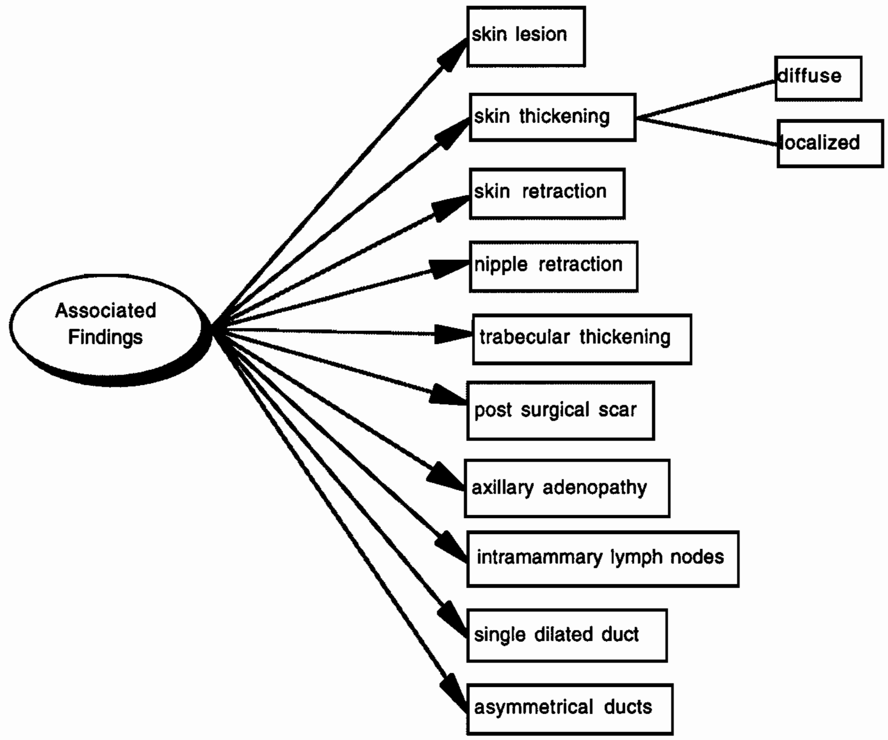



Figure 3.6. Mammographic Features of Associated Findings.

(f) post-surgical scar, (g) axillary adenopathy, (h) intramammary lymph nodes, and (i) single dilated duct, and/or (j) asymmetrical ducts.

The second part of the content analysis was conducted in order to derive classifications of benign and malignant breast diseases. The classification for benign breast diseases (see Figure 3.7) includes fibroadenoma, giant fibroadenoma, phylloides tumor, simple cyst, galactoceles, lymph nodes, haematoma, abscess, and granular cell tumor. Other infections and rare tumors include tuberculosis and filariasis. Benign stellate lesions include radial scar and fat necrosis. A multitude of calcifications are also included, such as calcifications with radiolucent centers, arterial calcifications, duct ectasia, calcified fibroadenoma, postsurgical calcification, milk-of-calcium in tiny benign cysts, and foreign body injection granulomas.

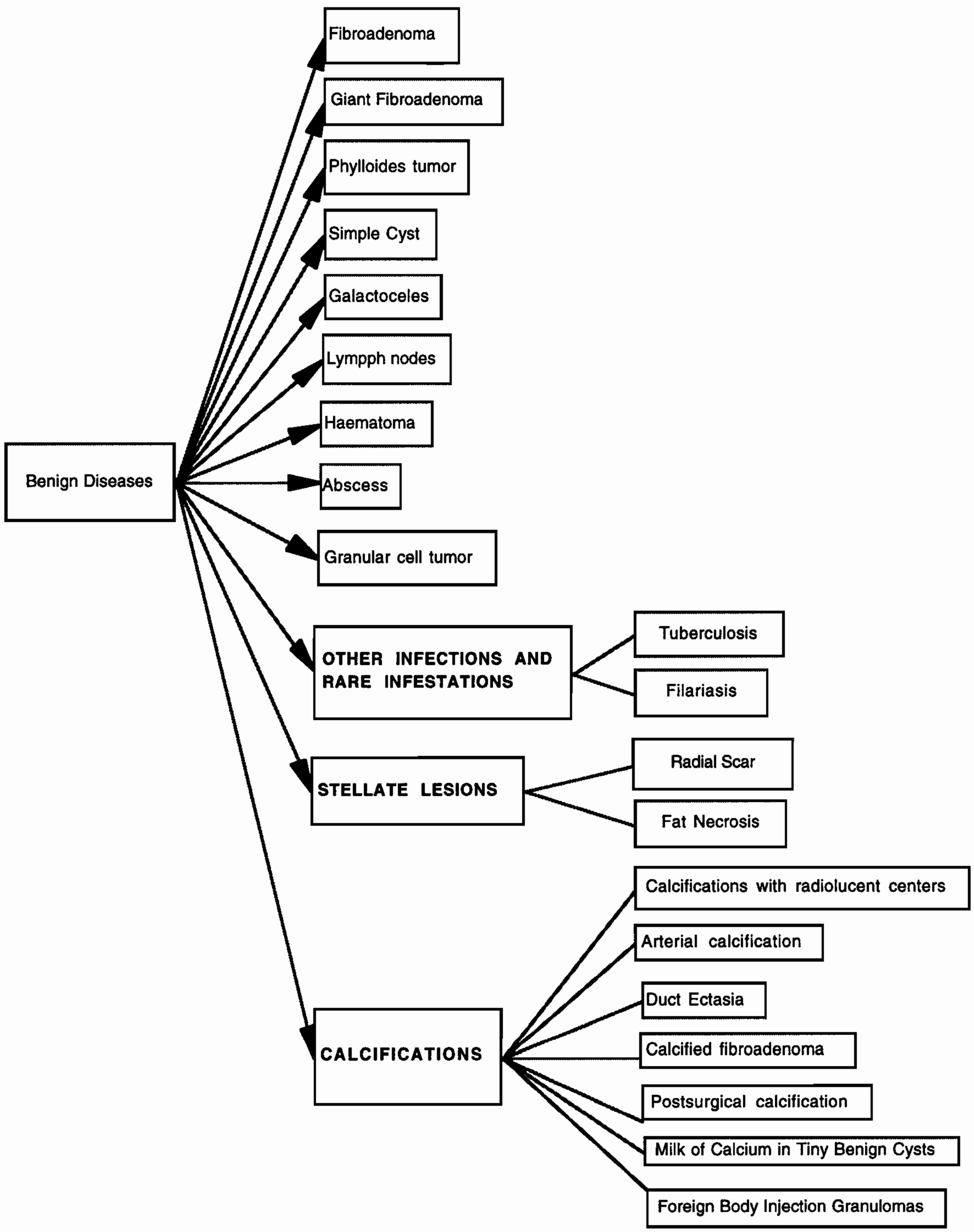



Figure 3.7. A Classification of Benign Breast Diseases Depicted on Mammograms.

The classification of malignant breast diseases (see Figure 3.8) includes ductal carcinoma-in-situ, lobular carcinoma-in-situ, invasive ductal carcinoma (not otherwise

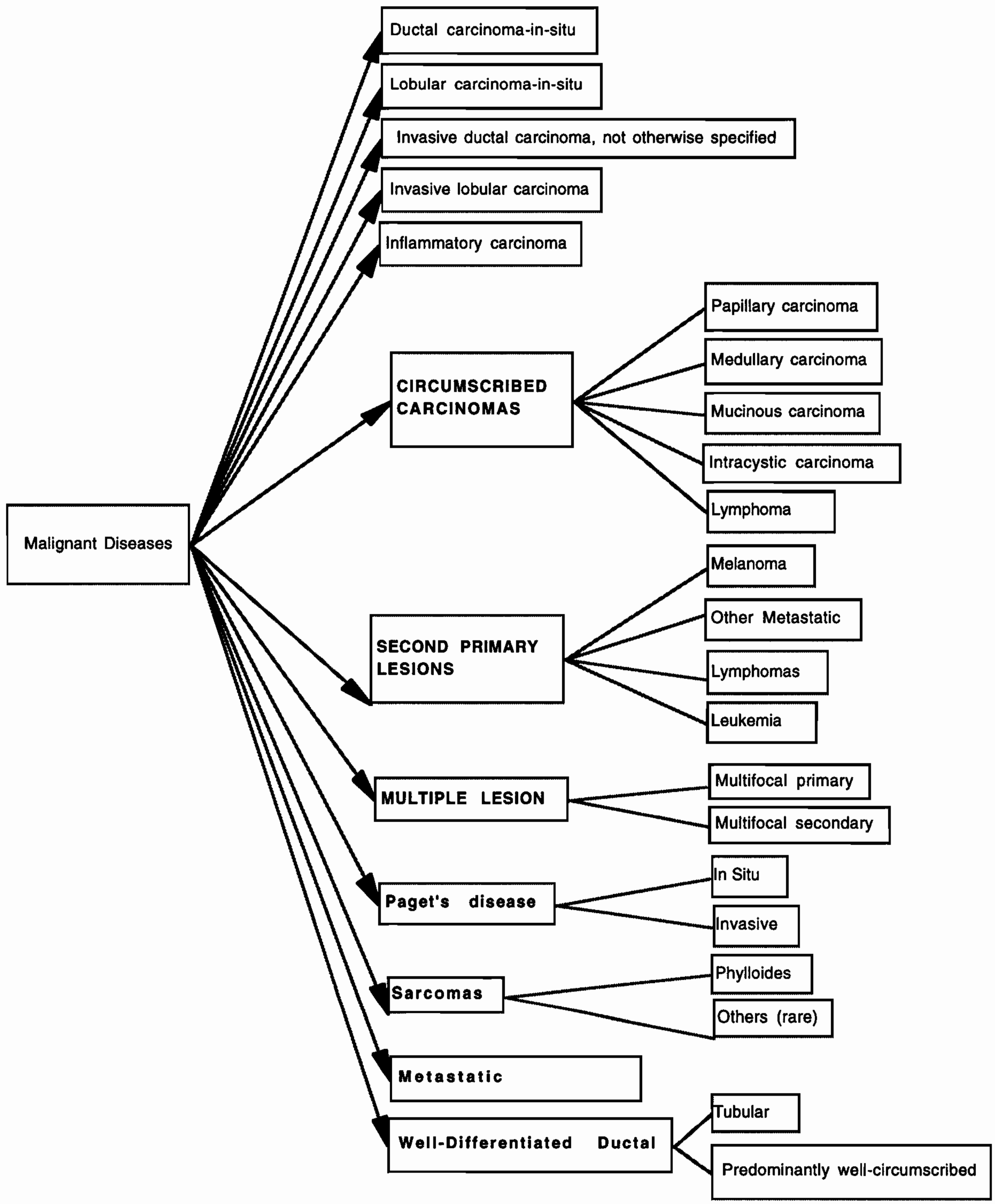



Figure 3.8. A Classification of Malignant Breast Diseases Depicted on Mammograms.

specified), invasive lobular carcinoma, and inflammatory carcinoma. Circumscribed carcinomas include papillary, medullary, mucinous, intracystic, and lymphoma. Second primary lesions include melanoma, other metastatic diseases, lymphomas, and leukemia.

Multiple lesions include multifocal primary and multifocal secondary. Paget's disease is either in-situ or invasive. Sarcomas include either phylloides or other rare types. There are also metastatic malignant diseases. Well-differentiated ductal carcinomas include tubular or predominantly well-circumscribed types.

The third part of the content analysis delineated the further investigations typically made in cases exhibiting suspicious findings or malignancy (see Figure 3.9). Typically, the clinical history and set of mammograms are sufficient for the radiologist to derive a diagnosis meaning that no further examination is required. The various further investigations used by radiologists to collect subsequent data allow them to follow-up or manage a case. In the case of malignancy or suspicious findings, a radiologist may ask for a follow-up which either entails a clinical correlation or performing subsequent mammograms. In the case of clinical correlation, the radiologist may request a repeat physical examination or additional relevant information such as the patient's history of trauma and/or the chronicity of the event.

Mammographically, a follow-up would include one of the following (a) comparing existing mammograms to the patient's previous mammograms, (b) performing a second set of mammograms in six months, (c) performing a second set of mammograms in one year, or (d) requesting additional views (such as a magnified view, spot compression, cleavage view, Cleopatra view, etc.). Ultrasound may also be performed to determine if a density is a cyst or a mass. If the mass is detected under ultrasound, the radiologist or surgeon may conduct either a needle (core) or excisional biopsy. Finally, a radiologist may decide to consult with his/her colleagues regarding the diagnosis of a set of mammograms.

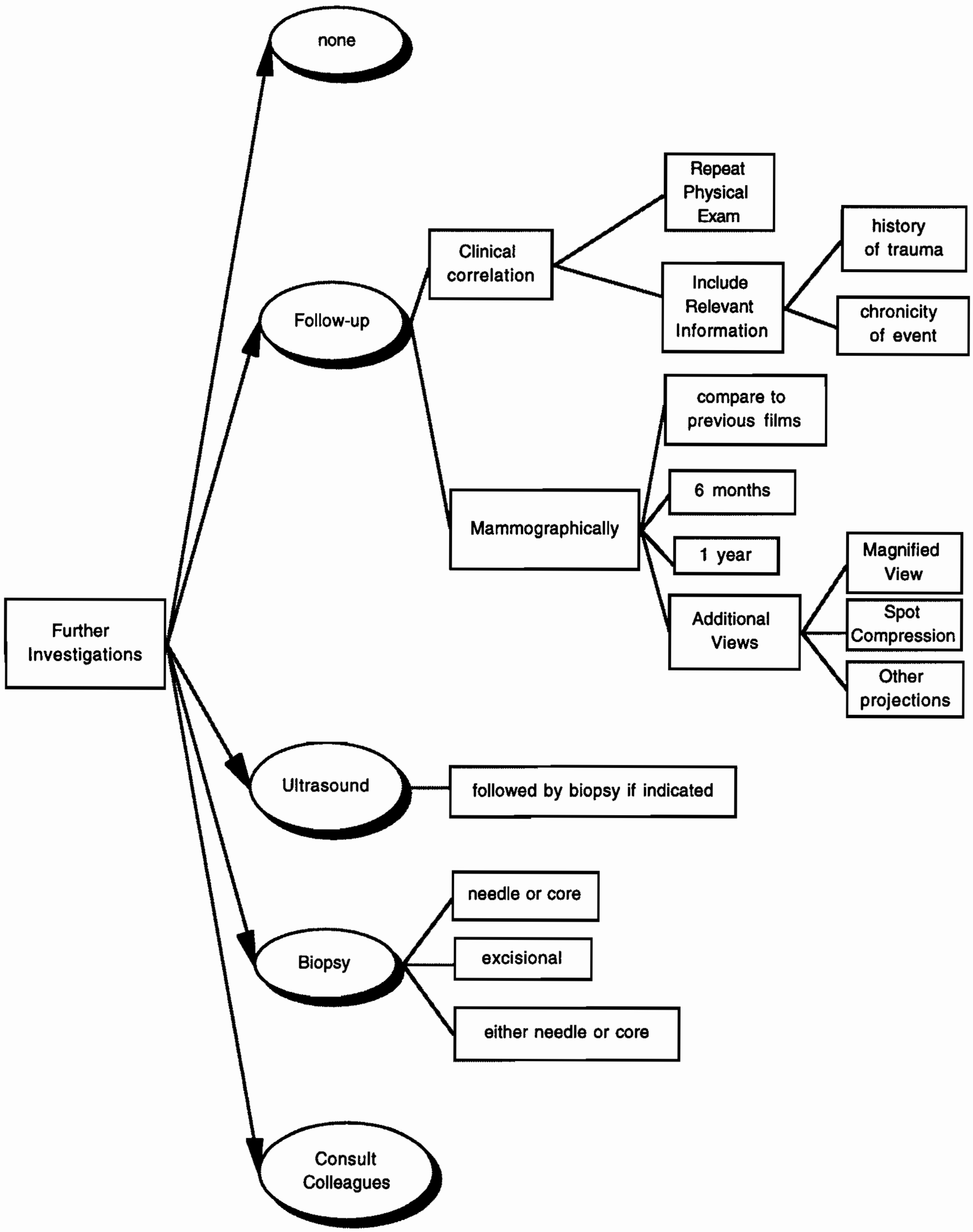



Figure 3.9. A List of Further Investigations Used by Radiologists to Either Follow-up or Manage Breast Diseases.

The fourth part of the content analysis delineated the mammographic features of malignant breast diseases (see Figure 3.10). The mammographic features of malignant breast diseases are sub-divided into major and minor signs. The first major sign is a stellate

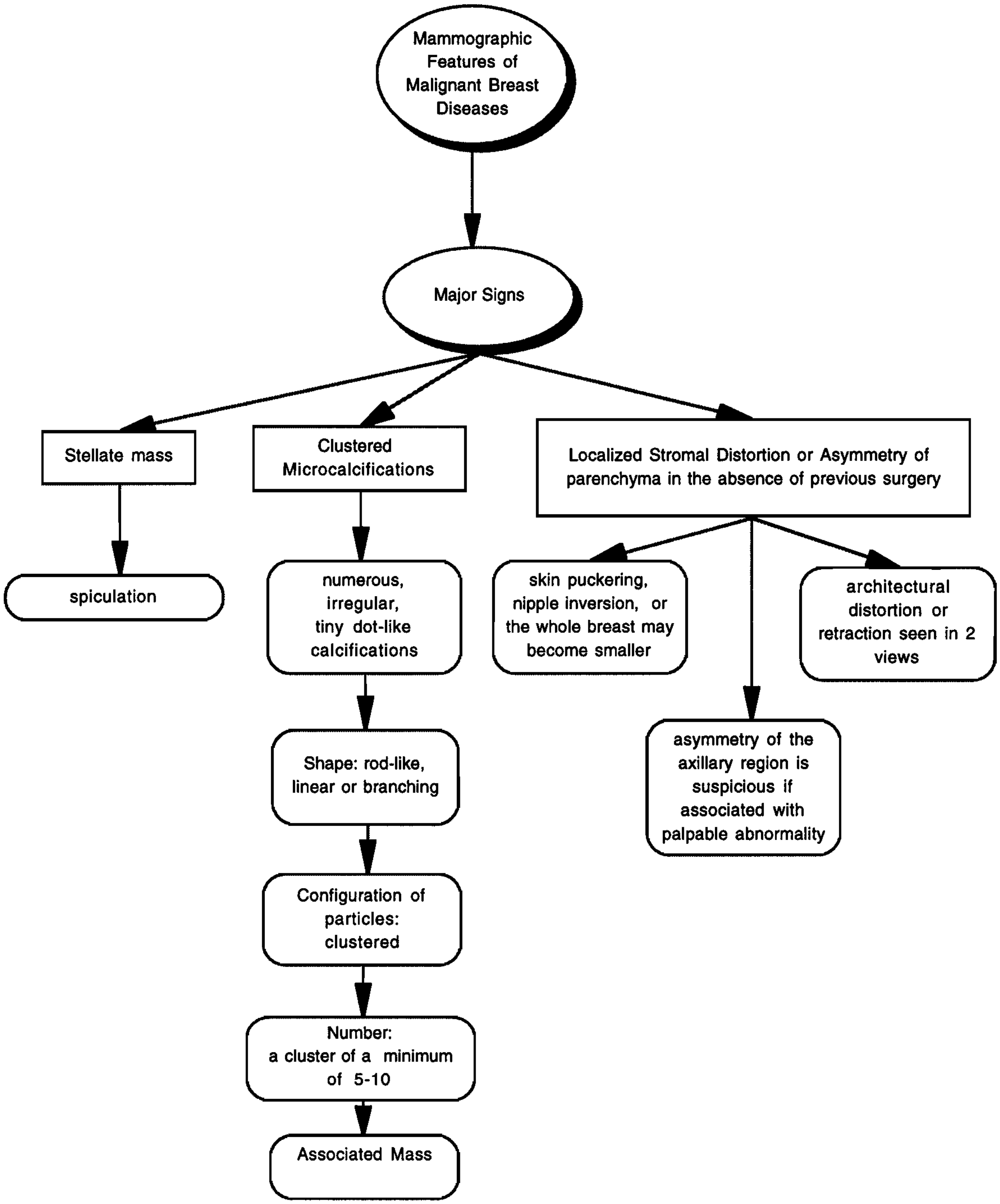



Figure 3.10. A Categorization of the Major Mammographic Features of Breast Diseases.

mass with spiculated margins, and the second is a cluster of numerous, irregular, tiny dot-like microcalcifications. Also included are clusters of a minimum of five to ten other types of microcalcifications (rod-shaped, linear or branching) associated with a dominant mass.

The third major sign is localized stromal distortion or asymmetry of the breast parenchyma in the absence of previous surgery. In this case, there may be skin puckering, nipple inversion, or the whole breast may become smaller. Also, suspicion for malignancy may be indicated by the presence of asymmetry of the axillary region which is associated with a palpable abnormality. Architectural distortion or retraction that is visible on two mammographic views is also a sign of malignancy. The minor mammographic signs of malignancy include associated skin or nipple changes along with skin thickening or retraction (see Figure 3.11). Further, the nipple may be enlarged, ulcerated, calcified or producing discharge. There may also be changes in vascularization, asymmetry in the ductal pattern, or enlarged round and dense axillary lymph nodes.

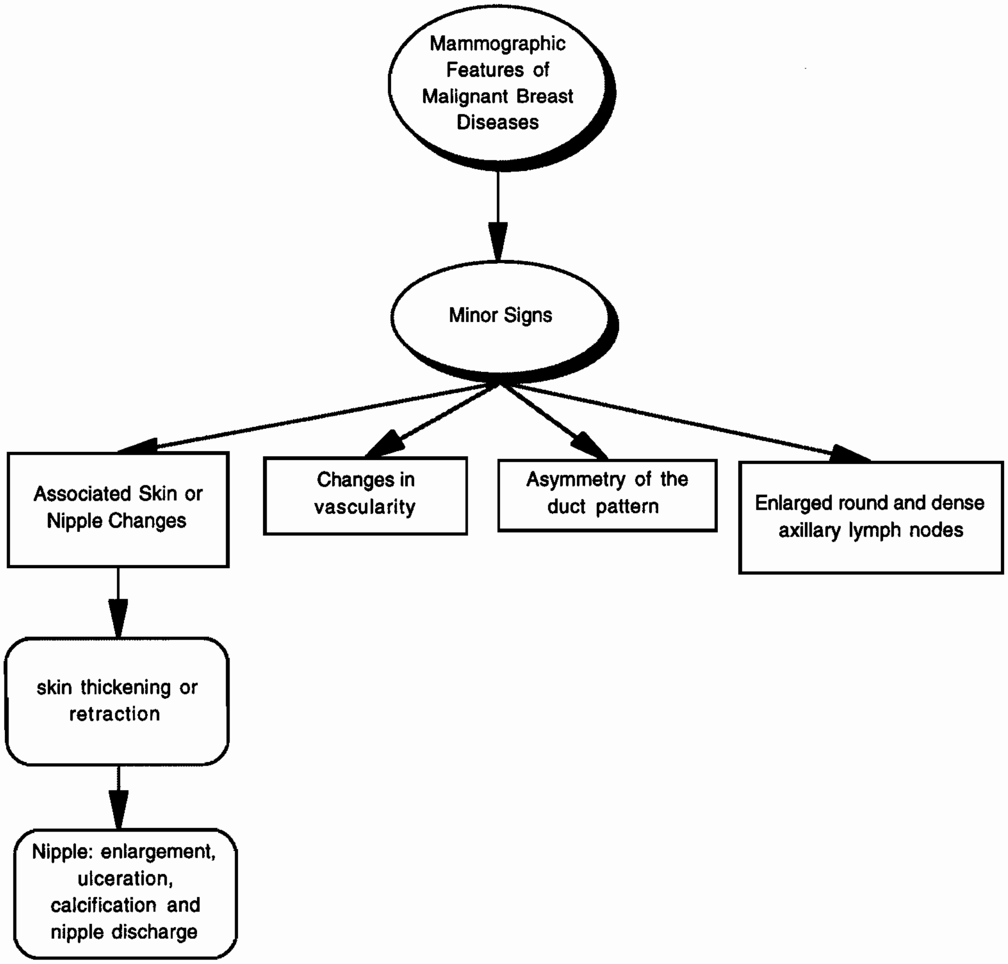



Figure 3.11. A Categorization of the Minor Mammographic Features of Breast Diseases

## Experimental Procedure

Data collection sessions were conducted at several teaching hospitals belonging to McGill University. The location for each data collection session depended on the physical site where each resident was on rotation and the location of each staff radiologist's office. The typical data collection session took place in an office or a conference room equipped with a mounted viewbox suitable for viewing a set of mammograms either in a single row or multiple rows. During data collection sessions certain environmental conditions (such as low lighting, reduced ambient noise, and minimized distractions) were maintained as far as possible in accordance with the standards set by the Canadian Association of Radiologists (CAR).

Participants were tested individually. Prior to the experimental sessions, each participant was greeted by the experimenter and asked to read and sign the consent form (see Appendix A) and complete the brief questionnaire (see Appendix B). The experimenter provided the participant with a one-page hand-out of instructions for the diagnostic task (see Table 3.1). The experimenter placed the materials in front of the participant, including the practice case, the ten experimental cases, the magnifying glass to inspect the mammograms, and the permanent marker to use for pointing (to the mammographic findings). The experimenter presented each participant first with the practice case and subsequently with the ten cases. Each case was comprised of an envelope containing a type-written clinical history and a set of mammograms. For each case, the experimental procedure involved having the participant: (a) read the clinical history, (b) display the mammogram set on a view-box, (c) point to the mammographic findings and observations, (d) provide a diagnosis (or a set of differential diagnoses), and (e) discuss subsequent further investigations (if necessary). The participant was instructed to "think out loud" throughout the entire diagnostic process.

The experimenter sat on a chair about two feet away either to the left or right of the participant, depending on whether the participant was left or right-handed. The experimenter then began to record the session using both video and audio equipment. During each session, the video camera was focused on the viewbox so that the mammograms and the participant's "pointing behavior" was captured on video. The experimenter then instructed the participant to begin the diagnostic task. The experimental procedure was repeated for each participant until he/she diagnosed all ten cases under the two experimental conditions (5 authentic and 5 augmented). No time constraints were imposed.

The ten cases were counterbalanced across experimental conditions and participants by using the 10X10 complete orthogonal Latin Squares set from Fisher and Yates (1963, p. 88) to avoid practice and order effects. Table 3.3 illustrates the case assignment used for each of the twenty participants in the study. The numbers 1 through 10 along the second row of the table represent the assignment of cases. Also, in the second column, the staff

Table 3.3

<u>Table Illustrating the Case Assignment for each Participant in the Study</u>

| **Experience Level** | **Participant Number** | **Case Assignment** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **Staff Radiologists** | **S 1** | 1 | **10** | 2 | **8** | 3 | **9** | 4 | 5 | **6** | **7** |
| | **S 2** | 9 | 7 | **4** | **2** | 10 | **1** | 6 | **3** | 8 | **5** |
| | **S 3** | 3 | **9** | **8** | 1 | **10** | 4 | **7** | **6** | 2 | 5 |
| | **S 4** | 9 | 7 | **4** | **2** | 10 | **1** | 6 | **3** | 8 | **5** |
| | **S 5** | 5 | 3 | **6** | 4 | **8** | 2 | **10** | **9** | **7** | 1 |
| | **S 6** | **5** | 10 | 7 | **4** | 9 | **3** | 8 | 6 | **2** | **1** |
| | **S 7** | **7** | 5 | **9** | **10** | **6** | 3 | 2 | 1 | 4 | **8** |
| | **S 8** | **3** | 9 | 8 | 6 | 7 | 10 | **5** | **2** | **1** | **4** |
| | **S 9** | **9** | **7** | **10** | **6** | 4 | 5 | 1 | 3 | **8** | 2 |
| | **S 10** | **1** | **5** | 7 | 8 | 6 | **4** | **3** | 9 | 10 | **2** |
| **Radiology Residents** | **S11** | 1 | **10** | 2 | **8** | 3 | **9** | 4 | 5 | **6** | **7** |
| | **S12** | 9 | 7 | **4** | **2** | 10 | **1** | 6 | **3** | 8 | **5** |
| | **S13** | 3 | **9** | **8** | 1 | **10** | 4 | **7** | **6** | 2 | 5 |
| | **S14** | 9 | 7 | **4** | **2** | 10 | **1** | 6 | **3** | 8 | **5** |
| | **S15** | 5 | 3 | **6** | 4 | **8** | 2 | **10** | **9** | **7** | 1 |
| | **S16** | **5** | 10 | 7 | **4** | 9 | **3** | 8 | 6 | **2** | **1** |
| | **S17** | **7** | 5 | **9** | **10** | **6** | 3 | 2 | 1 | 4 | **8** |
| | **S18** | **3** | 9 | 8 | 6 | 7 | 10 | **5** | **2** | **1** | **4** |
| | **S19** | **9** | **7** | **10** | **6** | 4 | 5 | 1 | 3 | **8** | 2 |
| | **S20** | **1** | **5** | 7 | 8 | 6 | **4** | **3** | 9 | 10 | **2** |

radiologists are represented as S1 to S10 and the radiology residents are represented as S11 to S20. For each participant, a number (1 to 10) denotes the case number, and the typeface (bold or normal) indicates the experimental condition (normal typeface indicates authentic experimental condition and bold indicates augmented experimental condition)[4].

---

[4] For example, the case assignment for participant #1 (S1) were as follows: (1) case 1 under the authentic condition, (2) case 10 under the augmented condition, (3) case 2 under the authentic condition, (4) case 8 under the augmented condition, (5) case 3 under the authentic condition, (6) case 9 under the augmented condition, (7) case 4 under the authentic condition, (8) case 5 under the authentic condition, (9) case 6 under the augmented condition, and (10) case 7 under the augmented condition. To summarize, each participant diagnosed a total of ten breast disease cases: five under the authentic condition and five under the augmented condition.

**The Treatment of the Think Aloud Verbalizations and Video Data**

The raw data collected for this study consisted of audio and video tape recordings from 20 participants diagnosing ten cases (recorded on 20 audio and video tapes). During the first phase of the data analysis, a graduate student with extensive experience in transcribing problem solving protocols was hired to transcribe the audio tapes. The tapes were transcribed according to the standard transcription conventions (Bracewell & Breuleux, 1993) presented in Table 3.4. These conventions ensure (as much as possible) that the accuracy of lexical and syntactic structures is maintained (Bracewell & Breuleux, 1993). A text file comprised of ten separate segments corresponding to the ten cases was created for each participant.

The second phase consisted of verification of the accuracy of the transcriptions by the experimenter. This process involved comparing each text file with the video tape of the participant. The experimenter used a Macintosh® computer with the text file open while running the tapes using Apple Video Monitor©. During this process the necessary changes to the text file were made. The majority of transcription errors consisted of misspelled medical terms which were flagged by the transcriber using double parentheses (see Table 3.4).

During the third phase, the experimenter segmented the transcripts and inserted corresponding video screen shots. Screen shots of a participant's verbal and non-verbal problem solving behavior were captured and inserted in the text file to facilitate subsequent coding. This was done for segments comprised of elided phrases. Ellipsis is one the most pervasive phenomena encountered in the English language and involves the omission of elements that would be expected in the full syntactically correct form of an utterance or phrase (Winograd, 1983). For example, when asking for “2 tickets” at the theater, the teller knows that “2” means two tickets and therefore the word tickets is omitted. Other segments for which screen shots were captured included a participant's pointing behavior and mammogram comparison. The next section presents a detailed description of the coding scheme and its application to a subset (50) of the 200 transcribed and segmented protocols.

Table 3.4

Transcription Conventions

| Mark | Significance | Example |
|---|---|---|
| . | Period: Used with utterances having normal (falling) intonation | It could represent a cyst of fibroadenoma. |
| ? | Question mark: Used with interrogative (rising) intonation | What would I do next? |
| ! | Exclamation point: Used with exclamatory (sharp rise at end of word) intonation | However, the presence of retraction is somewhat worrisome! |
| !xxx! | Bracketing exclamation points: Used with utterances that are stressed with loudness or highly pitched | This mass is suspicious. |
| :: | Colon(s): Used to indicate prolongation of syllables | The microcalcifications are spi::culated |
| ... | Multiple periods: Used to indicate pause by speaker | Each period counts 1 second |
| xxx-, | Hyphen plus comma: Used for utterances that are revised in the course of production | There is a round-, ill-defined mass |
| ((xxx)) | Double parentheses: Used where transcription may not be accurate | This represents could be either ((medelary)) or infiltrating carcinoma. |
| "xxx" | Bracketing asterisks: Used for all comments made by the transcriber | "verify next sentence for spelling of medical vocabulary." philoidis tumor |

**Coding Scheme.** The goal of the verbal protocol data collection was to obtain in-depth, detailed data from the participants while they were interpreting and diagnosing the breast disease cases. A coding scheme was constructed based on the content analysis described in a previous section, theoretical articles, methodological articles, and the results of previous studies in various relevant areas such as cognitive science (Ericsson & Simon, 1993;

Newell & Simon, 1972; Simon & Kaplan, 1989), medical cognition (Barosi, Magnani & Stefanelli, 1993; Evans & Gadd, 1989; Patel & Groen, 1986; Hassebrock & Prietula, 1992; Ramoni et al., 1992), technical writing (Bracewell & Breuleux, 1993; Breuleux, 1991; Smagorinsky, 1993), discourse processing (Frederiksen, 1975; 1986), syntactical analysis (Winograd, 1983), chest radiography radiology (Friedman, Cimino & Johnson, 1994; Friedman et al., 1995; Lesgold et al., 1981; 1988; Rogers, 1992; 1996), and artificial neural networks for mammogram interpretation (Lo, Baker, Kornguth, Iglehart & Floyd, 1997; Wu, Giger, Doi, Vyborny, Schmidt & Metz, 1993).

Fifty of the 200 protocols were used to refine an initial coding scheme into a more comprehensive one consisting of three major categories. The major categories included, *knowledge states*, *problem solving operators*, and *control processes* (Anderson, 1993b; Newell & Simon, 1972). Knowledge states in this domain were coded as radiological observations, radiological findings, and diagnoses. Problem solving operators were clustered around 11 classes and comprised a total of 30 operators. Control processes were comprised of diagnostic planning, goal verbalizations, and meta-reasoning. Each of the major categories is described below.

**Knowledge States.** Knowledge states represented the hierarchical nature of medical knowledge in breast diseases and mammography including radiological observations, radiological findings, and diagnoses (which include category hypotheses used in the classification of breast diseases) (Evans & Gadd, 1989). Each of the three types of knowledge states identified in the verbal protocols is described below. For each, the results presented in Figure 3.2 were used as a basis for coding protocols and each was numbered sequentially (e.g., "Radiological Observation #1, "Radiological Observation #2", and "Radiological Observation #3", etc.).g1

**Radiological observations** are units of information that are recognized as potentially relevant in the problem solving context, but do not constitute clinically useful facts (Evans & Gadd, 1989). These represent common findings encountered in both clinical histories and mammograms. For example, the fact that a set of mammograms presents *prominent fibroglandular tissue* is of no particular clinical significance in solving the case. Other examples of radiological observations may include: small axillary lymph nodes, several round benign-looking nodules, glandular tissue with nodular appearance, and asymmetrical breasts. In the think aloud verbal protocols, radiological observations were identified based on the results of the consulting radiologist's diagnosis of each case.

**Radiological findings** are composed of critical cues with particular clinical significance (Evans & Gadd, 1989). These include common findings encountered in both clinical histories and mammograms. For example, the fact that a set of mammograms presents *a large partially well-circumscribed mass* seen on two views is of critical importance since it typically indicates the presence of a malignant tumor. Other examples of radiological findings include: several small nodules, large well-circumscribed nodules, an area of retraction with some disorganization, round calcifications, a large well-circumscribed mass consisting of fat and containing encapsulated islands of fibroglandular tissue, a density associated with a loss of organization, architectural distortion and some retraction, a nodule in the center of the right breast, pleomorphic calcifications in a scar, and a large spiculated lesion. In the think aloud verbal protocols, a radiological finding was identified based on the results of the consulting radiologist's diagnosis of each case.

**Diagnoses** represent medical knowledge at different levels of abstraction. Pre-diagnostic labels and definitive diagnoses serve both to establish a context in which observations and findings are interpreted and provide a basis for anticipating and searching for confirming and disconfirming findings. The reason they were grouped together is that unlike many other radiology sub-specialties (e.g., chest radiography), mammography is comprised of fewer disease types and a limited number of mammographical manifestations. For example, a diagnosis of mucin-producing carcinoma subsumes the following levels of abstraction (from lowest to highest): a malignant process, neoplastic lesion, primary breast cancer, ductal carcinoma, and invasive (infiltrating) adenocarcinoma with productive fibrosis (Peters, Voegeli & Scanlan, 1989). Other examples include a pathological description such as *malignant process*, or categorical descriptions such as *neoplastic lesion*.

Two aspects of diagnoses need further explanantion. First, the level of abstraction as indicated by a participant's hypothesis is not entirely related to medical experience. In some cases the manifestation of mammographic findings and/or observations may only allow a participant to propose a lower level hypothesis such as suspicious for malignancy as opposed to stating that a set of mammograms show papillary carcinoma. Secondly, and most importantly from the viewpoint of mammography, the diagnoses of "benign" and "malignant" at different levels of abstraction typically lead to the same follow-up actions. For example, in the case of benign diseases, classifying a set of mammograms as normal or stating a diagnosis of calcium in small cysts does not lead to further examinations. Similarly, with malignant diseases, classifying a set of mammograms as suspicious or stating a diagnosis of inflammatory carcinoma leads to subsequent examinations.

**Problem Solving Operators.** The second category of representation in the scheme identified the *problem solving operators* used to generate or instantiate states of radiological knowledge. A problem solving operator is an inferred cognitive process which modifies, adds, and/or eliminates existing or currently active knowledge states and produces new, active knowledge states (Anderson, 1993a; Anderson, 1993b; Newell & Simon, 1972). Twelve basic types of operators were identified to characterize distinct segments of problem solving behavior. Each basic operator was further analyzed in order to provide a more detailed representation. The operators are presented in Figure 3.12.

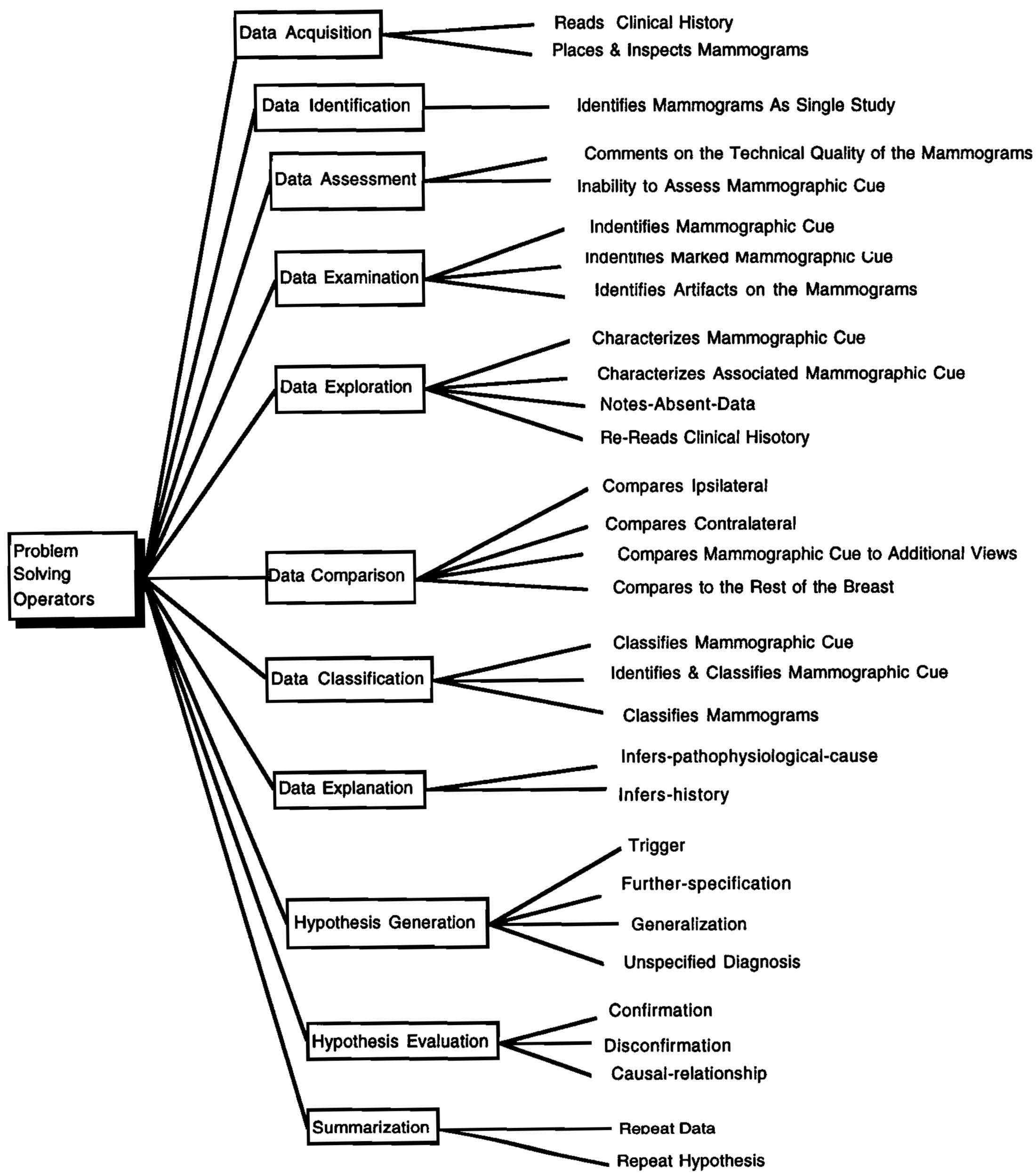



Figure 3.12. List of Problem Solving Operators Used by the Participants.

The problem solving operators used in this study are analogous to the goal-directed medical problem-solving behavior of physicians (Barrows & Tamblyn, 1980; Elstein et al., 1978; Hassebrock & Prietula, 1992: Kassirer & Gorry, 1978; Patel & Groen, 1986; Rogers; 1992; 1996). The problem solving operators reflect the knowledge and problem solving behaviors required to successfully complete the diagnostic task. Furthermore, given the visual nature of radiological diagnosis, most of the conceptual operations involve either **actions which are not concurrently accompanied by verbalizations** (e.g., scanning mammograms) or **actions which are concurrently accompanied by verbalizations** (e.g., using a magnifying glass to determined the morphology of microcalcifications while simultaneously verbally characterizing them). For example, the presentation of operators as discrete independent entities facilitates the characterization of these behaviors at a higher level. A detailed account of each problem solving operator is presented in Appendix N.

**Control Processes.** The third category of representation in the coding scheme consisted of *control processes*, generated by the participants. These included goal verbalization (the use of the future tense to indicate an intended action), diagnostic planning (the planning of subsequent examinations and their possible interpretations), and meta-reasoning (a participant conducts a self-evaluation of the quality of the evolving diagnostic strategy).

**Goals** can consist either of operations that are possible, postponed, intended, or of states that are expected to be obtained (Breuleux, 1991). Goals can be identified in that they have no reference to already existing states. According to Breuleux (1991) the characteristics of goal statements include (a) temporal antecedence over action (temporality marked as future), (b) conditionality (disjunctive relations between goals, interrogative truth value, or marked modality), (c) volition (volitional lexical identifier), and (d) abstraction (fragmentary, incomplete, or illegal actions). In coding each protocol, goals were identified based on the results of research by Breuleux (1991), Bracewell & Breuelux (1993) and Frederiksen (1975; 1986). For example, if a participant mentioned "*I will inspect the left breast for microcalcifications*" while solving a case, it would be coded as "GOAL [inspect left breast for microcalcifications]".

**Diagnostic planning** statements included verbalizations that indicated a need to perform subsequent radiological or medical examinations to acquire additional information, constrain a set of differential diagnoses, or clarify a working diagnostic hypothesis. A plan involves coordinating the selection of operators. Its execution involves making behavior conditional on the state of the problem and a hierarchy of goals and sub-goals (Simon, 1975). This involves obtaining more clinical information or using invasive methods (e.g., biopsy) which allow the radiologist to follow-up or manage a case. In either case, the objective is to constrain the number of differential diagnoses or further specify a working hypothesis at a higher level of abstraction. In coding the protocols, the first recommendation made by a participant was coded as "Recommendation #1". In cases where a participant mentioned several recommendations while solving a case, each one was numbered sequentially in ascending numerical order based on order of appearance in the protocol (e.g., "Recommendation #1, "Recommendation #2", etc.).

**Meta-reasoning** included any episodes in which the participant evaluated the diagnostic reasoning process itself. This included statements of plans, and/or self-evaluations of the quality of the evolving diagnostic strategy. A participant used this operator to criticize or critique specific diagnostic hypotheses or general diagnostic conclusions. For example, *"this one I can't even guess at, because I don't know what I'm looking at*". A second use of this operator was to recall case-specific information (including disease incidence or probability given the clinical history and mammographic manifestations) regarding a previous patient or specific clinical encounter. For example, "*I have never seen one of these in my practice....these are rare manifestations of a possible benign condition*".

## Application of the Coding Scheme

The coding scheme was applied to each protocol in the segment as described based on the identification of knowledge states and problem solving operators and control processes. However, each segment was also coded based on its semantic relations. This was accomplished by using domain specific semantic relations (obtained from the content analysis and the results of the consulting expert's diagnosis of each case) and Frederiksen's (1975; 1986) theory of propositional representations. Table 3.5 lists the domain-specific semantic relations which included appearance, associated mammographic cue, body location, certainty, change, classification, degree, density, description, distribution, margins, number, quantity, recommendation, region, shape, and size. Table 3.6 lists the semantic relations that were adopted from Frederiksen (1975) including dependency logical relations, derived logical relations, and derived logical relations for quantification. The semantic relations were identified and coded for each segment of a protocol.

Table 3.5

A Description of the Domain-Specific Semantic Relations

| Semantic Relation | Definition | Example |
|---|---|---|
| Appearance | Terms describing the appearance of features related to (micro)calcifications. | "... a cluster of microcalcifications in the left breast"<br><br>*Radiological Observation #1 [microcalcifications]*<br>    *Number [a]*<br>    *Appearance [cluster]*<br>    *BodyLocation [left breast]* |
| Associated Mammographic Component (ASSOC:) | Terms or descriptions of mammographic features that are associated with radiological findings or observations. | "... a lesion with fat around it"<br><br>*Radiological Finding #1 [lesion]*<br>    *Number [a]*<br>*ASSOC:*<br>    *Radiological Mammographic Component [fat around it]* |
| BodyLocation | Terms denoting a well-defined area of the body or body part. | "... a lesion in the right breast"<br><br>*Radiological Finding #1 [lesion]*<br>    *Number [a]*<br>    *BodyLocation [right breast]* |
| Certainty | Terms affecting the certainty of a finding. This class modifies status and change terms in addition to findings. | "... the calcification appears to be in the left breast"<br><br>*Radiological Finding #1 [calcification]*<br>    *Certainty [appears to be]*<br>    *BodyLocation [left breast]* |
| Change | Terms denoting a change in findings or observations where the change is an improvement or worsening of a finding or observation. These don't include terms denoting the beginning or end of a change. | "... there is increased fatty tissue in the left breast"<br><br>*Radiological Observation #1 [breast tissue]*<br>    *Change [increased]*<br>    *Density [fatty]*<br>    *BodyLocation [left breast]* |

| Semantic Relation | Definition | Example |
|---|---|---|
| Classification | Medical terms representing the hierarchical nature of medical classification and diagnosis. Terms classifying a mammogram, a set of mammograms, a mammographic finding, a mammographic observation, or a set of mammographic findings and/or observations at several levels of abstraction. | "... the mass is suspicious for malignancy"<br><br>*Radiological Finding #1 [mass]*<br>*Classification [suspicious for malignancy]* |
| Degree | Terms denoting the severity of a finding. These terms can also modify change, certainty, and other degree terms. | "... the right breast shows a moderate amount of fibroglandular tissue"<br><br>*Radiological Observation #1 [fibroglandular tissue]*<br>*BodyLocation [right breast]*<br>*Degree [moderate amount]* |
| Density | Terms characterizing the density of a mammographic feature. | "... the left breast is fatty"<br><br>*Radiological Observation #1 [left breast]*<br>*Density [fatty]* |
| Description | Terms characterizing the description of calcifications. | "... there are five microcalcifications in the inner outer quadrant"<br><br>*Radiological Observation #1 [microcalcifications]*<br>*Number [5]*<br>*Region [inner outer quadrant]* |
| Distribution | Terms characterizing the distribution of calcifications. | "... a cluster of calcifications in the right breast"<br><br>*Radiological Finding #1 [calcifications]*<br>*Distribution [cluster]*<br>*BodyLocation [right breast]* |
| Margins | Terms characterizing the margins of breast masses. | "... the lesion has well-defined borders"<br><br>*Radiological Finding #1 [lesion]*<br>*Margins [well-defined]* |

| Semantic Relation | Definition | Example |
|---|---|---|
| Number | Terms denoting a count for concepts. | "... there is one calcification in the left breast"<br><br>*Radiological Finding #1 [calcification]*<br>*Number [one]* |
| Quantity | Terms representing non-numeric quantitative information. | "... there are multiple lesions located in the left breast"<br><br>*Radiological Finding #1 [lesions]*<br>*Quantity [multiple]*<br>*BodyLocation [left breast]* |
| Recommendation | Terms denoting subsequent radiological or medical recommendations. | "... this mass needs to be excised"<br><br>*Radiological #1 [mass]*<br>*Recommendation [excise mass]* |
| Region | Terms denoting relative locations within a body location. | "... the lesion is in the upper outer quadrant of the right breast"<br><br>*Radiological Observation #1 [lesion]*<br>*Number [the]*<br>*Region [upper outer quadrant]*<br>*BodyLocation [right breast]* |
| Shape | Terms characterizing the shape of a mammographic feature. | "... the mass is oval"<br><br>*Radiological finding #1 [mass]*<br>*Shape [oval]* |
| Size | Terms denoting the size of radiological findings or observations. | "... a 2 cm lesion in the retroaereolar region"<br><br>*Radiological Finding #1 [lesion]*<br>*Number [a]*<br>*Size [2 cm]*<br>*Region [retroaereolar region]* |

Table 3.6

A List of the Semantic Relations

| Semantic Relation | Definition | Example |
|---|---|---|
| AND | Dependency logical relation: specifies union of concepts | "... a nodule in the left breast and another one in the right breast"<br><br>*Radiological Observation #1 [nodule]*<br>*Number [a]*<br>*BodyLocation [left breast]*<br>AND<br>*Radiological Observation #2 [nodule]*<br>*Number [one]*<br>*BodyLocation [right breast]* |
| CAUSAL | Dependency logical relation: reduced set of implications, never contraposes | "... the left breast is enlarged due to dysplasia"<br><br>*Radiological Observation #2 [dysplasia]*<br>CAU:<br>*Radiological Observation #1 [breast tissue]*<br>*BodyLocation [left breast]*<br>*Change [enlarged]* |
| CONDITION | Dependency logical relation: reduced set of implications, may not contrapose | "... the two mass lesions in the left breast represent benign lesions"<br><br>*Radiological Finding #1 [mass lesion]*<br>*BodyLocation [left breast]*<br>AND<br>*Radiological Finding #2 [mass lesion]*<br>*BodyLocation [left breast]*<br>COND:<br>*Classification [benign lesions]* |
| EQUIVALENCE | Derived algebraic logical relations: specifies concepts having the same value of a property | "... there is an area of increased density which may just be composite shadows"<br><br>*Radiological Finding #1 [breast tissue density]*<br>*Distribution [an area]*<br>*Change [increased]*<br>EQUIVALENCE<br>*[composite shadows]* |

| Semantic Relations | Definition | Example |
|---|---|---|
| IF...THEN... | Dependency logical relation: material implication, relation contraposes (e.g., valid inferences are if p then q, and if not-p then not-q) | "... if it's solid then the patient should have a biopsy"<br><br>IF *[solid]* THEN *[perform biopsy]* |
| OR-ALTERNATIVE | Dependency logical relation: specifies one or more concepts from among a set of concepts | "... not knowing what the diagnosis is, I would consult my colleagues or a text"<br><br>*Recommendation #1 [consult colleagues]*<br>OR-ALTERNATIVE<br>*Recommendation #2 [consult textbook]* |
| OR-EXCLUSIVE | Dependency logical relation: specifies one concept from among a set of concepts | "... it could represent a cyst or a fibroadenoma"<br><br>*Classification*<br>OR-EXCLUSIVE *[cyst] [fibroadenoma]* |
| ORDER | Derived algebraic logical relations: specifies concepts that differ in the value of a property, and orders the concepts with respect to the value | "... we could get more clinical information followed by an ultrasound"<br><br>*Radiological Finding #1 [mass]*<br>ORDER<br>*Radiological Recommendation #1 [clinical follow-up]*<br>*Radiological Recommendation #2 [ultrasound]* |
| PROXIMITY | Derived algebraic logical relations: specifies concepts with similar values of a property | "... the other abnormality is near the first lesion"<br><br>Radiological Finding #2 *[abnormality]*<br>PROXIMITY<br>Radiological Finding #1 *[lesion]* |

An example of the application of the coding scheme to a transcribed protocol from a staff radiologist is provided in Tables 3.7 and 3.8. Table 3.7 presents a transcribed and segmented protocol from a staff radiologist (S5) solving a malignant case (Case 1A) under the authentic condition. The transcription includes notes (in italics) regarding the participant's nonverbal behavior such as placing mammograms on a viewbox and the duration of this activity (22 seconds for this case). The case was then segmented according to the steps outlined in the mammogram interpretation model. The duration of the entire diagnostic episode (one minute and 44 seconds or 104 seconds in this case) was also recorded. Table 3.8 presents the same protocol after application of the coding scheme. The protocol was broken down into 10 segments. For each segment, the normal typeface illustrates the participant's verbalizations while the italicized text represents the coded verbalization. More specifically, the first line of each coded segment specifies the problem solving operator used in the segment. For example, segment three involves the use of the "*data examination*" operator. In this particular case the participant used a specific data examination operation called "*identifies mammographic cue*". The use of this operator led to the identification of a radiological finding (a mass), with a particular size (large), type of border (well-defined), region (retroaereolar), and location (right breast). Other operators present in this protocol include data exploration, data classification, hypothesis generation, and hypothesis evaluation. The participant used one control process (diagnostic plan) seen in segment nine where he ordered a biopsy. The participant's plan included the patient having a stereotaxic biopsy to confirm the benign nature of the lesion if the initial biopsy results were to indicate that it was solid (not a cyst).

Table 3.7

A Transcribed and Segmented Protocol from a Staff Radiologist (S5) Solving a Malignant Case Under the Authentic Condition (Case 1)

**Case 1A**
*Places Mammograms on Viewbox (00:22)*
1. Case 1A. A 60 year-old woman who presented with a mass in the right breast.
2. Ah both films are somewhat dark.
3. However, the obvious finding is a large, well-defined retroaereolar mass in the right breast
4. measuring 6 by 4.5 centimeters in diameter.
5. The edges are very well-defined
6. and this is most likely a benign lesion.
7. It could represent a cyst or a fibroadenoma.
8. I do not believe that this could represent a colloid carcinoma or a medullary carcinoma.
9. If old films are available the obvious next step would be to do an ultrasound to alert a cyst if it's not a cyst then the patient could be followed or have a stereotaxic biopsies just to confirm the benign nature of this lesion.
10. It could represent a fibroadenoma.
(01:44)

Table 3.8

A Coded Protocol from a Staff Radiologist (S5) Solving a Malignant Case Under the Authentic Condition (Case 1)

| S5 - 1A | | |
|---|---|---|
| **Segment #** | **Transcript** | **Time** |
| 1 | Case 1A. A 60 year-old woman who presented with a mass in the right breast.<br>*Data Acquisition ⇒ Reads Clinical History* | |
| 2 | Ah both films are somewhat dark.<br>*Data Assessment ⇒ Comments on the Technical Quality of Mammograms* | |
| 3 | However, the obvious finding is a large, well-defined retroaereolar mass in the right breast<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Size [large]*<br>*Borders [well-defined]*<br>*Region [retroaereolar]*<br>*BodyLocation [right breast]* | 00:22 |
| 4 | measuring 6 by 4.5 centimeters in diameter.<br>*Data Exploration ⇒ Measures Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Size [6 by 4.5 cm in diameter]* | |
| 5 | The edges are very well-defined<br>*Data Exploration ⇒ Characterizes Borders of Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Degree [very]*<br>*Borders [well-defined]* | |
| 6 | and this is most likely a benign lesion.<br>*Hypothesis Generation ⇒ Trigger*<br>*Radiological Finding #1 [mass]*<br>*Certainty [most likely]*<br>*Classification [benign]* | |
| 7 | It could represent a cyst or a fibroadenoma.<br>*Hypothesis Generation ⇒ Further Specification*<br>*OR [cyst] fibroadenoma]* | |
| 8 | I do not believe that this could represent a colloid carcinoma or a medullary carcinoma.<br>*Hypothesis Evaluation ⇒ Disconfirmation*<br>*Radiological Finding #1 [mass]*<br>*NEG:COND*<br>*OR [colloid carcinoma] [medullary carcinoma]* | |
| 9 | If old films are not available the obvious next step would be to do an ultrasound to alert a cyst if it's not a cyst then the patient could be followed or have a stereotaxic biopsy just to confirm the benign nature of this lesion.<br>*Diagnostic Plan ⇒ Subsequent Radiological Examinations*<br>*Radiological Finding #1 [mass]*<br>*Recommendation #1 [ultrasound]*<br>*IF [no previous available] THEN [perform ultrasound]*<br>*IF [not a cyst] THEN [followed mammographically] OR [stereotaxic biopsy]* | |
| 10 | It could represent a fibroadenoma.<br>*Hypothesis Generation ⇒ Further Specification*<br>*Radiological Finding #1 [mass]*<br>*Certainty [could]*<br>*Classification [fibroadenoma]* | 01:44 |

Inter-rater reliability was established by recruiting a graduate student with experience in the area of breast disease and mammography. The experimenter trained the student to use the coding scheme and she was instructed to independently code 20 protocols (which were selected randomly from the 200 protocols). There was agreement on 183 out of a total of 198 coded segments (20 protocols with approximately 10 segments each) yielding a reliability coefficient rating of .92 (Borg & Gall, 1996). The inconsistencies were resolved through discussion between the researcher and the student.

## Research Questions and Hypotheses

Several research questions and hypotheses were addressed in this study. Each is described below.

1) Identify a cognitive model of diagnostic problem solving in mammography interpretation.
2) Do staff radiologists and radiology residents use different problem solving strategies, operators, and control processes during mammogram interpretation? It was hypothesized that staff radiologists would solve the breast disease cases using a data-driven problem solving approach. In contrast, it was hypothesized that the residents would solve the cases by using either goal-driven or a mixed-strategy (combination of data-driven and goal-driven problem solving strategies).
3) What are the effects of the authentic and augmented experimental conditions on several aspects of the groups' performance, including: (a) number of mammogram findings, (b) number of observations, (c) number of diagnoses, (d) scanning time to construct an initial mental representation based on the clinical history and set of mammograms, (e) reading time required to solve each case, (f) accuracy ratings for diagnosis, (g) accuracy ratings for subsequent examination, and (h) accuracy ratings for overall diagnostic accuracy?

   It was hypothesized that in comparison to radiology residents, staff radiologists would be faster in scanning and reading mammograms, and would have better ratings for accuracy of diagnoses, subsequent examinations and overall diagnostic reasoning regardless of experimental condition.

   It was also hypothesized that, compared to the authentic condition, highlighting mammographic findings (in augmented condition) would facilitate residents' ability to identify findings and observations, provide correct diagnoses, decrease scanning and reading time, and increase the accuracy of diagnoses, subsequent examinations, and overall diagnostic accuracy.

4) What are the effects of the authentic and augmented experimental conditions on the frequency and types of errors committed by both.

It was hypothesized that in comparison to radiology residents, staff radiologists would commit less errors. It was also hypothesized that both groups would make less errors when solving cases presented under the augmented condition.

## Research Design

A mixed factorial design (Keppel, 1982) was employed to answer the above mentioned research questions and test the specific hypotheses. It consisted of ten participants nested across two levels of radiological expertise (staff radiologists and radiology residents) crossed with two experimental conditions (authentic and augmented). Two levels of radiological expertise (staff radiologists and radiology residents) comprised the between-participants factor. The experimental conditions (authentic and augmented) comprised the within-participants factor. In the authentic condition, each participant was given a brief clinical history and a set of mammograms. In the augmented condition, mammographic findings were highlighted on a second set of the same cases.

# CHAPTER 4

# RESULTS

This chapter is divided into four sections, each dealing with a particular research question and its corresponding hypotheses. The *first* section presents a cognitive model of diagnostic problem solving in mammogram interpretation based on a the content analysis, consulting radiologist's diagnosis of each case, and analysis of verbal protocols. The *second* section presents in-depth analyses of protocols from several cases in order to exemplify typical staff radiologists' and radiology residents' problem solving strategies. This section attempts to answer the question of whether staff radiologist and radiology residents use different problem solving strategies, operators, and control processes during mammogram interpretation. The *third* section presents frequency analyses, comparisons of descriptive statistics, non-parametric analyses, and inferential statistics. These analyses were performed to investigate whether level of expertise and experimental conditions had an effect on several performance measures (number of mammogram findings, observations, and diagnoses, scanning and reading time, accuracy ratings for diagnosis, subsequent examination, and accuracy ratings for overall diagnostic accuracy). The *fourth* section presents frequency analyses, describes the types of errors committed by participants, and provides results of in-depth analyses of protocols. These analyses were performed to investigate whether level of expertise and experimental conditions had an effect on the number and types of errors committed.

## A Cognitive Model of Mammogram Interpretation

The first objective of this study was to identify a cognitive model of diagnostic problem solving in mammography interpretation. The model specifies seven steps (see Figure 4.1). The first step usually involves reading a clinical history comprised of a letter from the referring physician (e.g., family doctor or surgeon) requesting a mammographic examination. The request is based on the results of a physician's (or surgeon's) examination or it may be a follow-up mammogram in which case the patient not only has a clinical history but also has a previous set of mammograms which can be used for comparison. In a clinical situation a radiologist might not have access to the patient's clinical history and therefore must make a diagnosis from the set of mammograms. The second step involves placing a set of mammograms on the viewbox followed by the identification of individual mammograms in the set. For example, the radiologist may state "I am presented with the left and right CC and MLO views".

The third step involves the visual inspection of the mammograms either with or without the use of a magnifying glass. The fourth step involves the identification of mammographic findings and observations. Depending on their presence, this step usually involves stating the location of these mammographic features. For example, "there is a mass in the retroaereolar region of the right breast". The fifth step involves the characterization of the mammographic findings and observations. Basically, this step involves elaborating on the results of the previous step. For example, "there is a large partly well-circumscribed round mass in the right retroaereolar region". In this example, the radiologist has provided the number, size, characterization of the margins, shape, and anatomical location of the mass. In mammography, this step is typically followed by the search for the same mass in the ipsilateral view (of the same breast). Other times the search is for the findings or observations in the contralateral view (of the opposite breast). For example, "there is dense fibroglandular tissue bilaterally". In other cases, the absence of a feature on the ipsilateral view limits the radiologist's ability to characterize the feature at the "perceptual level" (e.g., "the density is not seen in the CC view"). However, in most cases, the characterization of individual features provides the radiologist with enough

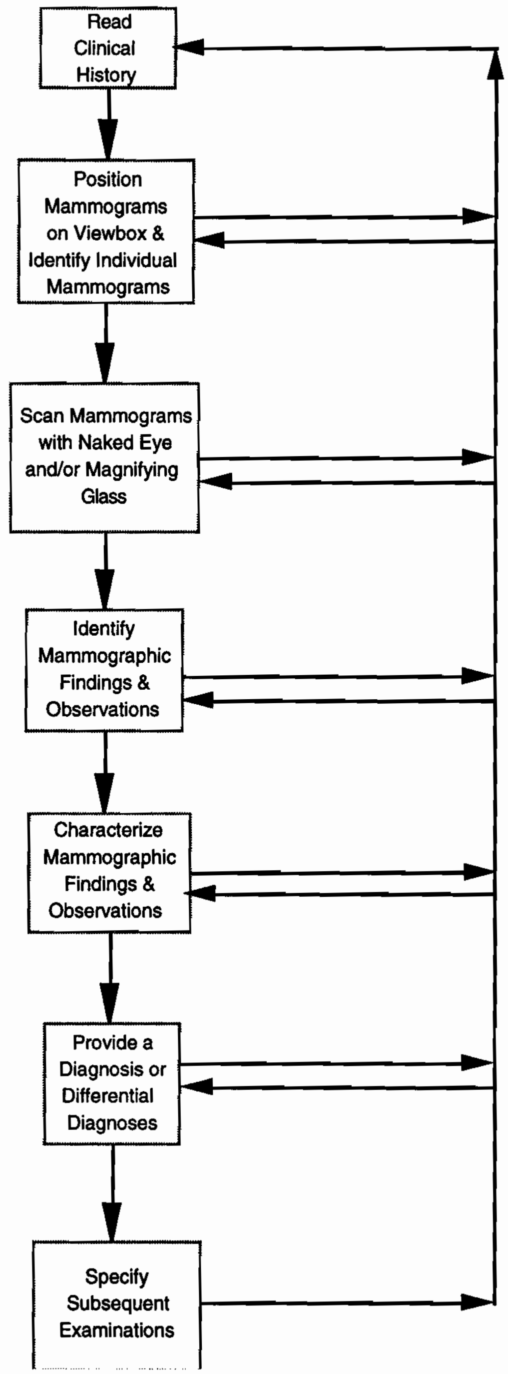



Figure 4.1. A Model of the Mammogram Interpretation Process.

information to construct a working hypothesis such as "this cluster of microcalcifications is suspicious for malignancy".

The sixth step involves providing a definitive diagnosis or a set of differential diagnoses. Typically, a definitive diagnosis is based on (a) a single feature (e.g., malignancy suspected based on the presence of an ill-defined mass), (b) a constellation of features (e.g., presence of punctate microcalcifications distributed throughout the breast), or (c) multiple features (e.g., two cancers in two separate and suspicious masses). It is important to note that ontologically, diagnoses are represented in a hierarchical format, and thus may be specified at different levels. However, in mammography, even though the diagnoses of "suspicious for malignancy" and "medullary carcinoma" represent different levels of specification they have the same implications (i.e., a biopsy should be performed).

However, radiologists cannot always specify a diagnosis and often provide a list of (ranked) differential diagnoses instead. Several factors can contribute to the inability to specify a single diagnosis: (a) the constellation of mammographic features may be representative of several diseases, (b) a lack of clinical information, (c) poor technical quality of films, (d) findings may be insufficient to specify a diagnosis, (e) a lack of clinical experience, and (f) a combination of these factors.

The last step involves specifying further investigations. In the case of a benign disease no further investigations are required. However, in the case of suspicious findings or malignancy a variety of examinations can be conducted to clarify the mammographic manifestations and provide additional information to specify a diagnosis. These examinations include follow-up (clinical correlation or additional mammograms), ultrasound, biopsy, and/or consultation with colleagues.

The model allows for a "linear approach" (from reading the clinical history to specifying subsequent examinations) or an “iterative approach” in which the results of a step may feed back to previous steps in the model. The model also comprises a certain amount of iteration as indicated by the vertical and horizontal arrows on the right side. For example, after positioning the mammograms on the viewbox the radiologist may refer back to the clinical history since it may indicate the location of the suspicious finding. Similarly, following the characterization of the mammographic findings the radiologist may re-scan the set of mammograms with a magnifying glass and identify new findings.

## Problem Solving Strategies, Operators and Control Processes Used by Staff Radiologists and Radiology Residents During Problem Solving

The second research question investigated whether staff radiologists and radiology residents used different problem solving strategies, operators, and control processes during mammogram interpretation. It was hypothesized that staff radiologists would solve the cases using a data-driven problem solving approach. In contrast, it was hypothesized that the radiology residents would solve the cases by using either goal-driven or a mixed-strategy (combination of data-driven and goal-driven problem solving strategies). This question was investigated by performing in-depth analyses of protocols from 40 breast disease cases to exemplify typical staff radiologists' and radiology residents' problem solving strategies.

Overall, the in-depth analyses of the same 40 protocols indicated that diagnostic reasoning in mammography is characterized by (a) the predominant use of a data-driven diagnostic strategy, (b) the use of a goal-driven strategy or a combination of data-driven and mixed strategies depending on case typicality and clinical experience, and (c) rapid schema-based problem-solving which facilitates search and characterization of mammographic features and integration of clinical history cues followed by accurate diagnosis and subsequent radiological recommendations. The evidence from this study does not wholly support the hypothesis that radiologists would solve the breast disease cases using a data-driven problem solving approach and that residents would solve the cases by using either goal-driven or a mixed-strategy. The results indicate that both staff and residents make use of data-driven and mixed problem solving strategies.

## Analyses of Diagnostic Problem Solving Strategies

This section is divided into two parts. The first consists of detailed examples of the two diagnostic problem solving strategies which were used extensively by all participants (data-driven and mixed-strategy). The second presents an in-depth examination of the problem solving strategies used by staff radiologists and residents in solving four cases (1, 5, 9, and 10) and provides detailed examples from the protocols. Appendix O contains a sample of the protocols that was used to develop the coding scheme.

### Analyses of Two Diagnostic Problem Solving Strategies

Illustrations of data-driven and mixed diagnostic problem solving strategies are presented below. Table 4.1 provides an example of a segmented protocol obtained from a staff radiologist (S10) solving a benign case (case 5) under the augmented condition. The protocol provides an example of a data-driven diagnostic strategy. The participant starts by reading the clinical history (segment #1). She immediately identifies a radiological observation by indicating that the breasts are very dense (segment #2). Next, she identifies, locates, and characterizes the critical mammographic finding (segment #3), and provides a diagnosis of hamartoma in segment #4. Since this is a benign lesion, she then proceeds to discuss the clinical management (segment #5), and provide a summary of her diagnosis (segment #6).

Table 4.1

A Segment Protocol From a Staff Radiologist (S10) Solving a Benign Case (case 5) Under the Augmented Condition.

1. Okay this is a young patient in her 30s who presented with large mass in the lateral portion of her.. left breast.
2. Okay so we have very dense breasts.
3. She has a very very large lesion here on the ah right which contains ah well-demarcated, large, very inhomogeneous, contains fat
4. And this has all the characteristics of a hamartoma on the breast.
5. And this is not a malignant lesion and as long as the mass is not bothering her I guess we'd leave it.
6. It's a benign lesion on the breast.

Table 4.2 provides an example of a segmented protocol from a staff radiologist (S3) solving a malignant case (case 6) under the augmented condition. This protocol provides an example of a mixed reasoning strategy involving bottom-up and top-down processes. The mixed strategy starts with a data-driven approach indicated by the participant's reading of the clinical history, identification of a mammographic observation, and assessment of the adequacy of the mammographic technique (segments #1-3). Subsequently, the strategy changes to a top-down approach illustrated by a goal statement (segment #4) where the participant is now searching for microcalcifications. During the search for microcalcifications, the participant identifies and classifies several other mammographic features (segments # 5-6). The search for microcalcifications ends with segment #7. The participant returns to a previous mammographic finding (segment #5), classifies it as indeterminate (segment # 8) and, due to the patient's clinical presentation, provides a tentative diagnosis (segment #9). Based on further assessment of the finding's features she changes the working hypothesis from carcinoma to indeterminate (segments # 10-13). Next, she reviews the findings and discusses other subsequent radiological examinations including stereotaxic biopsy and ultrasound (segments # 14-16). She then hypothesizes that the pathology results may indicate malignancy and that the mass should be excised (segments #17-18). However, the differential diagnoses of malignant and benign breast diseases are still active (segment #19). Segment #20 indicates that performing one of the subsequent examinations will clarify the nature of the mass.

Table 4.2

A Segmented Protocol From a Staff Radiologist (S3) Solving a Malignant Case (case 6) Under the Augmented Condition.

1. Okay, a woman in her 50's who presented with a mass in the upper outer quadrant of the right breast..
2. Right breast, left.
3. Immediately these breasts are dense, nodular, difficult to evaluate.
4. Looking for microcalcifications.
5. I see this area, asymmetric area outlined by the pen in both views in the upper outer quadrant.
6. And asymmetrical increased density.
7. No calcifications of any suspicion.
8. But this asymmetrical density is indeterminate,
9. and particularly since it's palpable, I think we have to rule out carcinoma.
10. It's not a difficult carcinoma.
11. But that doesn't mean it's not a spiculated mass but carcinomas can appear as asymmetrical densities and particularly if it doesn't have well-defined borders.
12. So it doesn't have any benign features either.
13. It's kind of indeterminate.
14. The fact that's it's palpable and ill-defined I would be concerned enough to suggest ah, it should probably be excised.
15. Another option would be to, the physician himself could biopsy it or it could be biopsied either under stereotaxic mammographic guidance or can see depending on the expertise.
16. We could do ultrasound and see if we can see it, biopsy it under ultrasound to know what the pathology is.
17. But ah I would be more suspicious that it's gonna be a carcinoma
18. and it's gonna have to come out anyway
19. unless there's I guess it's possible it could be benign
20. but it definitive needs a biopsy either excisional or under imaging guidance.

## Analyses of Diagnostic Problem Solving Strategies in Four Cases

The following section presents the detailed results of four cases (1, 5, 9, 10) from the analysis of the protocol data. The rationale for selecting these particular cases is based on several issues described below. First, they represent diverse mammographic manifestations including: a single visible finding (case 1), multiple visible findings but rare clinical incidence (case 5), perceptually difficult mammographic manifestations associated with the appearance of a possible second carcinoma (case 9), and numerous discrete findings with severe clinical implications (case 10). They also provide evidence of (a) use of different problem solving strategies, (b) influence of case type on scanning and reading times, and (c) identification of particular error types and their associations to case type. Each of these cases are discussed below in the following sequence: (a) brief case description, (b) typical staff protocol, (c) typical resident protocol, (d) problem solving protocol illustrating an error committed by a staff member, and (e) problem solving protocol illustrating an error committed by a resident.

**Case 1 - Mucin-Producing Carcinoma.** Case 1 is that of a 60 year old woman who has been diagnosed with a mucin-producing carcinoma (diagnosis confirmed by pathology report). The expert's protocol and four mammograms are presented in Appendix C. The mammogram depicts a radiological finding consisting of a large partly well-circumscribed mass in the right retroaereolar area and a radiological observation consisting of dense fibroglandular tissue bilaterally. This case was selected mainly because the mammographic finding was solitary, large, and visible to the naked eye. However, the characterization of the mass' margins is critical to making a diagnosis.

The protocols of a staff radiologist (S1 - 1A) and radiology resident (S13-1A) are presented below to illustrate the similarities between staff and residents when solving this case. Overall, staff and residents solve the problem in a similar manner. They: (a) read the clinical history (data acquisition), (b) identify the radiological finding (data examination), (c) characterize it (data exploration) according to size, margins, region and location, and (d) request subsequent examinations (diagnostic planning).

There are, however, several slight differences in how participants approached the case. The differences include the fact that the staff radiologist compared the finding on the ipsilateral view (data comparison), classified the left breast as normal (data classification), summarized the finding (summarization), associated the patient's age with the mammographic finding (data explanation), and proposed further recommendations (diagnostic planning) to rule out the possibility of a cyst.

**S1-1A - staff**
Okay first case here...clinical history a sixty year old woman who presented with a mass in the right breast.
Ah so we have just one.. set of films no previous.
Okay we have usual CC and MLO views, ah...general,
I don't know what the date is on here..an obvious finding is of a large almost completely well-circumscribed nodule just deep to the right nipple..ah, here..seen very well on both views,
okay the margins.are where I can see them are smooth and well-defined although I lose in the posterior portion of the CC view.
I lose the margin
and similarly, in the posterior portion of the MLO view I don't see the posterior margin clearly.
The background of the right breast is fairly dense for particularly a sixty year-old patient, and it seems taking into account technical differences somewhat denser than contralateral breast.
I don't see anything in the left breast which is suspicious.
So to summarize there is a approximately a 6 or 7 cm mostly well-circumscribed ah mass just deep to the right nipple.
A patient this age needs follow-up,
one compare to previous films and two this would be an excellent case for ultrasound.
If it's a cyst it should be aspirated if not if it's solid we could do a fine needle aspiration perhaps under ultrasound.

The resident identified (data examination), and characterized (data examination) the finding, and examined the rest of the right breast (data examination). She then proposed a diagnostic plan to rule out the possibility of a cyst (diagnostic planning), associated the mass density to a second hypothesis (cystosarcoma philloides), and repeated her original recommendation.

**S13-1A - resident**
Okay, a sixty year old woman who presented with a mass in the right breast.........
Okay. A mass in the right breast.
So here we are looking at a large well-defined mass in the central portion of the right breast.
It's large it measures 4 cm in diameter.
Ah... I don't see any ah calcifications elsewhere in the breast or any other abnormalities elsewhere .
Ah, this lesion I guess I would an ultrasound on it first ah just to make sure this doesn't represent a cyst
thought it's very dense and makes me think the first that I think it looks like is a cystosarcoma philloides but again it could I guess still be a cyst.
I think the ultrasound here is reasonable to do as a next step.

**<u>Case 5 - Hamartoma or Fibroadenolipoma</u>.** Case 5 is that of a young patient in her thirties who presented with a large mass in the lateral portion of the left breast. The expert's protocol and four mammograms are presented in Appendix G. The mammograms depict a finding which is a well-circumscribed mass consisting of fat and containing encapsulated islands of fibroglandular tissue. They also show two radiological observations: dense fibroglandular tissue and breast asymmetry which is typical for a patient of this age. This case was selected mainly because the mammographic finding was associated with multiple features which is a manifestation rarely encountered in a clinical situation.

The protocols of a staff radiologist (S3 - 5A) and radiology resident (S15-1A) are presented below to illustrate the similarities and differences between staff and residents. The staff radiologist read the clinical history (data acquisition), classified the left mammogram as being abnormal (classifies mammogram), and immediately proposed a diagnosis of fibroadenolipoma (hypothesis generation). He then identified and classified a benign microcalcification, identified the radiological finding (data examination), restated the original diagnosis of hamartoma (hypothesis generation), and proposed the removal of the mass for cosmetic reasons only (diagnostic planning).

**S3 - 5A - staff**
A young patient in her 30's who presented with a large mass in the lateral portion of the left breast.
Hhmm very abnormal. Very dense breasts.
This is a **adeno lipofibroadenolipoma**.
Okay, so very very dense breasts.
This is difficult to assess since smaller little masses could be hiding in typical young woman having dense fibroglandular tissue.
There's a benign microcalcification here.
There's a huge- there's a lesion but it consist of nodule but has fat in it.
Looks like a **fibroadenolipoma** which is a benign tumor.
It's radiologically, definitely benign.
There's a benign calcification.
But ah I suspect the breast is enlarged by it and probably for cosmetic reasons
and it's probably going to be removed but mammographically this is a benign lesion.
There are no suspicious calcifications and otherwise the breasts are difficult to assess because the rest of the tissue is dense and you can't really assess them radiographically which is unfortunate.

The resident solved the problem in a similar manner. However, the diagnosis is proposed at the end of the problem solving episode. The resident (a) read the clinical history (data acquisition), (b) assessed the technical quality of the films (data assessment) (c) identified the radiological finding (data examination), (d) characterized it according to its margins, density and associated components (data exploration), (f) proposed a diagnosis of hamartoma or adenofibroadenolipoma (hypothesis generation), and (g) requested no subsequent examinations because it was a benign mass (diagnostic planning).

**S15-5A - resident**
Case 5A. Clinical history. Young patient in her 30s who presented with a large mass in the lateral portion of the left breast...........
So we have bilateral mammograms in CC and MLO positioning.
These are older films ah demonstrating some of the chest wall,
which is not usual for these days.
But the quality of the examination is adequate.
The ah main abnormality is an obvious ah very large ah space occupying mass measuring greater than 14 centimeters by 12 centimeters in two dimensions.
The striking finding is that it is extremely well defined and has ah large lucent component as well as multiple ah soft tissue components within.
I think this is diagnostic of a hamartoma of the breast and adeno **fibroadenolipoma** tissue and this is benign.
And no follow up is required.

**Case 9 - Two carcinomas.** Case 9 is that of a 68 year old woman who had a palpable abnormality in the lateral portion of the left breast (upper outer quadrant). The expert's protocol and six mammograms are presented in Appendix K. The mammograms depict two radiological findings consisting of a large spiculated lesion having the appearance of malignancy, and a small area of retraction that may have represented a second carcinoma. This case was selected because in addition to an apparent carcinoma there was the suggestion of a second carcinoma (difficult to discern when comparing ipsilateral mammograms). The radiological manifestation is typical and routinely encountered during clinical practice.

The protocols of a staff radiologist (S10 - 9A) and radiology resident (S19-9B) are presented below. The staff radiologist identified and located the first cancer (data examination), verified its presence on the exaggerated CC view (data comparison), determined the nature of its margins (data exploration), provided a diagnosis of very suspicious for malignancy (hypothesis generation), and proposed a diagnostic plan (has to be pursued) of performing a biopsy. Subsequently, she immediately identified the second carcinoma (data examination), provided a hypothesis of "I'm worried" (suggesting malignancy), summarized the findings, and suggested a second biopsy for this second carcinoma (diagnostic plan).

**S10- 9A - staff**
....
Okay there's ah- we certainly have this **obvious lesion here in the tail of the breast** which we never get even on the exaggerated CCs,
which is this one here.
We never really see it on any other view.
This is a mag view, but anyway.
So **this is a lesion** which I'm not sure is spiculated.
Extreme, **very very high suspicion of malignancy**.
Ah and this- even though we can't see it on any other views, it **has to be pursued**.
There's another thing that worries me a bit which I think there may be a **second lesion over there**.
So on this one here **I'm worried** about.
So we have **I think there's two lesions**.
We have this one over here which has to be **biopsied** and those are tricky.
.........

The resident proceeded in a similar manner by identifying and locating the first carcinoma (data examination), identifying the second carcinoma (data identification), examining the margins of both carcinomas (data exploration), noting the absence of any associated findings, identifying a few more mammographic observations (data examination), classifying the rest of the left breast, proposing a diagnosis of multi-centric ductal carcinoma for both carcinomas (hypothesis generation), and proposing a biopsy for each cancer (diagnostic planning).

**S19-9B - resident**
.....
Ah the region of interest which is the **left upper outer quadrant** of the ah of the left breast which is the area where there's a palpable abnormality.
The ah breast demonstrates mainly fatty replacements in keeping with the mutations of age.
We note **two** ah areas of increased density which are fairly well defined.
Ahm however ah both of them seem to demonstrate spiculations.
I don't see any- there's no associated microcalcification in this lesion.
There are a few linear calcifications in the smaller lesion.. which is indeterminate in character as far as this microcalcification.
I think the remainder of the breast is normal.
And we have again, a normal lymph node in the axilla.
So, first you would obtain previous films but regardless of the previous film with a palpable abnormality and the lesion that is spiculated since there are **two** I have to consider a **multi-centric ductal carcinoma**.
First and foremost and these require a **biopsy**.

**Case 10 - Inflammatory Carcinoma.** Case 10 is that of a 60 year old who had experienced inflammation in her left breast for two weeks. The expert's protocol and four mammograms are presented in Appendix L. The mammograms depict four radiological findings, some asymmetry, no discrete mass in the left breast, skin thickening around the left breast, and tissue with a reticulated appearance. This case was selected mainly because, in addition to mammographic manifestations, there are findings which are extremely suggestive of inflammatory carcinoma and of many other breast diseases such as unilateral edema, metastatic disease, etc. A biopsy must be performed promptly since the inflammatory carcinoma is rapidly progressing and has a extremely poor prognosis. Clinical experience and knowledge of the clinical literature in mammography should facilitate the diagnosis.

The protocols of a staff radiologist (S8 - 10A) and radiology resident (S12-10A) are presented below. The staff radiologist read the clinical history (data acquisition), identified three of the four radiological findings (data examination), explained the changes in the left breast based on inflammation (data explanation), and provided two diagnoses (hypothesis generation). However, he could not totally rule out cancer even though there was no mass in the left breast. As such, he proposed a diagnostic plan of antibiotics and a follow-up examination. This is another case of a data-driven problem solving approach.

**S8-10A - staff**
And this patient here. A 60 year-old..
And we have mammogram done with an antiquated .
A there is some asymmetry in the ah breast here with left breast being denser with **thickening of the skin** throughout.
There is **no mass** that I can localize within this **left breast**.
Ahm and the **changes** that we have here on the left side as compared to the right are those of an inflamed breast.
And I see here that it does correspond with inflammation of her left breast for two weeks..
The basis for the inflammation ah is **not necessarily cancer**.. it may be... ah **mastitis**.
As **I do not see a large mass**.
The patient has had the inflammation for two weeks.
I think in a situation like this, the right breast being normal and there being no mass on the left side, I would suggest that the patient is put on a **course of antibiotics** and have a **follow-up clinical examination**.

The resident proceeded in a similar manner (data-driven) as the staff radiologist. However, the difference lay in the resident's use of data exploration after the identification of each finding. He read the clinical history (data acquisition), identified each radiological finding (data examination), and then characterized it (data exploration) until all four were completed. Subsequently, he proposed a diagnosis of mastitis based on an association with inflammation in the left breast. However, he then provided another diagnosis. This was followed by a diagnostic plan which involved antibiotic treatment to rule out mastitis. The resident suggested that if the condition was not resolved after antibiotic treatment and if the follow-up mammograms showed no change, then the left breast would have to be biopsied.

**S12-10A - resident**
Case 10A. A 60 year-old woman who presented with inflammation of the left breast for two weeks.....
So I'm seeing an **asymmetry** in the ah **breasts**.
Ah the tissue in the left breast just seems a little bit diffusely increased in density and I'm seeing increase in **skin thickness** in the left breast which is the one which is inflamed.
Ah the skin thickness is mostly over the ah areolar but extends ah a little bit to each side, also.
I don't see **any evidence of a mass** in the microcalcifications.
Ahm if this woman has inflammation of her left breast we could be dealing with a **mastitis**.
She may have a **mastitis** or ahm a condition of an **inflammatory carcinoma** which can be ah difficult to see sometimes.
I would recommend a ah repeat **ultrasound** in ah one month time **after having antibiotic treatment** ahm for a mastitis.
If she has appropriate treatment for mastitis and ah comes back with the situation having been resolved and the mammogram is normal I'd be happy to sign this off as a mastitis but if the ah symptoms are not resolved and the mammographic picture remains unchanged then we'd have to proceed to **biopsy** of the left breast.

In summary, the hypotheses that staff radiologists would solve the cases using a data-driven problem solving approach and that the radiology residents would solve the cases by using either goal-driven or a mixed-strategy (combination of data-driven and goal-driven problem solving strategies) were not wholly supported. The results indicate that both staff and residents used data-driven and mixed strategy problem solving strategies.

## Frequency of Operators Used by Participants

The second research question also investigated whether staff radiologists and radiology residents used different problem solving operators during mammogram interpretation. A definitive hypothesis was not posited since literature in this area of medical diagnosis is lacking. Overall, based on frequency analyses the results indicate that both groups used the same types of problem solving operators regardless of experimental conditions.

The frequency of operator use by participants across levels of expertise and experimental conditions is provided in Table 4.3. The number of operators used was calculated by selecting 40 protocols at random from a pool of 160 protocols (200 minus 40

Table 4.3
Frequency of Operator Use by Level of Expertise and Experimental Condition.

| Operators | Staff Radiologists | | Radiology Residents | |
|---|---|---|---|---|
| | Authentic Condition | Augmented Condition | Authentic Condition | Augmented Condition |
| **1. Data Acquisition** | | | | |
| 1.1 Reads Clinical History | 10 | 10 | 10 | 10 |
| 1.2 Places & Inspects Mammograms | 10 | 10 | 10 | 10 |
| **Total** | 20 | 20 | 20 | 20 |
| **2. Data Identification** | | | | |
| 2.1 Identifies Mammograms As Single Study | 2 | 3 | 3 | 5 |
| **Total** | 2 | 3 | 3 | 5 |
| **3. Data Assessment** | | | | |
| 3.1 Comments on Technical Quality of Mammograms | 3 | 3 | 0 | 2 |
| 3.2 Inability to Assess Mammographic Cue | 0 | 2 | 1 | 1 |
| **Total** | 3 | 5 | 1 | 3 |
| **4. Data Examination** | | | | |
| 4.1 Identifies Mammographic Cue | 21 | 21 | 23 | 27 |
| 4.2 Identifies Marked Mammographic Cue | 0 | 2 | 1 | 3 |
| 4.3 Identifies Artifacts on Mammograms | 0 | 1 | 0 | 0 |
| **Total** | 21 | 24 | 24 | 30 |
| **5. Data Exploration** | | | | |
| 5.1 Characterizes Mammographic Cue | 8 | 13 | 18 | 15 |
| 5.2 Characterizes Associates Mammographic Cue | 4 | 1 | 7 | 7 |
| 5.3 Notes-Absent-Data | 3 | 9 | 8 | 8 |
| 5.4 Re-Reads Clinical History | 2 | 0 | 2 | 4 |
| **Total** | 17 | 23 | 35 | 34 |
| **6. Data Comparison** | | | | |
| 6.1 Compares Ipsilateral | 1 | 2 | 8 | 6 |
| 6.2 Compares Contralateral | 0 | 2 | 2 | 1 |
| 6.3 Compares Mammographic Cue to Additional Views | 0 | 1 | 1 | 1 |
| 6.4 Compares to the Rest of the Breast | 0 | 0 | 1 | 0 |
| **Total** | 1 | 5 | 12 | 8 |
| **7. Data Classification** | | | | |
| 7.1 Classifies Mammographic Cue | 7 | 2 | 5 | 10 |
| 7.2 Identifies & Classifies Mammographic Cue | 3 | 3 | 0 | 3 |
| 7.3 Classifies Mammograms | 1 | 0 | 0 | 1 |
| **Total** | 11 | 5 | 5 | 14 |
| **8. Data Explanation** | | | | |
| 8.1 Infers-Pathophysiological-Cause | 0 | 0 | 1 | 1 |
| 8.2 Infers-History | 2 | 2 | 0 | 1 |
| **Total** | 2 | 2 | 1 | 2 |
| **9. Hypothesis Generation** | | | | |
| 9.1 Trigger | 9 | 11 | 10 | 11 |
| 9.2 Further-Specification | 8 | 2 | 3 | 6 |
| 9.3 Generalization | 2 | 2 | 3 | 2 |
| 9.4 Unspecified Diagnosis | 0 | 0 | 0 | 0 |
| **Total** | 19 | 15 | 16 | 19 |
| **10. Hypothesis Evaluation** | | | | |
| 10.1 Confirmation | 4 | 0 | 0 | 0 |
| 10.2 Disconfirmation | 0 | 1 | 0 | 3 |
| 10.3 Causal-Relationship | 0 | 0 | 0 | 0 |
| **Total** | 4 | 1 | 0 | 3 |
| **11. Summarization** | | | | |
| 11.1 Repeat Data | 1 | 5 | 3 | 3 |
| 11.2 Repeat Hypothesis | 0 | 1 | 1 | 0 |
| **Total** | 1 | 6 | 4 | 3 |
| **TOTAL** | 101 | 109 | 121 | 141 |

used for developing the coding scheme). The coding scheme was then applied to the 40 selected protocols, and a breakdown of the use of operators is presented in Table 4.3.

Overall, the results indicate that residents used more operators than the staff. Both groups used more operators when solving cases presented in the augmented condition. An analysis of the frequency of operator use by staff and residents during the diagnostic process in cases presented under both experimental conditions revealed a predominant use of the following operators (listed in order of descending frequency): (a) data examination, (b) data acquisition, (c) data exploration, and (d) hypothesis generation. These four operators account for 76% (88 out of 101 operators and 82 out of 109 operators) of all operators used by the staff in solving cases under both experimental conditions, 79% (107 out of 121 operators) of all operators used by residents solving cases presented under the authentic condition, and 72% (117 out of 143 operators) of all operators used by residents solving cases presented under the augmented condition.

**Frequency of Use of Control Processes**

The second research question also dealt with the question of whether staff radiologists and radiology residents used different control processes during mammogram interpretation. Similar to the use of problem solving operators, a definitive hypothesis was not posited since research in this area of medical diagnosis is lacking. Overall, results based on frequency analyses indicate that the use of control processes is based on level of expertise and experimental conditions.

The frequencies of control processes used by participants across levels of expertise and experimental conditions is illustrated in Table 4.4. The frequencies were calculated from the same 40 randomly selected protocols used to calculate the frequency of operator use. Overall, the table shows that staff used slightly more control processes than the residents (47 as compared to 41). However, staff used more control processes when solving cases under the augmented condition than they did under the authentic condition

(28 as compared to 19). In contrast, residents used more control processes when solving the cases presented under the authentic condition (23 as compared to 18). Diagnostic planning was the most often used control process regardless of experimental condition, followed by goals. The staff used two-thirds of the goal operators when solving cases under the augmented condition (10 as compared to 5). None of the residents in the sample used goals. The other two control processes used by participants included self-evaluation of diagnostic strategy and experiential memory.

Table 4.4

Frequency of Control Processes Used by Level of Expertise and Experimental Condition.

| | Staff Radiologists | | Radiology Residents | |
|---|---|---|---|---|
| Control Processes | Authentic Condition | Augmented Condition | Authentic Condition | Augmented Condition |
| Diagnostic Plans | 14 | 16 | 22 | 18 |
| Goals | 5 | 10 | 0 | 0 |
| Self-Evaluation of Diagnostic Strategy | 0 | 1 | 1 | 0 |
| Experiential Memory | 0 | 1 | 0 | 0 |
| **TOTAL** | **19** | **28** | **23** | **18** |

## Quantitative Analyses of Several Measures on Groups' Performance

The third research question investigated whether expertise effects and experimental condition effects influence several performance measures. These measures included: (a) number of mammogram findings, (b) number of observations, (c) number of diagnoses, (d) scanning time to construct an initial mental representation based on the clinical history and set of mammograms, (e) reading time required to solve each breast disease case, (f) accuracy ratings for diagnosis, (g) accuracy ratings for subsequent examination, and (h) accuracy ratings for overall diagnostic accuracy.

Two hypotheses were tested. First, it was hypothesized that, compared to the authentic condition, the augmented condition would facilitate residents' ability to identify findings and observations, provide fewer diagnoses, decrease their scanning and reading time, and increase their ratings on accuracy (for diagnoses, subsequent examinations, and overall diagnostic accuracy).

Secondly, it was hypothesized that, in comparison to radiology residents, staff radiologists would be faster in scanning and reading mammograms, and have better ratings for accuracy (for diagnoses, subsequent examinations and overall diagnostic reasoning regardless of experimental condition).

**Number of Radiological Findings, Observations, and Diagnoses**

The means and standard deviations for the number of radiological findings, radiological observations, and diagnoses per case are presented in Table 4.5. Initially, 3 repeated measures ANOVAs were to be performed to compare the means across the levels of expertise and experimental conditions on radiological findings, radiological observations, and number of diagnoses. However, the ANOVAs were not conducted given the minimal differences between the means and standard deviations (see Table 4.5). The non-statistical comparisons of the means suggests there are no differences between the mean number of radiological findings, observations and diagnoses between the groups across the two experimental conditions. On average, participants identified at least one radiological finding, made three radiological observations, and gave one diagnosis per case. Participants in both groups tended to identify more radiological observations than radiological findings. All participants provided at least one diagnosis per case.

Table 4.5

Means and Standard Deviations for the Number of Radiological Findings, Radiological Observations and Diagnoses by Level of Expertise Across Experimental Conditions per Case

| | Staff Radiologists | | Radiology Residents | |
|---|---|---|---|---|
| Dependent Measures | Authentic Condition | Augmented Condition | Authentic Condition | Augmented Condition |
| Radiological Findings | 1.1 (0.2) | 1.1 (0.3) | 1.1 (0.2) | 1.2 (0.1) |
| Radiological Observations | 2.9 (1.3) | 3.2 (1.0) | 3.4 (1.3) | 3.1 (1.0) |
| Diagnoses | 1.3 (0.3) | 1.3 (0.2) | 1.2 (0.3) | 1.3 (0.2) |

## Scanning Time for Data Acquisition

Scanning time was defined as the amount of time (in seconds) it took a participant to read the clinical history, remove the set of mammograms from the envelope, place them on the viewbox, inspect them with either the naked eye and/or the magnifying glass, produce their first utterance. This operational definition is "loose" in comparison to eye movement studies which provide precise operational definitions and use sophisticated equipment required to record scanning time. A repeated measures ANOVA was performed to calculate the level of statistical significance based on level of expertise and by experimental condition on the mean scanning time. Results indicated a significant main effect for expertise, $F\,(1,18) = 4.89\ p < .05$, although there was no significant main effect for experimental condition ($F = 0.63$, $p > .05$) and no interaction ($F = .05$, $p > .05$). As expected, the staff radiologists were significantly faster than residents in scanning the mammograms. Table 4.6 presents the means and standard deviations for scanning time by level of expertise across experimental conditions per case. The means for scanning time indicate that, on average, the residents took longer than staff to scan each breast disease case.

Table 4.6

Means and Standard Deviations for Scanning Time by Level of Expertise Across Experimental Conditions

| | Staff Radiologists | | Radiology Residents | |
|---|---|---|---|---|
| Dependent Measures | Authentic Condition | Augmented Condition | Authentic Condition | Augmented Condition |
| Scanning Time (across five cases) | 230.6 (29.5) | 236.1 (33.1) | 313.9 (29.5) | 327.9 (33.1) |
| Scanning Time (per case) | 46.1 (17.6) | 47.2 (22.2) | 62.8 (19.6) | 65.6 (19.6) |

## Mean Scanning Times Across Cases

An analysis of the mean scanning times for both group and experimental condition was conducted to verify whether the case characteristics (such as number of mammographic findings, observations, and number of mammograms in each case) had an impact on the scanning time. Figure 4.2 presents the mean scanning time per case by level of expertise and experimental conditions. The results seem to indicate that on average, staff scanned cases faster than residents. Cases 7 and 9 are exceptions and they are both cases of carcinoma with no mammographic observations. However, these cases have mammographic findings which are considered difficult to localize and characterize (e.g., pleomorphic microcalcifications and a possible second carcinoma).

The results suggest that, overall, participants were faster at scanning cases 1, 5, and 6 (cases 1 and 6 were malignant and case 5 was benign). All three cases contain one radiological finding (e.g., a large partly well-circumscribed mass) and one pertinent radiological observation (e.g., dense fibroglandular tissue). The findings consist of relatively large masses or densities associated with other major breast changes (such as skin retraction). Participants took longer to scan cases 3 and 9. These cases depict malignant diseases with mammographic manifestations (a poorly defined nodule or a suspicious second carcinoma) that may require the participant to spend more time scanning the mammograms.

The figure also shows that on average, staff were faster than residents in scanning cases 2, 4, 5, 6 and 8 (in the augmented condition). Alternatively, under the augmented condition, residents were slower than staff in scanning cases 2, 4, 5, 6, 9, and 10. In contrast, when presented under the authentic condition, staff were faster than residents in scanning cases 1, 3, 7, and 10. Residents were faster than staff in scanning case 9 when it was presented under the authentic condition.

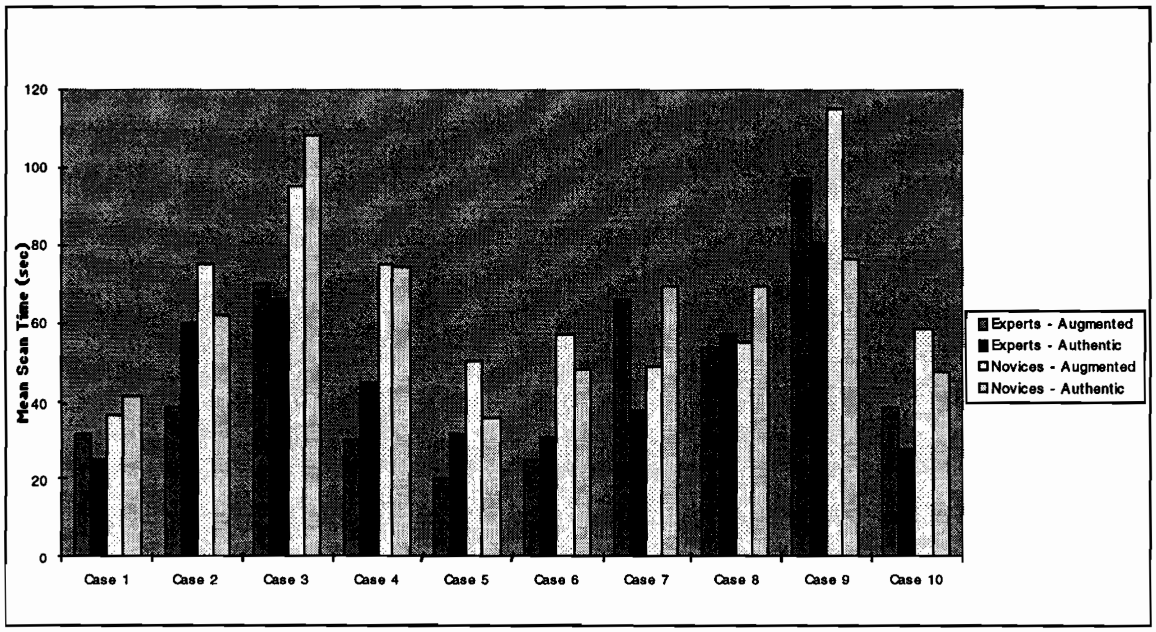



Figure 4.2. Mean Scanning Time Per Case by Level of Expertise and Experimental Conditions.

## Reading Time for Diagnosis

Reading time was defined as the total amount of time (in seconds) it took a participant to solve a breast disease case, from the initial reading of the clinical history until the last utterance made by the participant while solving the case. A repeated measures ANOVA was performed to calculate the level of statistical significance based on level of expertise and by experimental condition on the mean reading time. Results did not indicate a significant main effect for expertise ($F$ [1,18] = 1.57, $p$ >.05) or condition ($F$ = 0.11, $p$ >.05), and there was no interaction ($F$ = .0009, $p$ >.05). Overall, the results do not support the hypothesis that staff radiologists would read the cases faster than the radiology residents. Table 4.7 presents the means and standard deviations for reading time by level of expertise across experimental conditions across five cases (each participant solved five authentic and five augmented cases) and per case.

Table 4.7

Means and Standard Deviations per Case for Reading Time by Level of Expertise Across Experimental Conditions

| Dependent Measures | Staff Radiologists | | Radiology Residents | |
|---|---|---|---|---|
| | Authentic Condition | Augmented Condition | Authentic Condition | Augmented Condition |
| Reading Time (across five cases) | 883.5 (71.9) | 874.0 (63.2) | 998.1 (71.2) | 986.8 (63.2) |
| Reading Time (per case) | 176.7 (50.4) | 174.8 (45.5) | 199.6 (39.9) | 197.4 (33.5) |

**Mean Reading Times Across Cases**

A similar plotting of mean scanning times across levels of expertise and experimental condition was performed for the mean reading times. This analysis was conducted to verify if the case characteristics (such as number of mammographic findings, observations, and number of mammograms in each case) had an impact on the reading time. Figure 4.3 presents the mean reading time per case by level of expertise and experimental conditions. The figure tends to indicate that on average, staff were faster than residents in reading cases 1, 3, 4, 6, and 10. These include both benign and malignant cases and represent cases in which, according to the consulting radiologist, clinical experience would facilitate diagnosis. The results are similar to those presented on Figure 4.2 which suggest that participants were faster at scanning cases 1, 5, and 6.

Figure 4.3 shows that on average, under the augmented condition, staff were faster than residents in reading cases 2, 3, and 4. Alternatively, under the augmented condition residents were faster than staff in reading case 8 only. In contrast, under the authentic condition, staff were faster than residents in reading cases 1, 6, 7, 9, and 10. Alternatively, residents were faster than staff in reading case 5 when it was presented under the authentic condition.

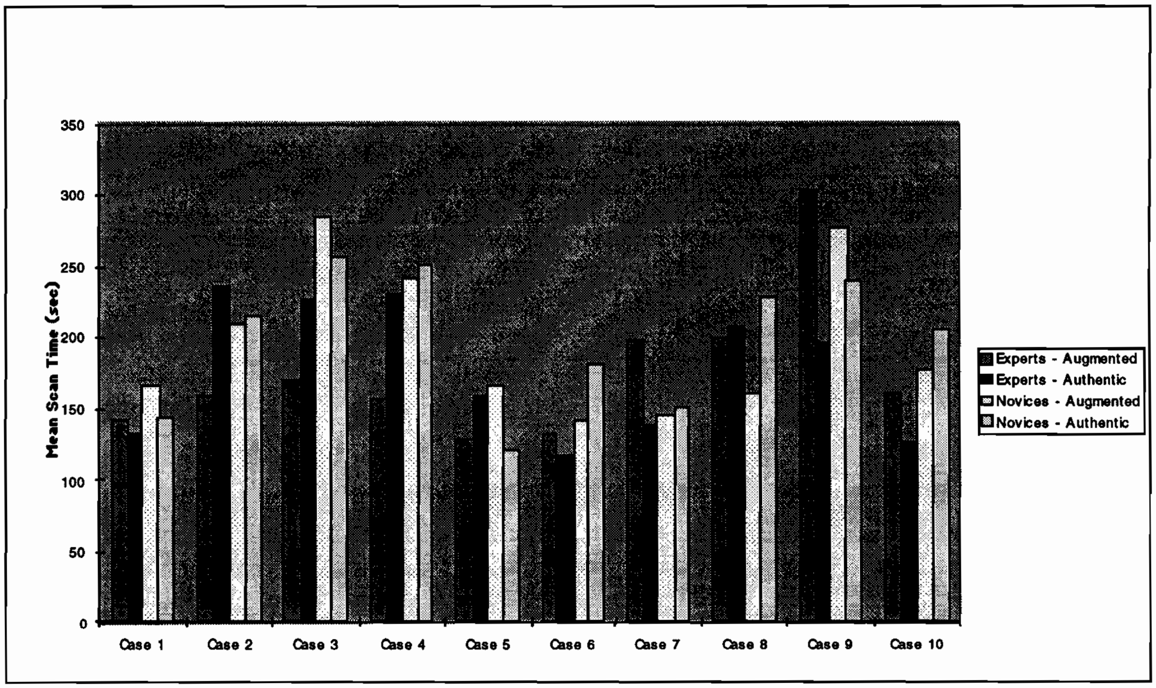



Figure 4.3. Mean Reading Time per Case by Level of Expertise and Experimental Conditions.

## Accuracy of Diagnoses

The levels of correctness (total number and percentage of cases) for diagnoses, radiological recommendations, and overall diagnostic accuracy by levels of expertise across experimental conditions are presented in Table 4.8. For these analyses, the consulting radiologist was provided with unmarked (e.g., all indications regarding expertise level, case number, and experimental condition were removed), transcribed verbal protocols. She was instructed to rate the diagnoses, radiological recommendations, and overall diagnostic accuracy as either *correct*, *indeterminate* or *wrong*. A diagnosis was deemed correct if it matched the consulting radiologist's diagnosis. A diagnosis at a level of abstraction different than the expert's was also considered correct. For example, if the diagnosis was carcinoma, a participant diagnosis of suspicious for malignancy would be coded as *correct*. *Indeterminate* was defined as a diagnosis which could not be classified in the same diagnostic category as the correct diagnosis. A diagnosis was rated as *incorrect* when it was fundamentally wrong. For example, a diagnosis of normal for a malignant case would be rated incorrect. The results of this verification process including a description of all diagnoses and recommendations for the ten cases by case number, experimental condition, level of expertise, and participant number are presented in Appendix P.

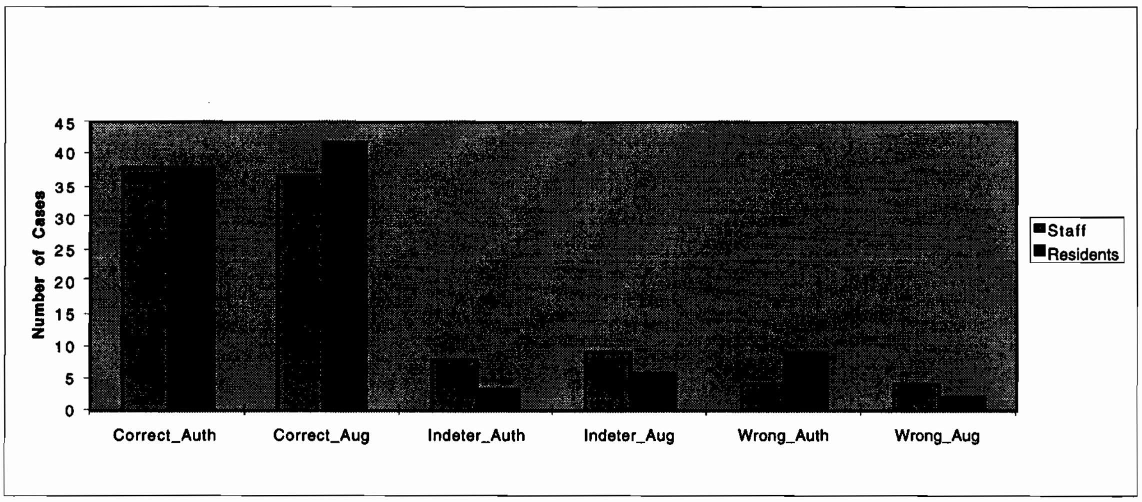
The minimal differences in the frequency of correct diagnoses across groups and experimental conditions indicates that the performance of the two groups is very similar. The frequency of correct diagnoses across groups and experimental conditions are presented in Table 4.8. The staff radiologists provided: (a) fewer correct diagnoses (38% as compared to 40%), (b) more indeterminate diagnoses (9% as compared to 5%), and (c) less inaccurate diagnoses (4% as compared to 6%). The results also seem to indicate that the residents benefited from the highlighted mammographic findings since they provided less incorrect diagnoses when solving cases presented under the augmented condition (1% as compared to 5%). The results are also presented in a histogram in Figure 4.4 which facilitates the visual inspection of the frequencies by level of correctness, expertise level, and experimental condition.

Table 4.8

<u>Frequency and Percentage of All Cases Illustrating the Levels of Correctness for Diagnoses, Radiological Recommendations and Overall Diagnostic Accuracy by Levels of Expertise Across Experimental Conditions</u>

| | Staff Radiologists | | | | | | Radiology Residents | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dependent Measures | Authentic Condition | | | Augmented Condition | | | Authentic Condition | | | Augmented Condition | | |
| | C | I | W | C | I | W | C | I | W | C | I | W |
| Diagnosis | 38 (19%) | 8 (4%) | 4 (2%) | 37 (19%) | 9 (5%) | 4 (2%) | 38 (19%) | 3 (2%) | 9 (5%) | 42 (21%) | 6 (3%) | 2 (1%) |
| Radiological Recommendations | 41 (21%) | 2 (1%) | 7 (4%) | 42 (21& %) | 5 (3%) | 3 (2%) | 42 (21%) | 2 (1%) | 6 (3%) | 46 (23%) | 1 (1%) | 3 (2%) |
| Overall Diagnostic Accuracy | 40 (20%) | 6 (3%) | 4 (2%) | 41 (21%) | 5 (3%) | 4 (2%) | 39 (20%) | 4 (2%) | 7 (4%) | 46 (23%) | 2 (1%) | 2 (1%) |

<u>Key:</u> C = Correct, I = Indeterminate, W = Wrong



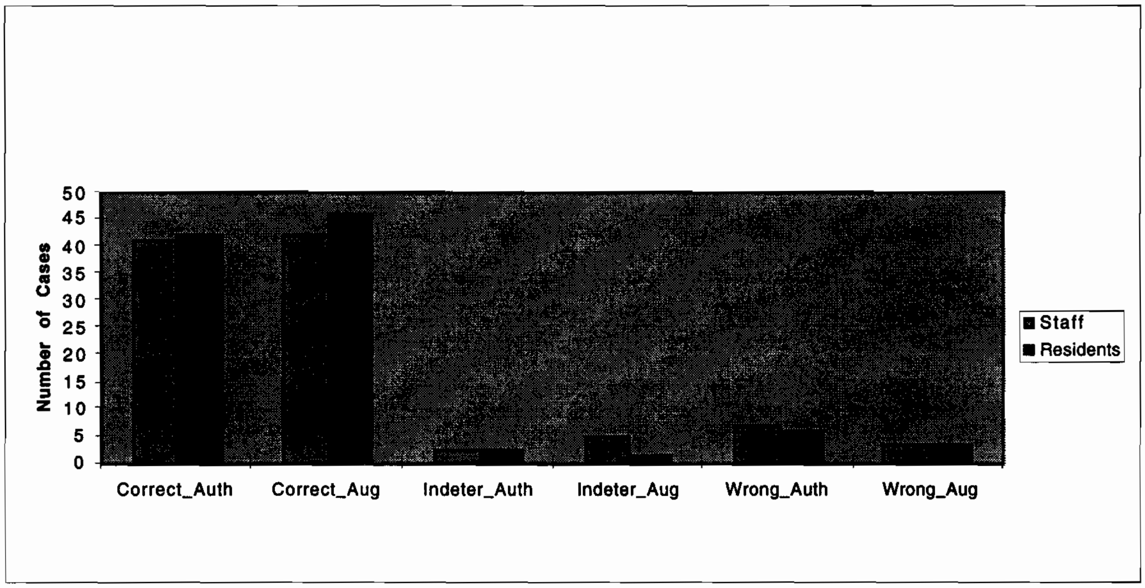

<u>Key:</u> Auth = Authentic, Aug = Augmented, Indeter = Indeterminate

<u>Figure 4.4.</u> Number of Correct, Indeterminate, and Wrong Diagnoses For All Cases By Level of Expertise and Experimental Conditions.

**Accuracy of Recommendations**

The frequency and the percentages of radiological recommendations provided by both groups across experimental conditions is provided in Table 4.8. Radiological recommendations include no further examinations (in the case of benign diseases), clinical follow-up, mammographic follow-up, ultrasound, biopsy or consultation with colleagues. A recommendation was correct if it was consistent with a participant's characterization of mammographic features or diagnostic hypothesis and the actual case. For example, if a participant, correctly gave a diagnosis of carcinoma, the subsequent recommendation would be a biopsy. Indeterminate recommendations were not adequate given a participant's characterization of mammographic features or diagnostic hypothesis. For example, for a diagnosis of carcinoma, a recommendation of ultrasound would be indeterminate since a more aggressive procedure (i.e., biopsy) is required. Wrong recommendations were those in which an incorrect subsequent examination was specified (based on the participant's characterization of mammographic features or diagnostic hypothesis). For example, a biopsy is an incorrect subsequent examination for a case diagnosed as normal.

The minimal differences in the frequency of recommendations across groups and experimental conditions indicates that the performance of the two groups is very similar. The staff radiologists provided: (a) slightly fewer correct radiological examinations (42% as compared to 44%), (b) more indeterminate radiological examinations (4% as compared to 2%), and (c) more inaccurate radiological examinations (6% as compared to 5%). The results also indicate that both groups benefited from the highlighted mammographic findings since they suggested fewer incorrect radiological examinations when solving cases presented under the augmented condition. The results are also presented in a histogram (see Figure 4.5).



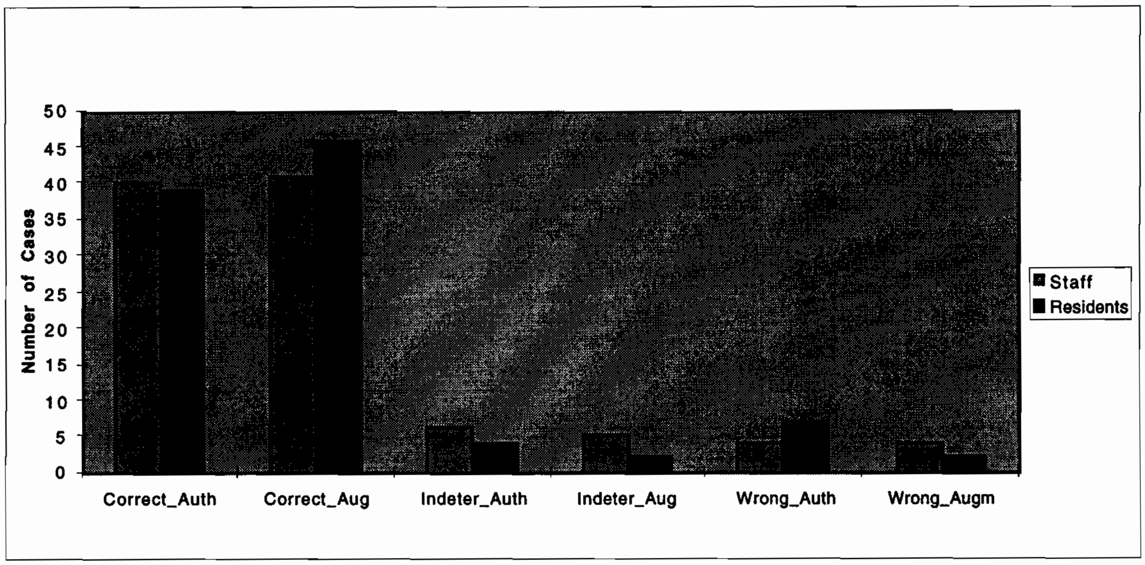

Key: Auth = Authentic, Aug = Augmented, Indeter = Indeterminate

Figure 4.5. Number of Correct, Indeterminate, and Wrong Recommendations for All Cases by Level of Expertise and Experimental Conditions.

**Overall Diagnostic Accuracy**

The total number and the percentages for overall diagnostic accuracy provided by both groups across experimental conditions is also presented in Table 4.8. Overall diagnostic accuracy includes the combination of diagnoses and radiological recommendations. For example, a diagnosis of a carcinoma followed-up by an excisional biopsy would constitute an *accurate* overall diagnosis. In contrast, a diagnosis of a benign lesion followed-up by a biopsy would constitute an *inaccurate* overall diagnosis. Twenty-five percent of the participants (5 out of 20), including three staff and two residents correctly diagnosed and provided the correct subsequent recommendations for the ten breast disease cases.

Again there were minimal differences in the frequency for overall diagnostic accuracy across groups and experimental conditions. The frequencies of overall diagnostic accuracy across groups and experimental conditions are presented in Table 4.8. The staff radiologists provided: (a) fewer correct overall diagnoses (41% as compared to 43%), (b) more indeterminate overall diagnoses (6% as compared to 3%), and (c) less inaccurate overall diagnoses (4% as compared to 5%). The small number of correct, indeterminate, and wrong cases was not sufficient to conduct log-linear analyses across level of expertise and by experimental condition. Therefore, a 2 X 2 Chi-Squares analysis was performed on the number of correct and wrong overall accuracy ratings across levels of expertise and experimental conditions (by collapsing indeterminate and wrong errors together). The analysis revealed a non-significant difference in the distribution of the number of cases across levels of expertise and correctness of overall diagnostic accuracy ($\chi^2$ [1, $N$ = 200] = .57, $p > .05$). The results are also presented in a histogram format in Figure 4.6.




Key: Auth = Authentic, Aug = Augmented, Indeter = Indeterminate

Figure 4.6. Number of Correct, Indeterminate, and Wrong Diagnosis and Recommendations for All Cases by Level of Expertise and Experimental Conditions.

The characteristics of the three staff members who achieved correct overall diagnostic accuracy for all ten cases are presented in Table 4.9. They had: (a) 16 to 32 years as physicians, (b) 10 to 26 years as staff radiologists, (c) diagnosed between 6,000 and 56,250 mammograms, and (d) 108 to 216 months of mammography experience. One staff member did not "see" or "read" mammograms on a weekly basis. However, the other two saw 60 and 75, and read 75 and 100 mammograms per week.

Table 4.9

The Characteristics of the Three Staff Radiologists Who Achieved Correct Overall Diagnostic Accuracy for All Ten Cases.

| Expertise Level | Years as Physician | Years as Radiologist | Total Number of Mammograms Seen & Read | Months of Mammography Experience | Mammo-grams Seen Per Week | Mammo-grams Read Per Week |
|---|---|---|---|---|---|---|
| Staff (S3) | 16 | 10 | 6,000 | 108 | 0 | 0 |
| Staff (S4) | 28 | 24 | 56,250 | 180 | 75 | 75 |
| Staff (10) | 32 | 26 | 35,000 | 216 | 60 | 60 |

The characteristics of the two fifth year residents who were given correct overall diagnostic accuracy ratings for all ten cases are presented in Table 4.10. They reported: (a) having read 250 to 1000 mammograms, (b) having one month of mammography experience, and (c) not seeing or reading mammograms on a weekly basis.

Table 4.10

The Characteristics of the Two Radiology Residents Who Achieved Correct Overall Diagnostic Accuracy for All Ten Cases.

| Expertise Level | Years as Physician | Residency Level | Range of Mammograms Seen & Read | Months of Mammography Experience | Mammograms Seen Per Week | Mammograms Read Per Week |
|---|---|---|---|---|---|---|
| Resident (S17) | 5 | 5th year | 250 - 1,000 | 1 | 0 | 0 |
| Resident (S19) | 5 | 5th year | 250 - 1,000 | 1 | 0 | 0 |

The worst overall diagnostic accuracy ratings were given to two staff and two residents. The two staff were the most junior members, having: (a) 5.5 and 8 years as physicians, (b) 5.5 and 3.5 years as staff radiologists, (c) experience diagnosing 250 to 1, 000 and 7,000 cases, and (d) mammography experience ranging from 5 months to 54 months (4.5 years). One staff member did not "see" or "read" mammograms on a weekly basis, while the other saw and read 30 mammograms per week. One resident was a third-year and the other a fifth-year, and both reported having seen and read between 100 and 250 mammograms. One reported to have one month of mammography experience while the other had none.

## Error Analyses: Frequency, Types, and Protocols Analyses

The fourth research question aimed at investigating the effects of the authentic and augmented experimental conditions on the frequency and types of errors committed by both groups. It was hypothesized that, in comparison to radiology residents, staff radiologists would commit fewer errors regardless of experimental condition. It was also hypothesized that residents would make fewer errors when solving cases presented under the augmented condition. These hypotheses were investigated by (a) conducting in-depth analyses of protocols from these cases, (b) identifying errors, and (c) classifying errors committed by participants during the problem solving task. Further, protocols are provided to exemplify errors committed by staff and resident.

## Frequency of Errors

Of the 200 protocols, 34 (17%) errors were committed by staff and residents across the two experimental conditions. The staff made slightly more errors than residents in the overall diagnostic accuracy (19 as compared to 15), and the residents received slightly more incorrect ratings than the staff (9 as compared to 8). The staff received almost twice the number of indeterminate ratings than the residents (11 as compared to 6). Figure 4.7 illustrates the number and types of errors committed by level of expertise and experimental conditions based on overall diagnostic accuracy.

Of the 34 errors, 17 were coded as wrong and 17 as indeterminate. The staff gave eight of the 17 wrong answers while the residents gave nine. The staff committed an

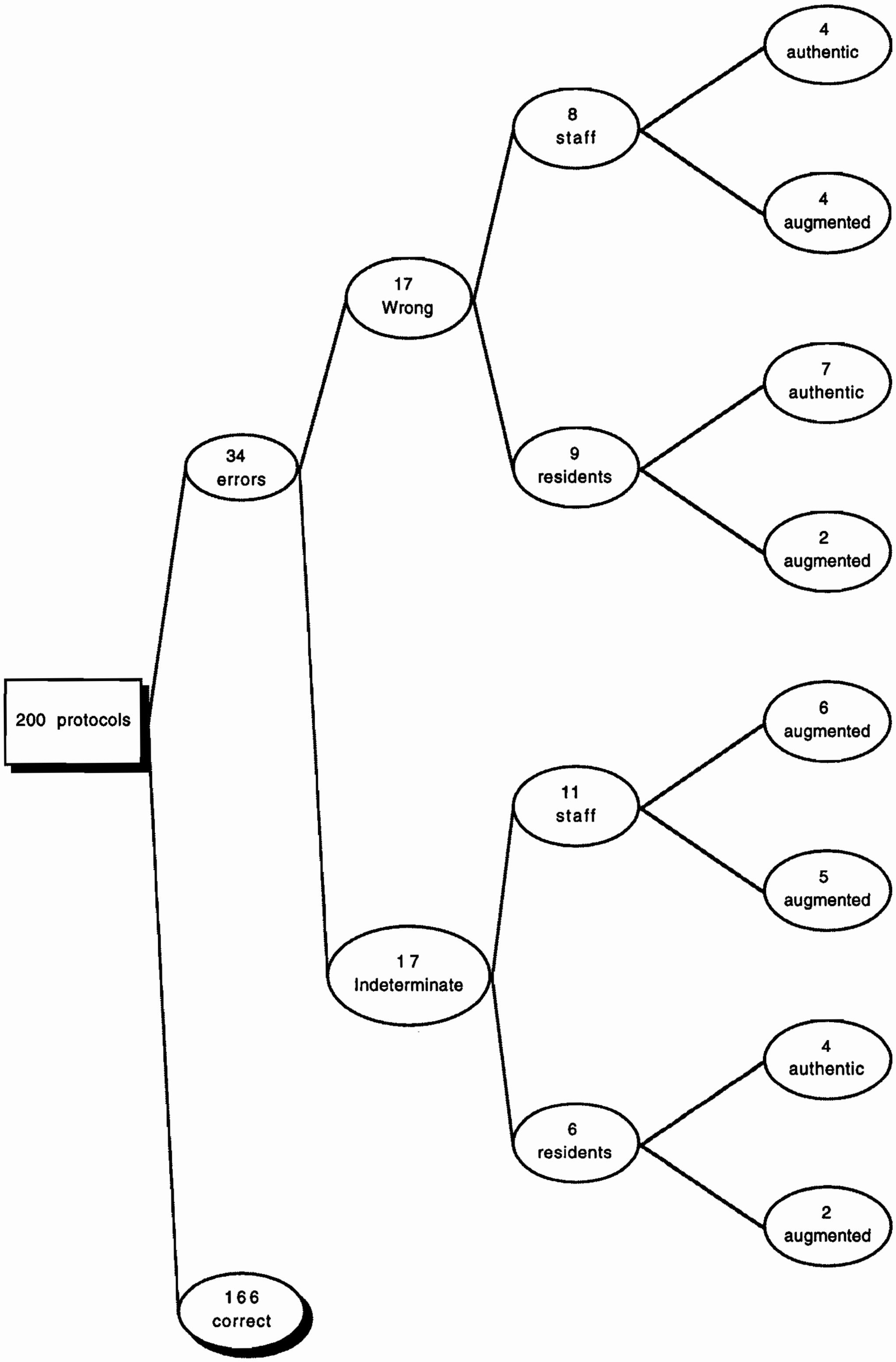



Figure 4.7. Frequency of Overall Accuracy Errors Committed by Levels of Expertise and Experimental Conditions.

equal number of errors regardless of experimental condition, while residents, committed seven errors in the authentic condition and two in the augmented condition.

Of the 17 indeterminate answers, the staff committed 11 and the residents committed six. The staff committed six errors in the augmented condition and five under the augmented condition. The residents, however, committed four errors in the authentic condition and two in the augmented condition.

The staff committed nearly the same number of wrong and indeterminate errors regardless of experimental condition. However, the results suggest that the residents benefited from the highlighting since they committed less errors (both incorrect and indeterminate) when the cases were presented under the augmented condition. For example, they received seven incorrect ratings under the authentic condition (as compared to 2 under the augmented condition), and 4 indeterminate ratings under the authentic condition (as compared to 2 under the augmented condition). These results suggest that the residents were less likely to commit errors when mammographic findings were highlighted. A detailed analysis of the types of errors committed by participants is presented in a subsequent section along with a discussion of the protocol analyses.

## Analysis of Errors Based on Overall Diagnostic Performance

An in-depth analysis of the 34 errors committed by participants based on overall diagnostic accuracy across experimental conditions was performed. The 34 coded protocols were analyzed to extract the medical problem solving elements and examine the nature of the errors. Figure 4.7 presented the frequency of errors committed by the participants according to level of expertise, participant number, case number, experimental condition, overall diagnostic accuracy, and type of error. Five types of errors were noted and each is described below along with the results from non-statistical comparisons based on error type, level of expertise, case number, and experimental conditions.

A **perceptual detection error** was coded when a participant failed to detect a mammographic finding. Perceptual detection is extremely important in diagnosing cases from mammograms since without detecting a mammographic cue, a participant cannot identify, characterize, diagnose and request subsequent examinations (if required). An example of this type of error involved participants' inability to detect a second carcinoma when solving case 9.

A **finding mischaracterization error** was coded when a participant incorrectly characterized a mammographic finding. For example, this error was coded where a participant characterized a mammographic cue as "well-defined retroaereolar mass" when in fact it was partly well-circumscribed. Another example was the characterization of "an abnormality ... distorting the underlying breast architecture associated with some puckering suggesting the presence of puckering" when in fact the mammogram exhibited a density. In another example, a participant characterized round and linear calcifications as "irregular with ill-defined margins" which he then incorrectly diagnosed as suspicious for malignancy. Finding mischaracterization errors are important because they have an impact on the ability to propose a diagnostic hypothesis and subsequent recommendations (if required).

A **no diagnosis error** was coded when a participant detected, correctly identified, and characterized a mammographic finding but failed to make a diagnosis. For example, when solving a classic hamartoma case (case 5) a participant found the mass in the upper outer quadrant and described it as made up of mixed elements, but could not make a diagnosis. Diagnosis level errors are typically associated with a lack of knowledge of diseases and their manifestations and may result in requesting an inappropriate subsequent examination (e.g., excise a hamartoma because it's palpable).

A **wrong diagnosis error** was coded when a participant detected, correctly identified, and characterized a mammographic finding but made a wrong diagnosis. When solving the same hamartoma case, a participant found the mass in the upper outer quadrant made up of mixed elements but incorrectly diagnosed it as a cystosarcoma (a malignant disease). Diagnosis level errors are typically associated with a lack of knowledge of diseases and their manifestations and may also result in requesting an inappropriate subsequent examination.

A **wrong recommendation error** was coded when a participant (a) correctly detected and characterized a mammographic finding, (b) provided a diagnosis at some level of abstraction (which may not have been entirely correct), but (c) suggested an inappropriate subsequent examination. For example, when solving a case of inflammatory carcinoma (case 10), a participant detected the asymmetry, skin thickening, and the reticulated appearance on the breast tissue. However, he requested antibiotic treatment with a follow-up in 6 months. The correct examination would be a biopsy since an inflammatory carcinoma is aggressive and would take the woman's life within six months. Wrong recommendation level errors are typically associated with a lack of knowledge of diseases and their manifestations.

The results presented in Table 4.11 suggest that the type of error committed is case-related. Nearly all participants regardless of expertise level and experimental condition committed the same error when solving cases 4, 5, 8, 9, and 10. For example, a staff radiologist and a radiology resident solving case 1 under both conditions each committed a finding characterization error. The results suggest that the clinical history, and more importantly, the mammographic manifestations are critical in determining the types of errors committed by radiology personnel. For example, only perceptual errors were committed when solving cases 3 and 9, probably because of the difficult mammographic findings exhibited in these cases. Similarly, wrong recommendation errors were committed by all participants across experimental conditions when solving cases 8 and 10, which suggests that the error lies in the determination of the most adequate examination to clarify the mammographic cues.

The frequency and types of errors by level of expertise and experimental condition are presented in Table 4.12. Overall and in descending order of frequency, the results indicate that most errors involved wrong recommendations (38%). These were followed by perceptual detection errors (26%), finding characterization errors (24%), no diagnosis (6%), and wrong diagnosis (6%). As indicated earlier, the staff made slightly more errors when solving cases under the authentic condition (10 as compared to 9). The residents made almost three times more errors when the cases were presented under the authentic condition (11 as compared to 4). The residents also made more perceptual detection errors (5 as compared to 4), and finding characterization errors (6 as compared to 2) than the staff. In contrast, the staff provided more no diagnosis errors (2 as compared to 0), and more wrong diagnosis errors (2 as compared to 0), and wrong recommendation errors (9 as compared to 4). Interestingly, unlike residents, staff did not provide a diagnosis on two occasions, provided the wrong diagnosis on another two occasions, and provided slightly more than twice as many wrong recommendations than the residents (9 as compared to 4).

Table 4.13 illustrates the types of errors committed by levels of expertise based on overall diagnostic accuracy (same results as Table 4.11). This table was created to determine whether error types were associated with individual participants. The results indicate that overall, errors are not attributable to individual participants, but rather to the characteristics of individual cases. The only exception to this general finding was found in the performance of S11 who committed a finding characterization error twice.

Table 4.11

Types of Errors Committed By Participants Based on Overall Correct Diagnostic Accuracy.

| Level of Expertise | Case Number | Experimental Condition | Overall Diagnostic Accuracy Code | Error Type |
|---|---|---|---|---|
| Staff (S5) | 1 | Authentic | Indeterminate | Finding Mischaracterzation |
| Resident (S18) | 1 | Augmented | Indeterminate | Finding Mischaracterzation |
| Resident (S11) | 2 | Authentic | Indeterminate | Finding Mischaracterzation |
| Staff (S2) | 3 | Augmented | Indeterminate | Perceptual Detection |
| Resident (S11) | 4 | Authentic | Wrong | Finding Mischaracterzation |
| Resident (S13) | 4 | Authentic | Indeterminate | Finding Mischaracterzation |
| Resident (S15) | 4 | Authentic | Indeterminate | Finding Mischaracterzation |
| Staff (S6) | 4 | Augmented | Indeterminate | Wrong Diagnosis |
| Resident (S20) | 4 | Augmented | Wrong | Finding Mischaracterzation |
| Staff (S1) | 5 | Authentic | Wrong | No Diagnosis |
| Staff (S5) | 5 | Authentic | Indeterminate | Wrong Recommendation |
| Staff (S7) | 5 | Authentic | Indeterminate | No Diagnosis |
| Staff (S9) | 5 | Authentic | Indeterminate | Wrong Recommendation |
| Resident (S13) | 5 | Authentic | Indeterminate | Wrong Recommendation |
| Staff (S2) | 5 | Augmented | Indeterminate | Wrong Recommendation |
| Staff (S6) | 5 | Augmented | Wrong | Wrong Diagnosis |
| Resident (S20) | 5 | Augmented | Wrong | Wrong Recommendation |
| Staff (S9) | 7 | Augmented | Wrong | Finding Mischaracterzation |
| Staff (S2) | 8 | Authentic | Indeterminate | Wrong Recommendation |
| Staff (S8) | 8 | Authentic | Indeterminate | Wrong Recommendation |
| Staff (S9) | 8 | Augmented | Indeterminate | Wrong Recommendation |
| Staff (S2) | 9 | Authentic | Wrong | Perceptual Detection |
| Staff (S6) | 9 | Authentic | Wrong | Perceptual Detection |
| Staff (S8) | 9 | Authentic | Wrong | Perceptual Detection |
| Resident (S12) | 9 | Authentic | Wrong | Perceptual Detection |
| Resident (S14) | 9 | Authentic | Wrong | Perceptual Detection |
| Resident (S16) | 9 | Authentic | Wrong | Perceptual Detection |
| Resident (S18) | 9 | Authentic | Wrong | Perceptual Detection |
| Resident (S20) | 9 | Authentic | Wrong | Perceptual Detection |
| Resident (14) | 10 | Authentic | Wrong | Wrong Recommendation |
| Staff (S1) | 10 | Augmented | Wrong | Wrong Recommendation |
| Staff (S7) | 10 | Augmented | Indeterminate | Wrong Recommendation |
| Staff (S9) | 10 | Augmented | Wrong | Wrong Recommendation |
| Resident (S13) | 10 | Augmented | Indeterminate | Wrong Recommendation |

Table 4.12

Frequency and Types of Errors by Level of Expertise and Experimental Condition.

| Control Processes | Staff Radiologists | | Radiology Residents | |
|---|---|---|---|---|
| | Authentic Condition | Augmented Condition | Authentic Condition | Augmented Condition |
| Perceptual Detection Error | 3 | 1 | 5 | 0 |
| Finding Characterization Error | 1 | 1 | 4 | 2 |
| No Diagnosis | 2 | 0 | 0 | 0 |
| Wrong Diagnosis | 1 | 1 | 0 | 0 |
| Wrong Recommendation | 3 | 6 | 2 | 2 |
| **TOTAL** | **10** | **9** | **11** | **4** |

Table 4.13

Types of Errors Committed By Levels of Expertise Based on Overall Diagnostic Accuracy.

| | | Case Number | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1A** | **1B** | **2A** | **3B** | **4A** | **4B** | **5A** | **5B** | **7B** | **8A** | **8B** | **9A** | **10A** | **10B** |
| **S** | S1 | | | | | | | ND | | | | | | | WR |
| **T** | S2 | | | | PD | | | | WR | | WR | | PD | | |
| **A** | S5 | FC | | | | | | WR | | | | | | | |
| **F** | S6 | | | | | | WD | | WD | | | | PD | | |
| **F** | S7 | | | | | | | ND | | | | | | | WR |
| | S8 | | | | | | | | | | WR | | PD | | |
| | S9 | | | | | | | WR | | FC | | WR | | | WR |
| **R** | S11 | | | FC | | FC | | | | | | | | | |
| **E** | S12 | | | | | | | | | | | | PD | | |
| **S** | S13 | | | | | FC | | WR | | | | | | | WR |
| **I** | S14 | | | | | | | | | | | | PD | WR | |
| **D** | S15 | | | | | FC | | | | | | | | | |
| **E** | S16 | | | | | | | | | | | | PD | | |
| **N** | S18 | | FC | | | | | | | | | | PD | | |
| **T** | S20 | | | | | | FC | | WR | | | | PD | | |
| **S** | | | | | | | | | | | | | | | |

Key: A = authentic, B = augmented
PD = perceptual detection error, FC = finding characterization error, ND = no diagnosis,
WD = wrong diagnosis, WR = wrong recommendation

**Analyses of Errors Based on the Diagnostic Problem Solving Strategies in Four Cases.**

The following section presents the detailed results of four cases (1, 5, 9, 10) from the analysis of the protocol data. As previously stated, these cases were selected for several reasons. They: (a) represent diverse mammographic manifestations, (b) provide evidence of the use of particular problem solving strategies, (c) demonstrate the influence of case type on scanning and reading times, and (d) allow for identification of particular error types and their associations to case type. Each of these cases is discussed below with reference to a protocol illustrating an error committed by a staff member followed by one illustrating an error committed by a resident.

**Case 1 - Mucin-Producing Carcinoma.** Two protocols (S5 - 1A and S18 - 1B) of a staff and resident are presented below to illustrate the errors for this case (see Table 4.11). First, the problem solving approach taken by the two participants is similar to the ones presented above. The differences in the length of the two protocols stem from the amount of time the resident (S18) spent on characterizing the mass (data exploration) and associating the woman's age, mass size, and lack of associated features with a diagnosis. The finding characterization errors committed by these participants (as discussed earlier in the chapter) stems from the inability to properly characterize the finding as having *partly* well-defined margins (which is suggestive of carcinoma especially in an elderly woman). Furthermore, the participants' diagnoses of benign diseases are correct and in keeping with their characterization of a well-defined mass.

**S5 -1A - staff**

Case 1A. A 60 year-old woman who presented with a mass in the right breast.
Ah both films are somewhat dark.
However, the obvious finding is a large, **well-defined** retroaereolar in the right breast measuring 6 by 4.5 centimeters in diameter.
The edges are **very well defined**
and this is most likely a benign lesion
It could represent a cyst or a fibroadenoma.
I do not believe that this could represent a correlate carcinoma or a medullary carcinoma.
If four old films are available the obvious next step would be to do an ultrasound to alert a cyst if it's not a cyst then the patient could be followed or have a stereotaxic biopsies just to confirm the benign nature of this lesion.
It could represent a fibroadenoma.

**S18 -1B - resident**

Okay we have a history. This is a case 1B. Clinical history. A60 year-old woman presented with a mass in the right breast...
The overall density in the breast, at least in left breast is quite fatty.
The right breast shows a moderate amount of fibroglandular tissue in a large region and it's hard to tell the density.
Additionally, ahm the technique is ah, well the technique, we see the pectoralis muscle quite low down and it's hard to say ah, it's hard to say if we see enough of the breast on the cranial caudal on the left breast.
But the main abnormality is in the right breast where there's a large **well-circumscribed** lesion which appears to have fat around it, or at least it's compressing the tissues around it, so it may be just normal tissue which is compressed.
This lesion is 5 by 4 centimeters, or 5 by 5 centimeters, it may be a little bit smaller on the other dimension..
Ahm... this is a very big lesion.
On the- it doesn't appear to have calcifications within it.
It doesn't have any calcification within it...
It's 60 year-old woman.
I think a mass on a 60 year-old woman could be cystic,
however it looks like a solid and compressed in the tissue.
I ah.... it's a big lesion.
I'm not sure what to do next.
I would ah... it doesn't have any ah- in terms of features of malignancy, in terms of calcifications or spiculation, it doesn't really have anything.
In terms of size, it's not a big lesion.
Ahm.. I'm not sure...
It's not presenting enough to help me with this lesion.
I mean, if it's cystic- I don't know if it's gonna be cystic..
I don't know, I can't do this one.
I'm gonna leave it.

**Case 5 - Hamartoma or Fibroadenolipoma.** There were eight errors committed by staff and residents across experimental groups when solving case 5 (presented in Table 4.11). Interestingly, the staff committed six out of the eight errors. Six staff committed two "no diagnosis" and four "wrong recommendation" errors, whereas two residents committed "wrong recommendation" errors. A junior staff member's (S2) protocol is presented below to illustrate her problem solving process. It also highlights the fact that she engaged in a mixed-strategy problem solving approach which included switching from goal-driven problem solving to data-driven problem solving. The adopted strategy seems to be based on the ability to characterize the finding using the data-driven approach. The goal-driven approach was used to search for typical mammographic signs that may be suggestive of malignancy (e.g., "Let's look at the size and get that out of the way."). However, the participant was unable to diagnose the case. She provided a poor characterization of the finding based on size and components, and this limited her to propose a working hypothesis of a very large lesion (an extremely low level of abstraction hypothesis).

**S2-5B - staff**
Okay.... This is a young patient in her 30s, large mass in the lateral portion of the right breast.
Let's see the outline anyway.
Very large breast.
Let's look at the size and get that out of the way.
Ahm calcium, nothing much at all.
Well to be honest, I'd say **I've never seen one of those** but the thing I can make up is it's a **very large lesion** and quite a nuisance.
It's got a major fatty component.
Ahm I suspect it's one of those **fatty tumors** I learned doing residency which I forgot the name of.
This is one case where I'd get my senior partners to confirm.
I don't see anything else in that lesion.
I suspect that's what's going on.
That one I would need help to make sure before I put that away.
But I suspect that's what it is.

The resident's (S20) protocol below highlights the fact that he adopted a data-driven problem solving approach which led him to propose the benign diagnosis and an incorrect recommendation. Once again, the resident identified and characterized the mammographic finding and its associated components, and checked the rest of the mammograms for other findings. He then summarized the findings, and similar to the staff radiologist, specified a diagnosis at a lower level of abstraction by stating that it was fat necrosis with multiple regions of hemorrhage secondary to a recent trauma. The diagnostic plan to find out more about the recent history of trauma and conduct an ultrasound was not required since it is a benign condition. The resident then proposed a few more benign conditions stating that it might not represent a philloides tumor.

For this case, both the staff radiologist and the resident were unable to specify a diagnosis. This seems \to be associated with a lack of clinical exposure to hamartoma-type benign lesions. Also, the protocol reveals the fact the generic radiological expertise facilitates the characterization of mammographic findings. However, the lack of specific expertise in dealing with hamartoma-type masses hampered the attempt to specify the correct diagnosis and led the resident to specify a set of differential hypotheses instead.

| |
|---|
| **S20-5B - resident** |
| Okay this young patient in her 30s who presented with a large mass in the lateral portion of the left breast............. |
| Hhm..Okay a mammographic study of a 30 year-old woman who presents with a large mass in the left breast demonstrates a **large ahm mixed-density mass** involving the ahm left upper-outer quadrant ahm of the breast extending from the level of the ah chest wall to the sub-cutenous tissues. |
| The ahm.. the skin surrounding this looks slightly thickened. |
| Ahm and the internal quality of the mass itself demonstrates ah multiple density characteristics range anywhere from fat to soft tissue. |
| Ahm there are no significant calcifications within the mass.. |
| And the adjacent ahm breast tissue appears to be compressed giving it a more dense appearance relative to the ahm right breast. |
| On the right breast there's a tiny chunky calcification in the ahm upper-outer quadrant and no significant ahm abnormality seen within. |
| Fibro, fatty breasts that are not exceedingly dense. |
| Ahm to conclude there's a **large ah mixed-density mass** in the ah upper-outer quadrant of the left breast but in the chest wall extending almost outward to involve the skin with what I think to be some involved skin thickening. |
| This ahm could represent ahm, with the mixed-density component to it, I'd have to wonder whether this represents some **fat necrosis with ah multiple regions of hemorrhage within it, secondary to a recent trauma**.. |
| The next test I would do is find out if there was a recent *history* of trauma and ah what the chronicity of this ah event was. |
| And ahm if it didn't fit for a trauma I would then consider *ultrasounding* the patient and ah considering something like a **giant fibroadenoma**. |
| Although there's no calcifications which would make me think less of that. |
| Ahm a **walled-off abscess** is a possibility again. |
| And **I don't think this represents a philloides tumor**. |
| It's not solid enough and uniform enough for me to think that. |
| But it could. |

**Case 9 - Two carcinomas.** Eight perceptual detection errors were committed by staff and residents when this case was presented under the authentic condition. As indicated earlier in Table 4.11, the staff radiologists committed three errors while the other five were committed by the residents. The protocol of a staff radiologist (S2) who committed a typical perceptual detection error is presented below and it highlights the fact that she used a mixed-strategy problem solving approach. She began with a data-driven approach as she read the clinical history (data acquisition), identified and classified a mammographic observation (data classification), and then identified the carcinoma (data examination). Then she employed a goal-driven approach ("just making sure there is nothing else") and proceeded to search for the patient's age by re-reading the clinical history. Next, she summarized the information relating to the finding (repeats data), and then provided a diagnosis (hypothesis generation) and a diagnostic plan (i.e., biopsy). The perceptual detection error stems from the failure to detect a mammographic finding as illustrated in the protocol provided below.

**S2-9A - staff**
68 year-old woman with palpable abnormality in the lateral portion of the left breast.
Figure out which one that is.
That's the left.....
That's the right views...
This is CC of the left and two for right with the marker on it.
Left breast...
Benign calcification on that side with vascular calcifications here
Actually that's the left, I'm not sure I have the right...
Okay... starting with **abnormality in the upper lateral**
**Just making sure there's nothing else first**...
Patient is 68..
So, what I have is a **very irregular stoich-like nodule** on the upper right quadrant, given the palpable finding.
It look like **cancer** and needs a **biopsy**
If it's palpable clinically they don't need us
I that's all we do on that.

The protocol of a resident (S20) who committed a perceptual detection error similar to the staff's is presented below. This protocol demonstrates that he also used a mixed-strategy problem solving approach. He began with a data-driven approach: reading the clinical history (data acquisition), immediately providing a diagnosis of carcinoma (hypothesis generation) for one of the lesions, re-reading the clinical history (data exploration), examining the margins of the lesion (data exploration), repeating the diagnosis (summarization), and proposing a surgical excision with needle localization (diagnostic planning). The participant then switched to goal-driven behavior ("any other masses in this breast?") and searched the mammograms, stating that there were no other areas of interest.

**S20-9A - resident**
68 year-old woman who has a palpable abnormality in lateral portion of the left breast, upper outer..........
That's a **carcinoma**. Exaggerated cranial caudal. I don't like this. This over here- these two images over here are enough to make a diagnosis on.
These are all views of the ah left ah left breast in this 68 year-old woman with a palpable abnormality in the upper-outer portion of the left breast that shows a spiculated calcified mass that represents **carcinoma** until proven otherwise.
The next step with this mass would be ahm **needle localization, surgical excision** with axillary node the section for staging.
That's it.
**Any other masses in this breast?**
There may be something else... Hhm...
No. I'm gonna pass it.
There's always one problem I'm not getting to.

**Case 10 - Inflammatory Carcinoma.** Five wrong recommendation errors were committed by staff and residents across experimental conditions. As indicated earlier in Table 4.11, the staff committed three errors while diagnosing the case under the augmented condition. The residents committed two errors, one under each experimental condition.

The protocol of a staff radiologist (S9) who committed a wrong recommendation error is presented below. He began by reading the clinical history (data acquisition), and commenting on the technical aspects of the mammograms (data assessment). He then identified two of the radiological findings (data examination), and proposed two diagnoses (hypothesis generation) and a diagnostic plan. The plan was comprised of antibiotic therapy for an abscess, and clinical re-examination followed by an ultrasound if a palpable lump was felt. The wrong recommendation error is based on the fact that even with a set of potentially plausible differential diagnoses, he failed to request a biopsy despite the fact that the disease was a progressing carcinoma with a poor prognosis.

**S9-10B - staff**
Next case. A 60 year-old woman presented with inflammation of the left breast for two week.
No previous examination is available for comparison.
Ah the density of these ah examinations and the contrast is somewhat low
but we do note some **asymmetric ah thickening** of the skin along the left breast....
On a second . . . we note slight increased ah diffuse density in the left breast I think which is partially technical but there may be slight increased amounts of ah density in the sub-aereolar portions ah..
**I do not see clearly defined abscess** but these may represent changes to the skin or ah deep ah inflammation such as **cellulitis or mastitis**..
The breasts are predominantly fatty and I do not appreciate a significant lesion underneath.
Ah **I don't really think this patient needs follow-up mammograms**.
Ah she may require appropriate **antibiotic therapy**, or other appropriate treatments
but ah I don't see evidence for an abscess.
Of course should she present with a palpable tender lump or something like that, she can always be re-examined.
But I recommend ah **ultrasound** at that point.

The protocol of a resident (S14) who committed a wrong recommendation error is presented below to illustrate his mixed-strategy driven problem solving process. The participant began with a data-driven process approach which terminated after the identification of a mammographic observation ("...she has pretty fatty breasts"). Subsequently, a goal-driven approach was assumed, which is characterized by goal specification followed by mammogram search cycles. The three goals have been highlighted (in bold and italics) to illustrate the pattern in the participant's problem solving process. For example, the participant set a goal to search for skin thickening but failed to identify the skin thickening in the left breast. She then reverted to a data-driven approach and identified and classified calcifications in both breasts. This was followed by specification of another goal to investigate the mammographic cues which may have suggested the causes of inflammation. However, nothing was detected that suggested the underlying causes of the inflammation. The participant failed to specify a hypothesis of malignant or benign. This was followed the subsequent scanning of the mammograms. Once again, the symmetry between breasts and the lack of calcifications were not suggestive of carcinoma. Next, the participant set another goal, to inspect the right breast for evidence of a lesion. She then proposed a diagnostic plan to compare the mammograms to previous ones, to re-examine the patient, and to have the patient return in three to four months for a follow-up mammogram. The wrong recommendation error is based on the fact that the participant failed to detect the mammographic findings, did not specify any plausible differential diagnoses, and more importantly, failed to request a biopsy despite the serious nature of and poor prognosis for this case.

**S14-10A - resident**
Okay so this is a 60 year-old woman who presented with inflammation of her left breast for two weeks.........
Okay so we're provided again with a CC and MLO views of both breasts and this woman has inflammation of her left breast for two weeks.
So first of all, let's take a quick look.
She has pretty glandular, ah sorry she has pretty fatty breasts.
She's 60 year old and... ***just want to see if there's any skin thickening***... which I don't really see.
We see that the left breast looks a little bit- I don't think it looks that much worse than the left. We're not positioning it that well with the intramammary wall is just pushed up a little bit.
There's some areas of benign, completely benign calcification in both breasts.
**Nothing that I would do anything about**...........
I don't really see any gross abnormality in these breasts.
***Ah usually people with ah breast inflammation could have like ah- they could have some areas of increased density.***
And grossly looking at the two breasts many there's a little bit more increased density in the left breast but not enough that I would be worried about.
Ah there's ***certainly nothing take looks malignant or suspicious*** in this breast.
Ah the retro-areolar tissue is pretty symmetrical in both breasts.
There's no areas of microcalcification,
and there's no area that looks like there's secretory calcifications or evidence for plasma cell mastitis...
I'm not sure why she has come for this mammogram........
***I'm just looking to make sure there's no lesion anywhere else in the right breast..***
Again some benign looking microcalcifications.
Nothing that looks too worrisome.
I would probably with anybody, she's 60 years old, I would probably **get some old mammograms compare them**, see if there's any change.
I'm sure she had them before.
Ahm then I would ask her if she felt a palpable mass or see if she felt any areas of lumps or bumps and go back and examine those areas.
But right now there's really nothing here that worries me.
I probably would tell her to come back in 3 or 4 months time, just for a **follow-up mammogram**, after she felt better.
But ah I don't see anything grossly abnormal here.
I would **compare them with a previous and ah examine the patient.**

## Summary

A seven step model of mammogram interpretation has been presented. Repeated measures ANOVAs revealed (a) that staff radiologists scanned the cases significantly faster than residents with no significant main effect for condition and no interaction, and (b) no differences between groups in reading time across experimental conditions. Non-statistical comparisons of means revealed no group differences in the number of radiological findings, radiological observations, and number of diagnoses across experimental conditions. Frequency analyses revealed that both groups regardless of experimental condition (a) used the same types of operators, control processes, diagnostic plans and goals, (b) committed the same number of errors, and (c) the errors committed were case-dependent. An additional analysis failed to reveal a significant correlation between the number of total mammograms diagnosed and the number of correctly diagnosed cases. Analyses revealed that mammography interpretation was characterized by a predominant use of data-driven or mixed strategies depending on case typicality and clinical experience.

# CHAPTER 5

# DISCUSSION

The results of this study are presented in the context of (a) the model of mammogram interpretation, (b) the use of different problem solving strategies, operators, and control processes during mammogram interpretation, (c) interpretation of several performance variables across groups and experimental conditions, and (d) error analyses. The theoretical and instructional implications of the study are also presented in the form of a follow-up study and a conceptual framework for the development of a computer-based environment for training residents to interpret mammograms. The contributions and limitations of the study, and conclusions are also presented.

The purpose of this study was to investigate the problem solving strategies used by staff radiologists and radiology residents during the process of interpreting difficult breast diseases depicted on mammograms. The results are discussed in the same sequence in which they were presented in the *Results* section. This chapter is divided into four sections, each dealing with a particular research question and its corresponding hypotheses and discussing the results in lieu of previous research. The *first* section presents a cognitive model of problem solving in mammogram interpretation based on a content analysis and the analysis of verbal protocols. The *second* section presents in-depth analyses of protocols from several cases in order to exemplify typical staff radiologists' and radiology residents' problem solving strategies. This section attempts to answer the question of whether staff radiologists and radiology residents use different problem solving strategies, operators, and control processes during mammogram interpretation. The *third* section presents frequency analyses, comparisons of descriptive statistics, non-parametric analyses, and inferential statistics. These analyses were performed to investigate whether level of expertise and experimental conditions had an effect on several performance measures (number of mammogram findings, observations, and diagnoses, scanning and reading time, accuracy ratings for diagnosis, subsequent examination, and accuracy ratings for overall diagnostic accuracy). The *fourth* section presents frequency analyses, describes the types of errors committed by participants, and provides results of in-depth analyses of protocols. These analyses were performed to investigate whether level of expertise and experimental conditions had an effect on the number and types of errors committed.

## Cognitive Model of Mammogram Interpretation

The cognitive model of diagnostic problem solving in mammogram interpretation was constructed from the content analysis and refined based on the verbal protocol analyses. Decomposition of the complex task of mammogram interpretation resulted in a model consisting of seven steps. These steps include: (a) reading a clinical history, (b) placing a set of mammograms on a viewbox and identifying individual mammograms in the set, (c) visually inspecting each of the mammograms either with or without the use of a magnifying glass, (d) identifying mammographic findings and observations, (e) characterizing mammographic findings and observations, (f) providing a definitive diagnosis or a set of differential diagnoses, and (g) specifying subsequent examinations.

The model allows for a "linear approach" (from reading the clinical history to specifying subsequent examinations) or an “iterative approach” in which the results of a step may feed back to previous steps in the model. The linear approach is indicated by the downward pointing arrows (see Figure 4.1), and can be exemplified by the use of a data-driven problem solving approach whereby a participant reads the clinical history, scans the set of mammograms, and provides a diagnosis. The model also comprises a certain amount of iteration as indicated by the vertical and horizontal arrows on the right side (see Figure 4.1). For example, after positioning the mammograms on the viewbox the radiologist may refer back to the clinical history to determine the location of the suspicious finding. Similarly, following the characterization of the mammographic findings the radiologist may re-scan the set of mammograms with a magnifying glass and identify new findings. It is important to note the model (see Figure 4.1) presents the idealized mammogram interpretation mode. As such, several steps may be omitted by a participant while using either the linear and iterative approach.

This model does share some features with those proposed by other researchers. For example, Lesgold and colleagues' (1981; 1988) model of diagnostic processing involves five components including film abnormality detection, abnormal feature characterization, anatomical localization, medical explanation, and overall case resolution. Their first (film abnormality detection) and second components (abnormality feature characterization) are similar to the model's fourth (identification of mammographic findings and observations) and fifth (characterization of the mammographic findings and observations) steps. The identification/detection and characterization steps are critical in radiology. It is important to note that the differences between the models are due to the fact that two different radiology sub-specialties (chest radiography and mammography) with different diseases and radiological manifestations are being compared.

The model corresponds more closely to Faremo's (1997) model of breast disease derived from the analysis of problem solving protocols of medical students and surgical residents. Faremo's model is comprised of five steps including (a) review of case history, (b) scan mammograms and identify findings (if any), (c) characterize findings and provide diagnosis/differential diagnosis, (d) order additional tests (if required), and (f) recommend an intervention for each finding that needs one. There is a certain amount of overlap between the models, and there are several reasons that explain the similarities and slight discrepancies between the two models. The similarities are based on the fact that Faremo's (1997) study also examined breast disease and the participants for both studies came from the same teaching hospital system. The slight discrepancies between the two models are based on the fact that Faremo (1997) focused on undergraduate medical students and surgical residents while the model constructed in this study was based on the investigations of staff radiologists and radiology residents.

More generally, the model is also consistent with Ramoni and colleagues' STModel (Stefanelli & Ramoni, 1992), which is a generic model of medical diagnosis. Similar to the mammogram interpretation model, the STModel includes a certain "linearity." More importantly, it also comprises iterativeness in the form of a process of proposing candidate solutions that cover symptoms and evaluating them against new information. Further, the STModel characterizes hypothesis generation and hypotheses into four basic inference types: abstraction, abduction, deduction, and induction. According to Patel and Joseph (1990), hypothesis generation is based on the given data and the diagnostician moves "forward" to identify a set of candidate hypotheses able to account for the situation. Hypothesis testing occurs after each evoked hypothesis is evaluated on the basis of its expected consequences. Therefore, the hypothesis drives the process of gathering data. However, unlike other more complex medical specialties such as nephrology, the experienced practitioner in mammography interpretation very rarely works with multiple hypotheses (i.e., set of differential diagnoses).

The mammogram interpretation model shares a few other similarities with the STModel. First, the abstraction phase results in identification of the relevant features of the problem. In the mammogram interpretation model, this is accomplished by: extracting relevant clinical history cues and mammographic cues, and identifying and characterizing mammographic findings and observations. During the abstraction process, data are not only filtered according to their relevance for the problem solution, but they are also chunked in schemata representing an abstract description of the problem at hand (Patel & Ramoni, 1997). Next, the hypotheses that can account for the current situation are related through a process of abduction, thus accomplishing the final part of the STModel's generation phase. This is similar to the sixth step in the mammogram interpretation model in which a definitive diagnosis or a set of differential diagnoses is provided. In the hypothesis testing phase, hypotheses are incrementally tested based on the extent to which they can account for a problem. This involves the inferencing process of deduction. A hypothesis is found to be "correct if the world following from it is the same as the most similar world to the real one" (Patel & Ramoni, 1997, p. 75). Based on the expected consequences requests for new information are frequently made which add to the initial case information. This is equivalent to the last step in the mammogram interpretation model which includes specifying subsequent examinations. However, it should be noted that in mammography, this step is used when there is a lack of clinical information and additional tests (e.g., biopsy or ultrasound) are required. The results of these tests are used to either narrow the number of diagnoses, identify a better candidate or achieve a definitive diagnosis. According to the STModel this is accomplished through the process of induction. The major feature of induction is the ability to falsify (i.e., rule out) those hypotheses whose expected consequences turn out to be in disagreement with the patient problem.

In summary, the seven step mammogram interpretation model shares many features with other diagnostic process models. However, it also has certain distinctive features which are domain specific. The shared features are based on the notion that any model of diagnostic reasoning, regardless of domain, typically involves certain generic phases (e.g., hypothesis generation and hypothesis testing). However, domain specificity also has an impact on the diagnostic process by stipulating unique steps involved in interpreting mammograms (e.g., characterizing mammographic findings and observations). Lastly, this seven step model has implications for learning and instruction in radiology and is presently being used to define the instructional sequencing for the mammogram interpretation tutoring environment.

As such, the emphasis on certain components of the "model of mammogram interpretation" is based on the (a) results of the content analysis, (b) level of granularity used for analyzing the verbal protocols, (c) experience level of participants (e.g., radiology residents versus staff radiologists), (d) methodological approaches (repeated probing of participants versus naturalistic problem solving task), (e) analytical methods used to construct the model, and (f) objectives of both the training program for each level of medical specialty (e.g., the objectives of mammography for surgical residents is non-existent while the residency training objectives for radiology residents is extensive) and their governing body (e.g., Canadian Association of Radiologists compared to the Radiological Society of North America).

## Problem Solving Strategies, Operators and Control Processes Used by Staff Radiologists and Radiology Residents During Problem Solving

The second research question investigated whether staff radiologists and radiology residents used different problem solving strategies, operators, and control processes during mammogram interpretation. It was hypothesized that staff radiologists would solve the cases using a data-driven problem solving approach and that the radiology residents would solve the cases by using either a goal-driven or a mixed strategy (combination of data-driven and goal-driven problem solving strategies). This question was investigated by performing in-depth analyses of protocols from several breast disease cases to exemplify typical staff radiologists' and radiology residents' problem solving strategies.

### Problem Solving Strategies

The evidence from this study does not wholly support the hypotheses that radiologists would solve the breast disease cases using a data-driven problem solving approach and that residents would solve the cases by using either goal-driven or a mixed-strategy. Instead, the results indicate that both staff and residents make use of data-driven and mixed problem solving strategies. Overall, the in-depth protocol analyses indicated that diagnostic reasoning in mammography is characterized by (a) the predominant use of data-driven diagnostic strategies, (b) the use of goal-driven strategies or a combination of both strategies depending on case typicality and clinical experience, and (c) rapid schema-based problem-solving which facilitates search, the characterization of mammographic features, integration of clinical history cues, and accurate diagnoses and subsequent recommendations.

The results can be interpreted in the context of other research investigating problem solving strategies (Lesgold, 1988; Lesgold et al., 1981; Lesgold et al., 1988; Patel & Groen, 1986; VanLehn, 1989; VanLehn, 1996). The use of data-driven problem solving is apparently based on extensive knowledge of a domain (Chi & Glaser, 1988). This knowledge permits rapid recognition of relevant features and rapid triggering of a schema which circumvents problem understanding and problem solving search. Therefore, experts have been found to use data-driven problem solving while novices use a combination of problem solving strategies given their limited experience in mammogram interpretation.

In this study, the problem solving strategies used by participants, regardless of experience, was a function of a few factors. First, the predominant use of a data-driven approach by all participants was directly related to the visual nature of mammogram interpretation. A clinical history is typically not sufficient to diagnose a case. For example, even in cases where the clinical cues suggest malignancy (e.g., family history, patient's age, previous history of malignancy) a diagnosis of a benign condition can occur. Therefore, participants rely on a data-driven (or bottom-up) approach by identifying and characterizing the mammographic findings and observations and proposing a diagnosis.

Second, the variations encountered in the problem solving approaches are more indicative of mammography experience than any other factor. For example, a staff radiologist with extensive mammography experience would read the clinical history, scan the mammogram and immediately propose a diagnosis. This is very characteristic of schema-based problem solving. In contrast, a resident with limited mammography experience may adopt the same data-driven approach but spend more time identifying, localizing, and characterizing each mammogram findings before providing the same diagnosis (than the expert).

Third, the use of a goal-driven problem solving approach by both groups indicates the use of strategic knowledge. Interestingly, both groups used a data-driven approach until either a few or no hypotheses were generated. Subsequently, they reverted to goal-driven problem solving to set goals and "segment the problem space" into a manageable size and search for specific mammographic signs of malignancy. For example, goals such as "I'm looking for an ill-defined mass", "I'm inspecting the mammograms for clusters of microcalcifications", etc., are examples involving specific domain knowledge.

The patterns of problem solving strategy used by the participants are indicative of the domain and several other factors. The visual aspect of mammogram interpretation basically forces a participant, regardless of experience, to use a data-driven approach to identify, localize and characterize the mammogram findings and observations and propose a diagnosis. Therefore, the nature of that task dictates the type of strategy that will be used by participants. This is comparable to studies in domains such as writing and computer programming where it has been shown that both experts and novices solve problems by using a goal-driven approach. This leads one to question the generalizations that are typically made between novices and experts and their "typical" problem solving approaches. In sum, it points to the nature of the domain and the specific task that participants are required to solve.

The results also provide conflicting evidence vis-à-vis the results of expertise studies in other domains. They challenge the operational definitions of expertise (including specific and general) used by various researchers (e.g., Patel & Groen, 1991). For example, in some studies of medical cognition experts are represented by physicians solving cases in their domain while the comparison group is comprised of sub-experts (i.e., physicians solving cases outside of their domain). So the question becomes - If one classifies staff radiologists as experts due to their extensive training in mammography, then what classification is reserved for residents? It is difficult to classify them based on the operational definition provided by Patel and Groen (1991) regarding the expert-novice continuum. They are not experts because an expert "is an individual with specialized knowledge of the domain" (Patel & Groen, p. 96). The residents may actually fall into the subexpert category which includes individuals with generic knowledge, but inadequate specialized knowledge of the domain. This classification may account for both groups using a data-driven approach and sometimes reverting to a goal-driven one. It may be that both groups have general expertise comprised of radiological knowledge which may include generic problem solving approaches to interpreting mammograms, CT scans, MRI images, etc. In addition, they may also have specific expertise with certain disease categories which also allow them to not only use data-driven approaches but to make a diagnosis soon after placing the mammograms on the view box. So the question then becomes - Why and when is a goal-driven approach used? Apparently, this approach is used when the participant has used a data-driven approach and either characterized all the findings and failed to propose a diagnosis or has proposed numerous diagnoses and cannot rule out any of them or cannot identify and characterize any of the mammographic features. They may have used a data-driven approach as far as possible and then adopted goal-driven approach due to the lack of specific radiological expertise with certain disease categories.

**Problem Solving Operators**

The second research question also investigated whether staff radiologists and radiology residents used different problem solving operators during mammogram interpretation. A definitive hypothesis was not posited since literature in this area of medical diagnosis is lacking. Overall, based on frequency analyses the results indicate that both groups used the same types of problem solving operators regardless of experimental conditions.

Overall, the results indicate that residents used more operators than the staff. Both groups used more operators when solving cases presented in the augmented condition. An analysis of the frequency of operator use by staff and residents during the diagnostic process in cases presented under both experimental conditions revealed a predominant use of the following operators (listed in order of descending frequency): (a) data examination, (b) data acquisition, (c) data exploration, and (d) hypothesis generation. These four operators account for 76% of all operators used by the staff in solving cases under both experimental conditions, 79% of all operators used by residents solving cases presented under the authentic condition, and 72% of all operators used by residents solving cases presented under the augmented condition.

The fact that the residents used more operators than the staff corresponds to the greater length of time required for residents to solve the cases. The fact that both groups used more operators when solving cases under the augmented condition suggests that the highlighting of mammographic findings distracted the participants, forcing them to spend more time considering the mammographic findings (especially if they were not part of a participant's diagnostic protocol). In general, participants' predominant use of the same clusters of operators (data examination, data acquisition, data exploration, and hypothesis generation) provides further evidence that both groups used the same operators. The results suggest that the domain of mammogram interpretation is well-constrained since participants with varying levels of mammographic experience used the same four clusters of operators. It is important to note once again, that the use of data examination, data acquisition and data exploration clusters of operators are also very indicative of the data-driven (i.e., bottom-up) approach used in the generation of a hypothesis. In contrast, Rogers' (1992) model incorporates problem solving operators that allow for both data-driven (bottom-up) and goal-driven (top-down) problem solving. It is important to note that the use of both problem solving operators is critical is the diagnosis of chest x-rays that exhibit more complexity in the "constellations" of findings and anatomical structures than mammography.

Further, the experimental conditions did not have an impact on the problem solving processes as indicated by the use of operators. However, these four clusters of operators clearly indicate that participants used these operators to acquire as much information from reading the clinical history, examining the mammograms, identifying and characterizing mammographic features, and proposing diagnoses. These four clusters are themselves indicative of and support the earlier findings of the predominant use of data-driven approach in mammogram interpretation.

## Control Processes

The second research question also dealt with whether staff radiologists and radiology residents used different control processes during mammogram interpretation. Similar to the use of problem solving operators, a definitive hypothesis was not posited since research in this area of medical diagnosis is lacking. Results based on frequency analyses indicate that the use of control processes is based on level of expertise and experimental conditions.

Overall, the results indicated that staff used slightly more control processes than the residents. However, the minimal differences in the frequencies across groups and experimental conditions indicates that the performance of the two groups is very similar. Overall, the results suggest that diagnostic plans are used more frequently than any other problem solving control process. In contrast to complex domains such as technical writing (Bracewell & Breuleux, 1993; Scardamalia & Bereiter, 1991) where goals and plans are characteristic of the domain, the use of diagnostic plans is a characteristic of mammography since cases suggesting malignancy and sometimes even benign cases warrant further examinations (by specifying diagnostic plans). Also, the fact that none of the residents used goals while staff sometimes did must be interpreted cautiously since the results presented here are based on a sample of the 200 protocols. This does not suggest that staff consistently engage in more goal-driven problem solving than residents.

## Radiological Findings, Observations, and Diagnoses

The third research question investigated whether expertise and experimental condition effects performance on several measures. These measures included: (a) number of mammogram findings, (b) number of observations, (c) number of diagnoses, (d) scanning time to construct an initial mental representation based on the clinical history and set of mammograms, (e) reading time required to solve each breast disease case, (f) accuracy ratings for diagnosis, (g) accuracy ratings for subsequent examination, and (h) accuracy ratings for overall diagnostic accuracy.

The first hypothesis stated that, compared to the authentic condition, the augmented condition would facilitate residents' ability to identify findings and observations, provide fewer diagnoses, decrease their scanning and reading time, and increase their ratings on accuracy (for diagnoses, subsequent examinations, and overall diagnostic accuracy). The planned statistical comparisons were not conducted given the minimal differences between the means and standard deviations. The non-statistical comparisons of the means suggests there are no differences between the mean number of radiological findings, observations and diagnoses between the groups across the two experimental conditions. The results did indicate that participants identified at least one radiological finding, made three radiological observations, gave one diagnosis per case, and tended to identify more radiological observations than radiological findings.

The second hypothesis stated that, in comparison to radiology residents, staff radiologists would be faster in scanning and reading mammograms, and have better ratings for accuracy (for diagnoses, subsequent examinations and overall diagnostic reasoning regardless of experimental condition). This hypothesis was partially confirmed in that the staff radiologists were significantly faster than the residents in scanning the mammograms. However, the results did not reveal that the staff radiologists were significantly faster in reading the mammograms nor did they achieve better accuracy ratings for diagnoses, subsequent examinations or overall diagnostic reasoning.

There are several possible explanations for the disparity in the results related to these two hypotheses. The minimal differences in the number of radiological findings, observations, and diagnoses can be interpreted in the context of the nature of the domain of mammography. In comparison to chest radiography, mammography is a well-constrained sub-specialty of radiology limited to one anatomical region and with substantially less mammographic "constellations" exhibiting breast disease. Furthermore, in mammography there are fewer disease classifications with relatively fewer findings and observations. Further, the cases used in the study presented anywhere from one to four radiological findings, and anywhere from zero to three radiological observations. This may have resulted in a ceiling effect, limiting the number of these factors identified by all participants regardless of level of expertise and experimental condition.

Another explanation may lie in the fact that residents (given their limited experience in mammography) are still exposed to monthly mammography teaching rounds during their five year training. The inclusion of seven fifth-year residents who were studying for their radiology certification examinations during the data collection phase may also have affected the results in that they may be considered to be have been "approaching expertise in mammography". In contrast, the study included a few staff radiologists who did not diagnose breast disease cases on a regular basis. These problems reflect the difficulty in recruiting enough participants since there are very few staff and residents with mammographic experience in the McGill University teaching hospital system.

The fact that staff were significantly faster at scanning the cases than the residents suggests that they were able to take advantage of their vast experience in mammography and attain a representation of the problem faster than the residents. This finding is supported by other research findings which suggest that experts are faster at abstracting the relevant features of a problem. However, both groups took longer to scan mammograms presented under the augmented condition. The highlighted mammographic findings may have distracted participants during the problem solving process. An explanation may lie in the fact that even though they highlight film findings for other physicians they do not do it for themselves. Another explanation may be based on the concept of integrating "loose ends." "Loose ends" are clinical cues that have not been integrated by a participant into their overall diagnostic process. As such, participants may have taken slightly longer in the augmented condition to integrate "loose ends" which were not part of their diagnostic process. Lastly, it must also be noted that scanning and reading times were affected by the number and size (e.g., visible to the naked eye) of mammographic findings and mammographic observations, and number of mammograms in the case.

Highlighting mammographic findings may be effective in eliciting more accurate diagnoses, radiological recommendations and providing overall diagnostic accuracy especially for radiology residents. The results seem to indicate that the highlighting of mammographic features has provided the residents with a kind of "perceptual scaffolding" which might trigger a diagnostic schema. Hypothetically, this scaffold may support the lower-level perceptual processes thus allowing them to concentrate on the characterization and classification of the mammographic findings, diagnosis, and radiological recommendations. Perhaps the use of even more difficult cases (i.e., atypical presentations) would be enough to actually elicit expert-novice differences.

The ratings indicate that the groups performed equally well based on the all these measures. The minimal differences in the frequency of correct diagnoses across groups and experimental conditions indicates that the performance of the two groups is very similar. This finding is explained by the fact that mammography is a well-constrained sub-specialty of radiology limited to one anatomical region and a few dozen diagnoses and half a dozen subsequent recommendations. Further, the levels of abstraction in diagnostic hypotheses are not considered important in mammography which may also have contributed to similar performance between the two groups.

## Error Analyses: Frequency, Types, and Protocols Analyses

The fourth research question aimed at investigating the effects of the authentic and augmented experimental conditions on the frequency and types of errors committed by both groups. It was hypothesized that, in comparison to radiology residents, staff radiologists would commit fewer errors regardless of experimental condition. It was also hypothesized that residents would make fewer errors when solving cases presented under the augmented condition. These hypotheses were investigated by (a) conducting in-depth analyses of protocols from these cases, (b) identifying errors, and (c) classifying errors committed by participants during the problem solving task.

Of the 200 protocols, 34 (17%) errors were committed by staff and residents across the two experimental conditions. This indicates that a few errors were committed by both groups across experimental conditions. The staff committed nearly the same number of wrong and indeterminate errors regardless of experimental condition. However, the results suggest that the residents benefited from the highlighting since they committed less errors (both incorrect and indeterminate) when the cases were presented under the augmented condition. This finding is extremely important for the design of the RadTutor (Azevedo, Lajoie, Desaulniers, Fleiszer & Bret, 1997), a mammography interpretation tutoring environment. More specifically, highlighting has important instructional implications and is presently being used as a form of "perceptual scaffolding" in the RadTutor.

Five types of errors were revealed by the in-depth analyses of the protocols. The results indicate that the type of error committed is case-related and that they were not based on level of expertise nor influenced by the experimental condition. The errors committed by the participants in this study are nearly identical to types identified by Rogers' (1992). For example, her detection and labeling errors are identical to the perceptual detection finding mischaracterization errors found in this study. Rogers' integration errors (involving the detection and correct labeling of an abnormality but failing to use it in the generation of a diagnostic hypothesis) reveal the complexity of chest radiography and the working memory overload that results when solving cases with multiple anatomical regions, findings, observations, and differential hypotheses. Due to the differences in radiology subspecialties, this type of error was not encountered in this study. Also, the other three errors that were encountered (no diagnosis, wrong diagnosis, wrong recommendation) indicate a lack of familiarity with specific types of breast disease.

The in-depth analyses of the problem solving protocols containing errors indicate that committing errors was related to a breakdown in data-driven problem solving approach. Furthermore, similar to Patel and colleagues (1994), the in-depth analysis of the 34 protocols revealed that errors were associated with the use of a mixed problem solving strategy. Interestingly, the error protocols revealed that participants used a mixed-strategy approach and not a pure data-driven approach. This finding has several implications relating to the role and structure of schemata. It suggests that the participants were not able to instantiate a schema. Furthermore, this finding has instructional implications and is presently being incorporated into the RadTutor as one major source of error.

## Limitations of the Study

This study has a number of limitations. One limitation was sample size which did not allow for an adequate test of all of the hypotheses since statistical power would have been violated. Instead, frequency analyses were conducted and comparisons between descriptive statistics between groups and conditions on most of the dependent variables were performed. Nevertheless, some of the results replicate some findings from previous studies in the areas of problem solving, expertise, medical cognition, and radiological diagnosis. A second limitation was the lack of additional expert mammographers. The incorporation of at least one other expert in rating the participants' accuracy of diagnoses, subsequent recommendations and overall diagnostic accuracy would have increased the reliability of the expert's ratings. A third limitation relates to the measurement of scanning and reading times which ideally would have been collected using the appropriate eye-movement apparatus. However, this equipment is extremely expensive and was not available for use within the university facilities, nor were there any funds available for purchasing such equipment. However, this issue will be considered in a follow-up study. A fourth limitation (also a criticism of present analytical tools for cognitive scientists) was the lack of sophisticated multimedia software for storing, analyzing, coordinating and coding verbal and video data. Such tools have the potential to accelerate the laborious process of analyzing protocols, to present data, and to teach others protocol analysis. A fifth limitation involved the relatively small numbers of radiological findings, observations, and diagnoses associated with the cases. This limitation may have caused a ceiling effect and thus masked any expertise effects on these three dependent measures. However, mammography is a relatively well-constrained domain and the radiological manifestations presented in the cases do generally represent "typical" cases. A sixth limitation relates to the generalizability of the results which, strictly speaking, directly apply to the visual medical domain of mammogram interpretation. The last limitation is the research methodology itself. It was appropriate for studying problem solving strategies and other performance measures. In contrast, longitudinal studies incorporating multiple groups (first, second, third, fourth, and fifth year residents and staff radiologists with varying levels of experience) would allow researchers to investigate the acquisition of radiological expertise through repeated testing.

## Original Contributions to Knowledge: Theoretical and Instructional Implications

This section presents the theoretical and practical implications of this dissertation. Two follow-up projects from this dissertation are presently being planned. One is an empirical study that will incorporate physiological data to broaden our understanding of the diagnostic model. This study will be conducted to investigate the interaction between perceptual and cognitive factors in mammogram interpretation by converging multiple sources of data including eye-movements, verbal protocols, and physiological data (i.e., cortical activity). The second project is the design of the RadTutor, a computer-based prototype for training radiology residents to interpret mammograms. The RadTutor will incorporate the model of mammogram interpretation, problem solving strategies, and error analyses from the dissertation. It will also be based on other research in instructional psychology, cognitive science, artificial intelligence and education.

### Directions for Future Empirical Research

This dissertation contributes to the emerging cognitive literature on radiological expertise by providing a performance model focusing mainly on cognitive factors.

However, the visual aspects related to diagnostic problem solving were not "captured" and need to be studied and incorporated into the model. A systematic effort employing a combination of analytical and perceptual methodologies is required to clarify the contributions of cognitive and perceptual factors in the development of radiological expertise (e.g., verbal, eye-movement, and physiological measures).

Additional research on the interpretation of mammograms can add to the results of this dissertation. The convergence of verbal, eye-movement and physiological data requires the integration of two theoretical frameworks. The first is based on IPT of cognitive psychology. The basic assumption behind this framework is that the mind is a computational system that constructs, manipulates, and represents symbols. In general, participants with varying levels of expertise are asked to verbalize their "thinking processes" as they attempt to carry-out a specific task (e.g., diagnose a medical case). The verbal protocols are then segmented and participanted to knowledge representation analyses in order to formalize the underlying cognitive components of expertise, such as the types of knowledge used (declarative and procedural), the problem solving strategies (data-driven, goal-driven, mixed-strategy, etc.), the coherence of problem solving (local and global), the inferencing approach and the overall performance. The assumption underlying this approach is that the novice-expert progression in any domain is characterized by both quantitative and qualitative changes in knowledge representation and use. These differences can therefore be used to provide an understanding of cognitive development in the specific domain and thus also have numerous pedagogical implications.

Second, recent advances in brain imaging have facilitated the construction of comprehensive models of cognitive processes through the convergence of physiological and psychological research (for an extensive review refer to Gazzaniga, 1995). Non-invasive brain imaging techniques such as functional magnetic resonance imaging (fMRI) have been instrumental in (a) resolving debates in cognitive science (e.g., interactionism versus modularity) (Just, Carpenter & Keller, 1996), and (b) providing adequate models of cognitive processes including multi-level analyses (e.g., computational, algorithm, implementation) of data (e.g., physiological, process-tracing, performance) based on levels of organization and processing (Just, Carpenter, Keller, Eddy & Thulborn, 1996).

Future research has the potential to provide a comprehensive cognitive model of mammogram interpretation by isolating the various "levels" of perceptual and cognitive processes based on expert-novice differences. The proposed methodological approach would be similar to the one used in this study in terms of levels of expertise and dependent variables. Verbal, eye-movement and physiological data would be collected concurrently during the study. The experimental procedure would involve: (a) positioning and securing the participant's head in the scanner (fMRI), (b) presenting the patient's clinical history to the participant, (c) displaying the mammogram set, (d) instructing the participant to "think out loud" while he/she is diagnosing the case, and (e) instructing the participant to provide a formal diagnostic report. The participant's eye-movements and the activity of his/her cortical regions (e.g., pre-frontal cortex active during goal-driven problem solving) would also be captured using an eye-tracking system that is embedded in the functional magnetic resonance imaging (fMRI) unit which enables whole brain imaging.

Quantitative analyses would be conducted to determine the statistical significance between the levels of expertise on a number of dependent measures (e.g., scan time, reading time, etc.). The verbal data would be analyzed using the same coding scheme present in this study. Image analyses would be conducted on the eye-movement data and fMRI images collected from each participant. Off-line computer processing for the reconstruction and display of functional images would be performed by custom written computer programs presently used by cognitive scientists (Eddy, Fitzgerald, Genovese, Mockus & Noll, 1996). The objective of these analyses would be to (a) reveal the underlying cortical regions that are active during the various stages of diagnostic problem solving, and (b) calculate the duration of saccades, fixations, and regressions for each participant while he/she scans each set of mammograms.

In summary, the contributions of this study would include: (a) a comprehensive cognitive model of diagnostic reasoning in radiology based on the convergence of verbal, eye-movement and physiological data, (b) extended replication data which could be compared to the present data, (c) enhanced understanding of the perceptual and cognitive processes underlying mammography interpretation, (d) an initial theory of learning in ill-structured domains, and (e) rich experimental data necessary to run a cognitive simulation model of radiological expertise.

## The Cognitive Basis for the Design of the RadTutor[5]

This section outlines a conceptual framework for the development of the RadTutor. As previously mentioned, it incorporates the results of this dissertation including the model of mammogram interpretation, the problem solving strategies used by staff radiologists and radiology residents, and the typical case-related errors. Furthermore, the framework is based on: (a) a critical assessment of the haphazard nature of radiology residency training programs, (b) a review and critique of existing computer-based radiology training environments, (c) an analysis of authentic radiology resident teaching rounds, and (4) instructional principles for the design of the mammography tutor.

The purpose of this section is to briefly present a cognitively-based and empirically-derived approach for the design of the RadTutor, a prototype computerized tutor to train radiology residents in diagnosing mammograms exhibiting breast diseases. A plethora of computer-based radiology training environments has recently been developed with the objective of supporting the acquisition of radiological expertise. In general, these systems have failed to reach this objective since they: (a) lack a theoretical framework incorporating the empirical evidence on medical cognition and radiological expertise, (b) are not based on an adequate model of instruction, and (c) are based on comparison studies (between

---

[5] A more complete paper on the design of the RadTutor has been published as Azevedo et al. (1997). RadTutor: The theoretical and empirical basis for the design of a mammography interpretation tutor. In B. du Boulay & R. Mizoguchi (Eds.), *Frontiers in Artificial Intelligence and Application* (pp. 386-393). Amsterdam: IOS Press.

computerized instruction and traditional lectures or several CBI typologies) which lack both methodological and statistical rigor. In contrast, the design of the RadTutor is based on the results of this dissertation and attempts to use a more empirically-driven instructional model.

**Computer-Based Environments for Radiology Training.** The problem of inconsistency in medical (including radiology) residency training programs has recently been addressed by the widespread proliferation and dissemination of computer-based training programs. However, these systems typically suffer from some major instructional deficits (for an extensive review of these instructional issues refer to Azevedo et al., 1997). There has been a general increase in interest in the application of ITSs in the area of medical education in recent years. The rationale for computer-based instruction is based on the assumption that the learner's cognitive processes can be modeled, traced, and corrected in the context of problem-solving (Anderson, Corbett, Koedinger & Pelletier, 1995; Derry & Lajoie, 1993; Greer & McCalla, 1994). In recent years, several ITSs have been developed for radiology training.

Examples include the CT Brain Tutor for training radiology residents to diagnose brain tumors from CT scans (Macura, Macura, Toro, Binet, & Trueblood, 1994), and a tutor for training radiology residents to diagnose neurological MRI images (Sharples, duBoulay, Teather, Teather, Jeffrey & duBoulay, 1995). The extensive work of Sharples and colleagues (Sharples, 1988; 1991; 1997; Sharples & duBoulay, 1992; Sharples, duBoulay, Teather, Teather, Jeffrey & duBoulay, 1994) in developing the CT and MRI tutors focuses on accounts of professional practice and skill development and how these issues influence the design of their tutors. They have used statistically-based principles and a structured image description language for teaching radiological image interpretation and the diagnosis of cerebral diseases. Their approach to visual concept tutoring is based on grouping exemplars. Their tutoring approach facilitates the novice to expert transition by assisting the residents in the progression from visual to structural schemas (facilitating rapid pattern matching) and therefore ensuring transfer of skills and learning (Sharples, 1991). Lastly, their tutors aim at training radiologists to view and describe images in a systematic manner. In sum, these ITSs are based on cognitive science principles of expertise development and incorporate tutoring interventions and tutorial dialogues that are based on analyses of human interactions.

An ITS approach offers consistency and standardization in the training of mammogram interpretation. Therefore, this section presents a cognitively-based and empirically-derived approach for the design of the RadTutor, a computerized tutor for training radiology residents to diagnose breast diseases from mammograms.

**Instructional Principles Underlying the RadTutor.** This section briefly delineates the instructional principles adapted for teaching breast disease diagnosis. This adaptation involved integrating the results of the dissertation with the empirical research on medical cognition, radiological expertise, mammogram interpretation, and analyses of radiology rounds. The successful integration of these several sources of evidence was critical to the design of a theoretically-based and empirically-derived prototype version of the RadTutor. Each of the four instructional principles posited is supported by existing theoretical and empirical evidence described previously (Azevedo et al., 1997). These instructional principles are: (a) multiplicity, (b) activeness, (c) accommodation and adaptation, and (d) authenticity (for an extensive overview refer to Koschmann, Kelson, Feltovich & Barrows, 1996).

The *principle of multiplicity* is based on the concept that knowledge is complex, context-sensitive, and inter-related. It's implication for instruction is that multiple perspectives, representations and strategies should be promoted. This principle is based on the theory of cognitive flexibility (Spiro, Feltovich, Jacobson & Coulson, 1991) in medicine which emphasizes the use of multiple knowledge representations and repeated exposure to instructional content in diagnosis. The RadTutor will provide the resident with a stock of breast disease cases that can be accessed in a structured manner according to diagnostic categories, specific mammographic manifestations (i.e., findings and observations), and relevant clinical history cues.

The *principle of activeness* is based on the concept that learning is an active process, requiring mental construction and manipulation of subsymbolic (e.g., gray-scale densities exhibited on mammograms) and symbolic representations (e.g., clinical findings exhibited on mammograms and relevant clinical history cues). In the RadTutor, instruction will foster knowledge construction through meaningful problem-solving activities which facilitate skill acquisition (VanLehn, 1996) and the development of expertise (Ericsson, 1997; Ericsson & Lehmann, 1996).

The *principle of accommodation and adaptation* is based on the concept that the learning process is to a large degree affected by the extent of the learner's existing knowledge. In the RadTutor, instruction will facilitate adaptability by building upon the learner's existing knowledge, monitoring learner progress and rectifying errors when they arise, and fostering the development of metacognitive skills. A rule-based domain knowledge module and a student modelling approach are presently being considered based on the well-constrained nature of the domain of mammography.

The *principle of authenticity* is based on the concept that learning is sensitive to contextual factors which determine the usability of what is learned and the extent of skill transfer (Clancey, 1997; Greeno, 1994). In the RadTutor, instruction will provide learning activities that are required in the domain, that are valued in the real-world context, and that emulate the real-world environment as much as possible. In the RadTutor, the problem-solving activities will resemble what is routinely encountered in a resident's work environment. For example, it will provide the tools typically used to solve mammogram cases (such as a magnifying glass and a ruler to measure masses and lesions).

**Analyses of the Radiology Teaching Rounds: Eliciting Teaching Strategies.** The tutoring strategies and levels of instructional scaffolding adopted in the prototype are based on the authentic analyses of radiology teaching rounds (Azevedo, Lajoie, Desaulniers

& Fleiszer, 1996). These analyses examined the diagnostic problem solving processes and teaching methodologies employed by a staff radiologist teaching six residents during two one-hour mammography rounds. However, this section briefly presents the teaching methods used by the staff radiologist during the mammography teaching rounds.

The teaching methods used by the staff radiologist during the breast disease rounds including *coaching*, *scaffolding* and *fading*, and *articulation* (similar to the teaching methods advocated by the proponents of cognitive apprenticeship [e.g., Brown et al., 1989]) are presently being incorporated into the RadTutor. Occasionally, the radiologist would also provide *coaching* and support. In the case of intermediate residents, the radiologist would provide scaffolding during the diagnostic process in the form of hints, redirecting their viewing process, and subsequently fading all instructional support when the resident demonstrated the capability to pursue the task on his/her own. In the RadTutor, *coaching* and hints are provided through text messages, pop-up text messages, and highlighting of mammographic findings and observations. In addition, multiple levels of instructional *scaffolding* have also been delineated based on the results of the dissertation and interviews with the expert staff radiologist. The finest illustrations of *articulation* were observed when the radiologist externalized her reasoning process beginning with the assignment of probabilities to pathological features, followed by the systematic elimination of competing differential diagnoses until the definitive diagnosis was achieved. This teaching method was especially valuable since all residents may potentially benefit from the externalization of the expert's diagnostic problem solving. In the RadTutor, articulation is being implemented by using digitized video clips of the expert staff radiologist diagnosing a case while the tutoring system highlights the mammographic findings and observations on the digitized mammograms.

In summary, this section has presented a cognitively-based and empirically-derived approach for the design of a prototype computerized tutor to train radiology residents in diagnosing mammograms. The design approach is based on the results of the dissertation including the model of mammogram interpretation, the problem solving strategies used by staff radiologists and radiology residents, and typical case-related errors.

## Bridging the Gap between Cognitive Science and Learning in Medicine

Understanding how expertise is acquired poses a great challenge to learning theory (Glaser, 1996). The real challenge in relation to the acquisition of expertise is in terms of the learning phenomena involved and the conditions for optimal acquisition of competence. Most of the studies conducted in medical expertise (and other domains) have described the acquired properties of expert performance. Cognitive task analyses based on performance theory have significantly contributed to these descriptions and to an understanding of human cognition. Only indirect attempts have been made to infer what the properties of attained expertise might mean for the acquisition of competence. According to Glaser (1996), the recent book edited by Ericsson (1996) has begun to ask the question that follows cognitive task analysis: *What are the conditions for development, learning, and acquisition of expertise?* The study of learning is proceeding at various levels of analysis. At a microlevel there are studies involving the dynamics of memory and verbal learning, and neurophysiology of elementary learning processes. An intermediate level is characterized by studies of machine learning and AI. At the molar level, instructional experiments are being conducted on the design of educational and training environments for optimizing learning and improving performance (Brown, 1994; Glaser & Bassok, 1989; Frederisken & Donin, 1996). According to Glaser (1996), further research at this molar-instructional level should place an emphasis on the development of a learning theory for designing instructional environments both to understand improvements in performance and to further refine theories applicable to the design of conditions for learning.

Glaser's (1996) goal in terms of conducting research at the *molar level* has been addressed in some recent research studies in cognitive science. Charness, Krampe and Mayr's (1996) recent model of factors supporting expertise/skill acquisition is one example. The factors include: (a) external social factors (parental/spousal support, coaches, role models, cultural support, financial support, competing demands), (b) internal motivational/personality factors (introversion/extroversion, attention span, compulsivity, competitiveness), (c) external information factors (discipline organization and dissemination channels), (d) practice (intensity [deliberate vs. causal], duration and content), and (e) the cognitive system (software [knowledge base - chunk size, retrieval structures], problem solving processes [representation, search mechanism], and hardware [working memory capacity, speed of processing, learning rate, and forgetting rate]). These factors can be used to study medical expertise at different "levels" and to gain a more comprehensive understanding of the conditions of learning that support the acquisition of competence.

The research from novice-expert differences and cognitive task analyses based on IP theory have produced a voluminous amount of rich data that has the potential to contribute to and improve learning and instruction in medicine. However, the contributions of cognitive science are much more prominent in participant-matter domains such as math and computer programming. An example is Anderson, Corbett, Koedinger and Pelletier's (1995) research program involving the development of ACT-R as a theory of cognition and its implementation in numerous computerized tutors. This research program has recently led to the widespread dissemination of computerized tutors to several Pittsburgh schools. It is an excellent example of how cognitive science can contribute to the improvement of learning and instruction.

A large body of cognitive science literature on medical expertise exists. As is the case with this type of research in other domains, it has yet to provide direct implications for the improvement of learning and instruction in medicine (Lillehaug & Lajoie, in press). The difficulty is in deriving learning strategies, instructional methods, and assessment techniques based on the existing research findings. Secondly, whereas various cognitive research groups such as the CTGV (1990; 1992; 1993; 1996) have not only analyzed subject-matter domains such as mathematics, but they have an adopted an alternative perspective of cognition (i.e., constructivistic approach). They have established links with schools, and the research conducted has revealed impressive learning gains. Recently, medical researchers have incorporated components of situated action into the study of decision making in dynamic clinical environments (Lajoie, Azevedo & Fleiszer, in press; Patel et. al., 1996). These environments are characterized by ill-structured problems, uncertainty, shifting ill-defined or competing goals, action feedback loops, time stress and high risk, and multiple players. However, unlike the recent shift from the traditional IP to the use of situated cognition in other domains, the epistemological and analytical shifts in cognitive research in medicine is largely an academic exercise and not one that is directly aimed at impacting learning and instruction in medicine. For example, the results of cognitive studies examining the *effects of* problem-based learning (PBL) curricula on medical reasoning have not been used to modify the PBL curricula.

In summary, brief descriptions of cognitive science research in areas such as expert-novice differences and medical expertise have been provided along with some of their implications for teaching and learning in medicine. Previous studies have attempted to delineate the "psychological boundaries" of the human mind, to provide models of human competence in medicine, and to outline the differences between medical experts and novices. The results of this research have been instrumental in making specific instructional prescriptions and has served as the basis for the development of numerous computer-based medical learning environments.

## Conclusion

The present study investigated the problem solving strategies used by staff radiologists and radiology residents during the process of diagnosing difficult breast diseases depicted on mammograms. It was designed to investigate the diagnostic problem solving strategies used by staff radiologists and radiology residents during the interpretation of difficult breast disease cases depicted on mammograms. The results indictated that staff radiologists scanned the cases significantly faster than, radiology residents. No group differences were found in the number of radiological findings, radiological observations, and number of diagnoses across experimental conditions. Frequency analyses revealed that both groups regardless of experimental condition (a) used the same types of operators, control processes, diagnostic plans and goals, (b) committed the same number of errors, and (c) committed case-dependent errors. Analyses revealed that mammography interpretation was characterized by a predominant use of data-driven or mixed strategies depending on case typicality and clinical experience. The fact that few differences were found between the groups on the various measures may be due to the fact that mammogram interpretation is a well-constrained task. The theoretical implications of the study include the need for further studies for the purposes of building a more detailed, comprehensive model of the perceptual and cognitive processes underlying mammogram interpretation. Finally, the results have been applied to develop a basic conceptual framework for the development of the RadTutor, a computer-based tutor for training radiology residents to interpret mammograms. In sum, the results of this study contribute to the literarature on problem solving, human expertise, medical cognition, radiological expertise, and instructional psychology.

## REFERENCES


Abrams, H.L. (1981). Research in diagnostic radiology: A holistic perspective. *Clinical Radiology, 32*, 121-128.

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory & Cognition, 9*, 422-433.

Allen, S.W., & Brooks, L.R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120*, 3-19.

Anderson, J.R. (1983a). Skill acquisition: Compilation of weak method problem solution. *Psychological Review, 94*, 192-210.

Anderson, J.R. (1983b). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J.R. (1987). Methodologies for studying human knowledge. *Behavioural and Brain Sciences, 10*, 467-505.

Anderson, J.R. (1987). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J.R. (1990). *Cognitive psychology and its implications.* New York: W. H. Freeman.

Anderson, J.R. (1993a). Problem solving and learning. *American Psychologist, 48*(1), 35-44.

Anderson, J.R. (1993b). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J.R., Corbett, A.T., Koedinger, K.R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4* (2), 167-207.

Anderson, J.R., Farrell, R., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science, 8*, 87-129.

Anderson, J.R., Pirolli, P., & Farrell, R. (1988). Learning to program recursive functions. In M. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 153-184). Hillsdale, NJ: Erlbaum.

Anzai, Y. (1991). Learning and use of representations for physics research. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 64-92). Cambridge, MA: Cambridge University Press.

Arocha, J.F., & Patel, V.L. (1995). Construction-integration theory and clinical reasoning. In C. A. Weaver, S. Mannes & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 359-381). Hillsdale, NJ: Erlbaum.

Arocha, J.F., Patel, V.L., & Patel, Y.C. (1993). Hypothesis generation and the coordination of theory and evidence in novice diagnostic reasoning. *Journal of Medical Decision Making*, *13*, 198-211.

Atkinson, P., & Hammersley, M. (1994). Ethnography and participant observation. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 248-261). Thousand oaks, CA: Sage Publications.

Azevedo, R., Lajoie, S.P., Desaulniers, M., & Fleiszer, D.M. (August, 1996). *Tutoring complex visual concepts in radiology*. Paper presented at the 26th International Congress of Psychology, Montreal.

Azevedo, R., Lajoie, S.P., Desaulniers, M., Fleiszer, D.M., & Bret, P.M. (1997). RadTutor: The theoretical and empirical basis for the design of a mammography interpretation tutor. In B. du Boulay & R. Mizoguchi (Eds.), *Frontiers in Artificial Intelligence and Application* (pp. 386-393). Amsterdam: IOS Press.

Azevedo, R., Lajoie, S.P., & Fleiszer, D.M. (1996, August). *Complex clinical decision making in all ill-structured task.* Paper presented at the 26th International Congress of Psychology, Montreal, Canada.

Barosi, G., Magnani, L., & Stefanelli, M. (1993). Medical diagnostic reasoning: Epistemological modeling as a strategy for design of computer-based consultation programs. *Theoretical Medicine, 14*, 43-55.

Barrows, H.S., & Tamblyn, R. (1980). *Problem-based learning: An approach to medical education*. NY: Springer.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: The Syndics of the Cambridge University Press.

Berbaum, K. (1988). Impact of clinical history on fracture detection with radiography. *Radiology, 168*, 508-511.

Berbaum, K., Franken, E.A., & Smith, W.L. (1985). The effect of comparing films upon resident interpretation of pediatric chest radiographs. *Investigative Radiology, 20*(2), 124-128.

Berbaum, K., Franken, E.A., Anderson, K.L., Dorfman, D.D., Erkonen, W.E., Farrar, G.P., Geraghty, J.J., Gleason, T.J., MacNaughton, M.E., Phillips, M.E., Renfrew, D.L., Walker, C.W., Whitten, C.G., & Young, D.C. (1993). The influence of clinical history on visual search with single and multiple abnormalities. *Investigative Radiology, 28*(3), 191-201.

Berbaum, K., Franken, E. A., Dorfman, D.D., Rooholamini, S.A., Coffman, C.E., Cornell, S.H., Cragg, A.H., Galvin, J.R., Honda, H., Kao, S.C.S., Kimball, D.A., Ryals, T.J., Sickels, W.J., & Smith, T.P. (1991). Time course of satisfaction of search. *Investigative Radiology, 26*(7), 640-648.

Bogdan, R.C., & Biklen,,S.K. (1992). *Qualitative research for education: An introduction to theory and methods* (2nd ed.). Boston, MA: Allyn and Bacon.

Boreham, N. (1995). Error analysis and expert-novice differences in medical diagnosis. In J.M. Hoc, P.C. Cacciabue & E. Hollnagel (Eds.), *Expertise and technology* (pp. 93-105). Hillsdale, NJ: Erlbaum.

Borg, G., & Gall, M.D. (1996). *Educational research: An introduction* (6th ed.). White Plains, NY: Longman.

Boshuizen, H.P.A. & Schmidt, H.G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates, and novices. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 517-523). Montreal, Quebec, Canada.

Bracewell, R.J., & Breuleux, A. (1993). Substance and romance in analyzing think-aloud protocols. In P. Smogorinsky (Ed.), *Speaking about writing: Reflections on Research Methodology* (pp. 55-88). Newbury Park, CA: Sage.

Braccio, A. (1988). *On-line analysis of novice problem solving in medicine.* Unpublished master's thesis, McGill University, Montreal, Quebec, Canada.

Breuleux, A. (1991). The analysis of writers' think-aloud protocols: Developing a principled coding scheme for ill-structured tasks. In G. Denhiere & J.P. Rossi (Eds.), *Text and text processing* (pp. 333-362). North-Holland: Elsevier Science Publishers.

Brown, A.L. (1994). The advancement of learning. *Educational Researcher, 23*(8), 4-11.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Brown, J.S., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science, 4*(4), 379-426.

Camerer, C.F., & Johnson, E.J. (1991). The process-performance paradox in expert judgment: How can the experts know so much and predict so badly? In K.A. Ericsson & J. Smith (Eds.), *Towards a general theory of expertise: Prospects and limits* (pp. 195-217). Cambridge, MA: Cambridge University Press.

Cantor, N., Smith, E.E., French, R., & Mezzich, J. (1980). Psychiatric diagnoses as prototype categorization. *Journal of Abnormal Psychology, 89*, 181-193.

Carlson, R.A., Khoo, B.H., Yaure, R.G., & Schneider, W. (1990). Acquisition of a problem-solving skill: Levels of organization and use of working memory. *Journal of Experimental Psychology: General, 119*, 193-214.

Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology, 33*, 1-16.

Charness, N. (1989). Expertise in chess and bridge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 183-208). Hillsdale, NJ: Erlbaum.

Charness, N. (1991). Expertise in chess: The balance between knowledge and search. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 39-63). Cambridge, MA: Cambridge University Press.

Charness, N., Krampe, R., & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: An international comparison of life-span chess skill acquisition. In K.A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 51-80). Mahwah, NJ: Erlbaum.

Chase, W.G., & Ericsson, K.A. (1981). Skilled memory. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141-189). Hillsdale, NJ: Erlbaum.

Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Chi, M.T.H., Glaser, R., & Farr, M.J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Chi, M.T.H., de Leeuw, N., Chiu, M-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Christensen, E.E., Murray, R.C., & Holland, K. (1981). The effect of search time on perception. *Radiology, 138*, 361-365.

Claessen, H.F., & Boshuizen, H.P.A. (1985). Recall of medical information by students and doctors. *Medical Education, 19*, 61-67.

Clancey, W.J. (1997). The conceptual nature of knowledge, situations and activity. In P.J. Feltovich, K.M. Ford & R.R. Hoffman (Eds.), *Expertise in context* (pp. 247-291). Mahwah, NJ: Erlbaum.

Clement, J. (1991). Nonformal reasoning in experts and in science students. In J.F. Voss, D.N. Perkins & J.W. Segal (Eds.), *Informal reasoning and education* (pp. 345-362). Hillsdale, NJ: Erlbaum.

Cooper, R.A. (1992). Mammography. In J.H. Isaacs (Ed.), *Textbook of breast disease* (pp. 47-86). St. Louis, MO: Mosby.

Corbett, A.T., Anderson, J.R., & Patterson, E.G. (1990). Student modeling and tutoring flexibility in the Lisp intelligent tutoring system. In C. Frasson & G. Gauthier (Eds.), *Intelligent tutoring systems: At the crossroads of artificial intelligence and education* (pp. 83-106). Norwood, NJ: Ablex Publishing Company.

CTGV. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19* (5), 2-10.

CTGV. (1992). The Jasper series as an example of anchored instruction: Theory, program description, and assessment data. *Educational Psychologist, 27* (3), 291-315.

CTGV. (1993). Toward integrated curricula: Possibilities from anchored instruction. In M. Rabinowitz (Ed.), *Cognitive science foundations of instruction* (pp. 33-55). Hillsdale, NJ: Erlbaum.

CTGV. (1996). Multimedia environments for enhancing learning in mathematics. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), *International perspectives on the design of technology-supported learning environments* (pp. 285-306). Mahwah, NJ: Erlbaum.

Custers, E. J. (1995). *The development and function of illness scripts.* Netherlands: Universitaire Par Mastriche.

Custers, E.J., Boshuizen, H.P.A., & Schmidt, H.G. (1996). The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Memory & Cognition*, *24*(3), 384-399.

Cutler, P. (1979). *Problem solving in clinical medicine: From data to diagnosis.* Baltimore, MA: Williams & Wilkins.

de Groot, A.D. (1965). *Thought and choice in chess.* The Hague: Mouton.

de Groot, A.D. (1946/1978). *Thought and choice in chess.* Den Haag, The Netherlands: Mouton.

Derry, S.J., & Lajoie, S.P. (1993). A middle camp for (un)intelligent instructional computing: An introduction. In S.P. Lajoie & S.J. Derry (Eds.), *Computers as cognitive tools* (pp. 1-11). Hillsdale, NJ: Erlbaum.

Donegan, W., & Spratt, J.S. (1995). *Cancer of the breast.* Philadelphia, PA: W.B. Saunders.

Dreyfus, H.L., & Dreyfus, S.E. (1986). Why skills cannot be represented by rules. In N.E. Starkey (Ed.), *Advances in cognitive science. Volume 1.* Chichester, GB: Ellis Horwood, Ltd.

Dubilet, P. & Hermann, P. G. (1981). Integration of radiographs: Effect of clinical history. *American Journal of Radiography, 137*, 1055-1058.

Duncker, K. (1945). *On problem-solving.* Westport: Greenwood Press.

Eddy, W.F., Fitzgerald, M., Genovese, C.R., Mockus, A., & Noll, D.C. (1996). Functional image analysis software-computational olio. In A. Prat (Ed.), *Proceedings in computational statistics.* Heidelberg, Germany: Physica-Verlag.

Ehrlich, K., & Soloway, E. (1984). An empirical investigation of the tacit plan knowledge in programming. In J.C. Thomas & M.L. Schneider (Eds.), *Human factors in computer systems* (pp. 113-133). Norwood, NJ: Ablex.

Elio, R., & Scharf, P.B. (1990). Modeling novice-to-expert shifts in problem solving strategy and knowledge organization. *Cognitive Science, 14*, 579-639.

Elstein, A.S., Shulman, L.S., & Sprafka, S.A. (1978). *Medical problem solving. An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Engle, R.W., & Buckstel, L.H. (1978). Memory processes among bridge players of differing expertise. *American Jounal of Psychology, 91*(4), 673-689.

Ericsson, K.A. (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Mahwah, NJ: Erlbaum.

Ericsson, K.A., & Charness, N. (1997). Cognitive and developmental factors in expert performance. In P.J. Feltovich, K.M. Ford & R.R. Hoffman (Eds.), *Expertise in context* (pp. 1-41). Mahwah, NJ: Erlbaum.

Ericsson, K.A., Krampe, R.T., & Tesch-Romer, C. (1993). The role of deliberate practice in acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Ericsson, K.A., Lehmann, A.C. (1993). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305.

Ericsson, K.A., & Oliver, W.L. (1988). Methodology for laboratory research on thinking: Task selection, collection of observations, and data analysis. In R.J. Strenberg & E.E. Smith (Eds.), *The psychology of human thought* (pp. 392-428). NY: Cambridge University Press.

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data (2nd ed.)*. Cambridge, MA: MIT Press.

Ericsson, K.A., & Smith, J. (1991a). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 286-300). Cambridge, MA: Cambridge University Press.

Ericsson, K.A., & Smith, J. (1991b). *Toward a general theory of expertise*. Cambridge, MA: Cambridge University Press.

Ericsson, K.A., & Staszewski, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kovosky (Eds.), *Compex information processing: The impact of Herbert A. Simon* (pp. 235-267). Hillsdale, NJ: Erlbaum.

Evans, D., & Gadd, C.S. (1989). Managing coherence and context in medical problem-solving discourse. In D.A. Evans & V.L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 211-255). Cambridge, MA: MIT Press.

Evans, D.A., & Patel, V.L. (1989). *Cognitive science in medicine: Biomedical modeling*. Cambridge, MA: The MIT Press.

Faremo, S. (1997). *Novice diagnostic reasoning in a visual medical domain: Implications for the design of a computer-based instructional system for undergraduate medical education.* Unpublished master's thesis, Concordia University, Montreal, Quebec, Canada.

Feigenbaum, E.A., & Simon, H.A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8*, 305-336.

Feltovich, P.J., & Barrows, H.S. (1984). Issues of generality in medical problem solving. In H.G. Schmidt, & M.L. De Volder (Eds.), *Tutorials in problem based learning: New directions in training for the health professions* (pp. 128-142). Assen, The Netherlands: Van Gorcum.

Feltovich, P.J., Ford, K.M., & Hoffman, R.R. (1997). *Expertise in context.* Menlo Park, CA: MIT Press.

Feltovich, P.J., Johnson, P.E., Moller, J.H., & Swanson, D.B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W.J. Clancey and E.H. Shortliffe (Eds.), *Readings in medical artificial intelligence* (pp. 275-319). Reading, MA: Addison-Wesley.

Feltovich, P.J., Spiro, R.J., & Coulson, R.L. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D.A. Evans, & V.L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling,* (pp. 111-172). Cambridge, MA: MIT Press.

Feltovich, P.J., Spiro, R.J., & Coulson, R.L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. In P.J. Feltovich, K.M. Ford & R.R. Hoffman (Eds.), *Expertise in context* (pp. 125-146). Mahwah, NJ: Erlbaum.

Fisher, R.A., & Yates, F. (1963). *Statistical tables for biological, agricultural and medical research* (6th ed.). London, UK: Oliver and Boyd.

Forsythe, G.B., & Barber, H.F. (1992, April). *Military strategic thinking: Expert-novice differences in the structure and content of cognitive representations.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Frederiksen, C.H. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7*, 371-458.

Frederiksen, C.H. (1986). Cognitive models and discourse analysis. In C.R. Cooper & S. Greenbaum (Eds.), *Wrritten communication annual, Vol. 1: Studying writing: Linguistic approaches* (pp. 227-267). Beverly Hills, CA: Sage Publications.

Frederiksen, C.H., & Donin, J. (1996, August). *Conceptual processing in text comprehension and learning through discourse.* Paper presented at the 26th International Congress of Psychology, Montreal, Canada.

Friedman, C., Cimino, J.J., & Johnson, S.B. (1994). A schema for representing medical anguage applied to clinical radiology. *Journal of the American Medical Informatics Association, 1(3)*, 233-248.

Friedman, C. Huff, S.M., Hersh, W.R., Pattison-Gordon, E., & Cmino, J.J. (1995). The Canon Group's effort: Working towards a merged model. *Journal of the American Medical Informatics Association, 2*(1), 4-18.

Gale, J., & Marsden, P. (1983). *Medical diagnosis.* Oxford: Oxford University Press.

Gaudette, L.A., Silberberger, C., Altmayer, C.A., & Gao, R.N. (1996). Trends in breast cancer incidence and mortality. *Health Reports, 8*(2), 29-37.

Gazzaniga, M. (1995). *The cognitive neurosciences.* Cambridge, MA: MIT Press.

Gentner, D.R. (1988). Expertise in typewriting. In M. Chi, R. Glaser & M.J. Farr (Eds.), *The nature of expertise* (pp. 1-22). Hillsdale, NJ: Erlbaum.

Gick, M.L., & Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Gilhooly, K.J. (1990). Cognitive psychology and medical diagnosis. *Applied Cognitive Psychology, 4*, 261-272.

Gilhooly, K.J., & Green, A.J.K. (1988). The use of memory by experts and novices. In A.M. Colley & J.R. Beech (Eds.), *Cognition and action in skilled behavior* (pp. 379-395). Amsterdam: Elsevier Science Publishers.

Glaser, R. (1986). On the nature of expertise. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 915-928). Amsterdam: Elsevier Science Publishers.

Glaser, R. (1996). Changing the agency for learning: Acquiring expert performance. In K.A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 303-311). Mahwah, NJ: Erlbaum.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology, 40*, 631-666.

Glaser, R., & Chi, M.T.H. (1988). Overview. In M.T.H. Chi, R. Glaser & M.J. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.

Goergen, S.K., Evans, J., Cohen, G.P.B., & MacMillan, J.H. (1997). Characteristics of breast carcinomas missed by screening radiologists. *Radiology, 204*, 131-135.

Greeno, J.G. (1994). Gibson's affordances. *Psychological Review, 101*(2), 336-342.

Greeno, J.G., & Simon, H.A. (1988). Problem solving and reasoning. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey & R.D. Lee (Eds.), *Steven's handbook of experimental psychology* (pp. 589-672). NY: John-Wiley.

Greer, J.E., & McCalla, G.I. (1994). *Student modelling: The key to individualized knowledge-based instruction.* NATO ASI Series (Vol. 125). Berlin: Springer-Verlag.

Groen, G.J., & Patel, V.L. (1985). Medical problem-solving: Some questionable assumptions. *Medical Education, 19*, 95-100.

Hassebrock, F., & Prietula, M.J. (1992). A protocol-based coding scheme for the analysis of medical reasoning. *International Journal of Man-Machine Studies, 37*, 613-52.

Hatano, G. (1996). Cognitive development and the acquisition of expertise. In E. de Corte & F.E. Weinert (Eds.), *International encyclopedia of developmental and instructional psychology* (pp. 273-276). Oxford: Pergamon.

Hatano, G., & Inagaki, K. (1996). Two courses of expertise. In H.W. Stevenson, H. Azuma & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262-272). NY: W.H. Freeman.

Hillard, A., Myles-Worsley, M., Johnston, W., & Baxter, (1985). The development of radiological schemata through training and experience: A preliminary communication. *Investigative Radiology, 18*(4), 422-425.

Holyoak, K.J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 286-300). Cambridge, MA: Cambridge University Press.

Holyoak, K.J., & Spellman, B.A. (1993). Thinking. *Annual Review of Psychology, 44*, 265-315.

Holyoak, K.J., & Thagard, P.R. (1989). Analogical reasoning as a mechanism in knowledge acquisition: A development perspective. In S. Vosniadu & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 242-266). Cambridge, UK: Cambridge University Press.

Homer, M.J. (1991). *Mammographic interpretation: A practical approach.* NY: McGraw-Hill.

Jackson, V. (1995). Breast disease. *Journal of Diagnostic Radiology, 6*(4), 123-156.

Johnson, P.E., Duran, A.S., Hassebrock, F., Moller, J., & Prietula, M. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science, 5*, 235-283.

Joseph, G.M., & Patel, V.L. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Journal of Medical Decision Making, 10*, 31-46.

Just, M.A., Carpenter, P.A., & Keller, T.A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103*, 773-780.

Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., & Thulborn, K.R. (1996). Brain activiation modulated by sentence comprehension. *Science, 274*, 114-116.

Kassirer, J.P., & Gorry, G.A. (1978). Clinical problem solving: A behavioral analysis. *Annals of Internal Medicine, 89*, 245-255.

Kaufman, D.R., & Patel, V.L. (1988). The nature of expertise in the clinical interview: Interactive problem solving. In *Proceedings of the tenth annual conference of the Cognitive Science Society* (pp. 461-467). Hillside, NJ: Erlbaum.

Kaufman, D.R., Patel, V.L., & Magder, S.A. (1992) Progressions of conceptual models of cardiovascular physiology and their relationship to expertise. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 885-890). Hillsdale, NJ: Erlbaum.

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163-182.

Kintsch, W. (1992). A cognitive architecture for comprehension. In H.L. Pick, P. Van den Broek & D.C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 143-163). Washington, DC: American Psychological Association.

Kintsch, W. & Van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1-55.

Koedinger, K.R., & Anderson, J.R. (1990). Abstract planning and conceptual chunks: Elements of expertise in geometry. *Cognitive Science, 14*, 511-540.

Kohler, W. (1925). *The mentality of apes*. NY: Harcourt, Brace, and World.

Koschmann, T.D., Kelson, A.C., Feltovich, P.J., & Barrows, H.S. (1996). Computer-supported problem-based learning: A principled approach to the use of computers in collaborative learning. In T.D. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 83-124). Hillsdale, NJ: Lawrence Erlbaum.

Kotovsky, K., & Simon, H.A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology, 4*, 399-424.

Krupinski, E.A., Nodine, C.F., & Kundel, H.L. (1993). Perceptual enhancement of tumor targets in chest x-ray images. *Perception & Psychophysics, 53*(5), 519-523.

Kuipers, B. & Kassirer, J. P. (1984). Causal reasoningin medicine: Analysis of a protocol. *Cognitive Science*, *8*, 363-385.

Kundel, H.A. (1985). Comment. *Investigative Radiology, 20*, 110-111.

Kundel, H.A., & Nodine, C.F. (1975). interpreting chest x-rays without visual search. *Radiology, 116*, 527-532.

Kundel, H.A., Nodine, C.F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology, 13*, 175-181.

Kushniruk, A.W., Patel, V.L., & Fleiszer, D.M. (1995). Complex decision making in providing surgical intensive care. In J.D. Moore & J.F. Lehman (Eds.), *Proceedings of the seventeenth annual conference of the Cognitive Science Society* (pp. 287-292). Hillsdale, NJ: Erlbaum.

Laird, J.E., & Newell, A. (1993). In P.S. Rosenbloom, J.E., Laird & A. Newell (Eds.) *The SOAR papers: Research on integrated intelligence* (pp.245-292). Cambridge, MA: The MIT Press.

Laird, J.E., Newell, A., & Rosenbloom, P.S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence, 33*(1), 1-64.

Lajoie, S.P., Azevedo, R., & Flesizer, D.M. (in press). Cognitive tools for assessment and learning in a high information flow environment. *The Journal of Educational Computing Research.*

Lajoie, S.P., & Derry, S.J. (1993). *Computers as cognitive tools*. Hillsdale, NJ: Erlbaum.

Lajoie, S.P., & Lesgold, A. (1989). Apprenticeship training in the workplace: Computer coached practice environment as a new form of apprenticeship. *Machine-Mediated Learning, 3*, 7-28.

Lancaster, J., & Kolodner, J. (1988). Varieties of learning from problem solving experience. *Proceedings of the tenth annual conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Larkin, J.H. (1983). The role of problem representation in physics. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 75-100). Hillsdale, NJ: Erlbaum.

Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335-1342.

Lawrence, J.A. (1988). Expertise on the bench: Modeling magistrates' judicial decision-making. In M. Chi, R. Glaser & M.J. Farr (Eds.), *The nature of expertise* (pp. 229-260) . Hillsdale, NJ: Erlbaum.

Lesgold, A. (1983). *Acquiring expertise* (Tech. Rep. No. PDS-5). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Lesgold, A. (1984). Human skill in a computerized society: Complex skills and their acquisition. *Behaviour, Research Methods, Instruments, & Computers, 16*(2), 79-87.

Lesgold, A. (1988). Problem solving. In R.J. Sternberg & E.E. Smith (Eds.), *The psychology of human thought* (pp. 188-214). Cambridge, MA: Cambridge University Press.

Lesgold, A., Feltovich, P.J., Glaser, R., & Wang, Y. (1981). *The acquisition of perceptual diagnostic skill in radiology* (Tech. Rep. No. PDS-1). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Lesgold, A., & Lajoie, S.P. (1991). Complex problem solving in electronics. In R.J. Sternberg & P.A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 287-316). Hillsdale, NJ: Erlbaum.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 311-342) . Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Lillehaug, S.I., & Lajoie, S.P. (in press). AI in medical education: Another grand challenge for medical informatics. *Journal of Artificial Intelligence in Medicine.*

Lo, J.Y., Baker, J.A., Kornguth, P.J., Iglehart, J.D., & Floyd, C.E. (1997). Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features. *Radiology, 203*, 159-163.

Lusted, L.B. (1984). ROC recollected. *Medical Decision Making, 4*, 131-135.

Macura, R.T., Macura, K.J., Toro, V.E., Binet, E.F. , Trueblood, J.H., & Ji, K. (1994). Computerized case-based instructional system for computed tomography and magnetic resonance imaging of brain tumors. *Investigative Radiology, 29*(4), 497-506.

Mayer, R. (1989). Non-adversary human problem solving. In K.J. Kilhooly (Ed.), *Human and machine problem solving*, (pp. 191-224). NY: Plenum Press.

Mayer, R. (1990). Problem solving. In M.W. Eysenck (Ed.), *The Blackwell dictioanry of cognitive psychology* (pp. 284-288). Oxford: Basil Blackwell.

Mayer, R. (1992). Cognition and instruction: Their historic meeting with educational psychology. *Journal of Educational Psychology*, *4,* 405-412.

Mayer, R. (1994). Problem solving. In R.J. Sternberg (Ed.*), Encyclopedia of human intelligence* (pp. 841-845). NY: Macmillan.

Murphy, G.L., & Wright, J.C. (1984). Changes in conceptual structures with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 144-155.

Neufeld, V.R., Norman, G.R., Feightner, J.W., & Barrows, H.S. (1981). Clinical problem solving by medical students: A longitudinal and cross-sectional analysis. *Medical Education*, *15*, 315-327.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 135-183.

Newell, A. (1989). Putting it all together. In D. Klahr & K. Kovosky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 399-440). Hillsdale, NJ: Erlbaum.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Newell, A., & Simon, H.A. (1976). Computer science as empirical inmquiry: Symblobs and search. *Communications of the ACM, 19*, 111-126.

Norman, G.R. (1988). Problem-solving skills, solving problems and problem-based learning. *Medical Education, 22*, 279-286.

Norman, G.R., Brooks, L.R., & Allen, S.W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, Cognition, 15*, 1166-1174.

Norman, G.R., Coblentz, C.L., Brooks, L.R., & Babcook, C.J. (1992). Expertise in visual diagnosis: A review of the literature. *Academic Medicine, 67*(10), S78-S83.

Norman, G.R., Jacoby, L.L., Feightner, J.W., & Campbell, G.J.M. (1979). Clinical experience and the structure of memory. *Proceedings of the 18th Annual Conference on Research in Medical Education,* (pp. 214-218). Washington, DC: American Association of Medical Colleges.

Norman, G.R., Rosenthal, D., Brooks, L. R., & Muzzin, L.J. (1989). The development of expertise in dermatology. *Archives of Dermatology, 125*, 1063-1068.

Novick, L.R. (1988). Analogical transfer, problem similarity and expertise. *Journal of Experimental Psychology: Learning, Memory, Cognition, 14*, 510-520.

Nurcombe, B., & Fitzhenry-Coor, I. (1982). How do psychiatrists think? Clinical reasoning in the psychiatric interview: A research and education project. *Australian and New Zealand Journal of Psychiatry, 16*, 13-24.

Olson, J.R., & Biolsi, K. J. (1991). Techniques for representing expert knowledge. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 240-285). Cambridge, MA: Cambridge University Press.

Patel, V.L., & Arocha, J.F. (1988). Cognitive diagnosis of clinical reasoning. In I. Hart, R. Harden & H. Walton (Eds.), *Newer developments in assessing clinical knowledge* . Montreal: Heal Publications.

Patel, V.L., & Arocha, J.F. (1995). Novice diagnostic reasoning in medicine: Accounting for evidence. *The Journal of the Learning Sciences, 4*(4), 355-384.

Patel, V.L., Arocha, J.F., & Groen, G.J. (1987). *Changes in expert reasoning as a function of task difficulty*. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Patel, V.L., Arocha, J.F., & Kaufman, D.R. (1994). Diagnostic reasoning and medical expertise. In D. Medin (Ed.), *The psychology of learning and motivation (Vol. 31)*(pp. 187-252). San Diego, CA: Academic Press.

Patel, V.L., & Coughlin, L.D. (1985). Cognitive science and the assessment of clinical competence. In *Proceedings of the International Conference on Newer Developments in Assessing Clinical Competence*, Ottawa, Canada.

Patel, V.L., Evans, D.A., & Groen, G.J. (1989). Biomedical knowledge and clinical reasoning. In D.A. Evans & V.L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 53-112). Cambridge, MA: MIT Press.

Patel, V.L., Evans, D.A., & Kaufman, D. (1989). Cognitive framework for doctor-patient interaction. In D. A. Evans & V. L. Patel (Eds.), *Cognitive Science in Medicine: Biomedical Modeling,* (pp. 253-308). Cambridge, MA: MIT Press.

Patel, V.L., Groen, G.J., & Arocha, J.F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition*, *18*(4), 394-406.

Patel, V.L., & Frederiksen, C.H. (1984). Cognitive processes in comprehension and knowledge acquisition by medical students and physicians. In H. G. Schmidt & M. L. de Volder (Eds.) *Tutorials in problem-based learning: New directions in training for the health professions*. Mastricht, The Netherlands: Van Gorcum.

Patel, V.L., & Groen, G.J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science, 10*, 91-116.

Patel, V.L., & Groen, G.J. (1991). The general and specific nature of medical expertise: A critical look. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93-125). New York: Cambridge University Press.

Patel, V.L., Groen, G.J., & Norman, G.R. (1993). Reasoning and instruction in medical curricula. *Cognition and Instruction, 10*(4), 335-378.

Patel, V.L., Groen, G.J., Ramoni, M.F., & Kaufman, D.R. (1992). Machine depth versus psychological depth: A lack of equivalence. In E. Keravnou (Ed.), *Deep models for medical knowledge engineering,* (pp. 249-272). Netherlands: Elsevier.

Patel, V.L., Groen, G.J., & Scott, H.M. (1988). Biomedical knowledge in explanations of clinical problems by medical students. *Medical Education, 22*, 398-406.

Patel, V.L., Kaufman, D.R., & Magder, S. (1991). Causal reasoning about complex physiological concepts by medical students. *International Journal of Science Education, 13*(2), 171-185.

Patel, V.L., Kaufman, D.R., & Magder, S. (1996). The acquisition of medical expertise in complex dynamic environments. In K.A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 127-165). Mahwah, NJ: Erlbaum.

Patel, V.L., & Ramoni, M.F. (1997). Cognitive models of directional inference in expert medical reasoning. In P.J. Feltovich, K.M. Ford & R.R. Hoffman (Eds.), *Expertise in context: Human and machine* (pp.67-99). Menlo Park, CA: AAAI Press.

Peters, M.E., Voegeli, D.R., & Scanlan, K.A. (1989). *Handbook of breast imaging*. NY: Churchill Livingstone.

Posner, M. (1989). *Foundations of cognitive science*. Cambridge, MA: MIT Press.

Powell, D.E., & Stelling, C.B. (1994). *The diagnosis and detection of breast disease*. St. Louis, MI: Mosby.

Reimann, P., & Chi, M.T.H. (1989). Human expertise. In K.J. Gilhooly (Ed.), *Human and machine problem solving* (pp.161-191). New York: Plenum Press.

Resnick, L.B., Levine, J.M., & Teasley, S.D. (1991). *Perspectives on socially shared cognition*. Washington, DC: American Psychological Association.

Riddererikhoff, J. (1991). Medical problem solving: An exploration of strategies. *Medical Education, 25*, 196-207.

Rogers, E. (1992). *Visual interaction: A link between perception and problem solving*. Unpublished dissertation, Georgia Institute of Technology, Atlanta, GA.

Rogers, E. (1995a). A cognitive theory of visual interaction. In B. Chandrasekaran, J. Glasgow & N. H. Narayanan (Eds.), *Diagrammatic reasoning: Cognitive and computational perspectives* (pp. 481-500). Menlo Park, CA: AAAI Press.

Roger, E. (1995b). VIA-RAD: A blackboard-based system for diagnostic radiology. *Artificial Intelligence in Medicine, 7*, 343-360.

Rogers, E. (1996). A study of visual reasoning in medical diagnosis. In G.W. Cottrell (Ed.). *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 213-218). Mahwah, NJ: Erlbaum.

Ramoni, M., Stefanelli, M., Magnani, L., & Barosi, G. (1992). An epistemological framework for medical knowledge-based systems. *IEEE Transacions on Systems, Man, and Cybernetics, 22*(6), 1361-1374.

Rubin, A. (1989). Contributions of cognitve science and educational technology to training in radiology. *Investigative Radiology, 24*(9), 729-732.

Rumelhart, D.E., & Norman, D.A. (1978). Accretion, tuning, and restructuring: Three modes of learning. In R. Klatzky & J.W. Cotton (Eds.), *Semantic factors in cognition* (pp. 37-60). Hillsdale, NJ: Lawrence Erlbaum Associates.

Salthouse, T.A. (1991). Expertise as the circumvention of human processing limitations. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 286-300). Cambridge, MA: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 172-194). Cambridge, MA: Cambridge University Press.

Schmidt, H.G., & Boshuizen, H.P.A. (1993). On acquiring expertise in medicine. *Educational Psychology Review, 5*, 205-221.

Schmidt, H.G., & Boshuizen, H.P.A., & Hobus, P.P.M. (1988). Transitory stages in the development of medical expertise: The "intermediate effect" in the clinical representation studies. *Proceeding of the tenth annual conference of the Cognitive Science Society* (pp. 139-145). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Schmidt, H.G., Norman, G.R., & Boshuizen, H.P.A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65*(10), 611-621.

Schyns, P.G. (1991). A modular neural network model of concept acquisition. *Cognitive Science, 15*, 461-508.

Sharples, M. (1988). The radiology tutor: Computer-based teaching of visual categorization. In D. Bierman, J. Breuker & J. Sandberg (Eds.), *Artificial intelligence and education* (pp. 252-259). Amsterdam: IOS.

Sharples, M. (1991). Computer-based tutoring of visual concepts: From novice to expert. *Journal of Computer-Based Learning, 7*, 123-132.

Sharples, M., & du Boulay, B. (1992) Heart and Head: The radiology tutor and beyond. *Rivista di Neuroradiologia, 5*, 465-471.

Sharples, M., du Boulay, B., Teather, B.A., Teather, D., Jeffrey, N., & du Boulay, G.H. (1995). The MR tutor: Computer-based training and professional practice. In J. Greer (Ed.), *Proceedings of AI-ED'95* (pp. 429-436). Charlottesville, VA: ACCE.

Sharples, M., du Boulay, B., Teather, D., Teather, B.A., Jeffrey, N., & du Boulay, G.H. (1994). The cognitive basis for an MRI tutor. In *Proceedings of the East-West Conference on Computer Technologies in Education (part 1)* (pp. 214-219). Crimea: Ukraine.

Sharples, M., Jeffrey, N., Teather, D., Teather, B., & Du Boulay, G. (1997). A socio-cognitive engineering approach to the development of a knowledge-based training system for neuroradiology. In B. du Boulay & R. Mizoguchi (Eds.), *Frontiers in Artificial Intelligence and Application* (pp. 402-410). Amsterdam: IOS Press.

Simon, D.P., & Simon, H.A. (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 325-348). Hillsdale, NJ: Erlbaum.

Simon, H.A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*, 268-288.

Simon, H.A. (1979). *Models of thought*. New Haven, CT: Yale University Press.

Simon, H.A. (1989). The scientist as a problem solver. In D. Klahr & K. Kovosky (Eds.*), Compex information processing: The impact of Herbert A. Simon* (pp. 375-398). Hillsdale, NJ: Erlbaum.

Simon, H.A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1-19.

Simon, H., & Kaplan, C. (1989). Foundations of cognitive science. In M. Posner (Ed.), *Foundations of cognitive science* (pp. 1-48). Cambridge, MA: MIT Press.

Sloboda, J. (1991). Musical expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 153-171). Cambridge, MA: Cambridge University Press.

Smagorinsky, P. (1993). Think-aloud protocol analysis: Beyond the black box. In P. Smagorinsky (Ed.), *Speaking about writing: Reflection on research methodology* (pp. 3-19). Newbury Park, CA: Sage.

Smith, E.E. (1995). Concepts and categorization. In E.E. Smith & D.N. Osherson (Eds.), *Thinking* (pp. 3-34). Cambridge, MA: MIT Press.

Smith, M.V. (1991*). Toward a unified theory of problem solving: Views form the content domains*. Hillsdale, NJ: Erlbaum.

Soloway, E., Adelson, B., & Ehrlich, K. (1988). Knowledge and proccesses in the comprehension of computer programs. In M. Chi, R. Glaser & M. J. Farr (Eds.), *The* *nature of expertise* (pp. 129-152). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Spiro, R.J., Feltovich, P.J., Jacobson, M.J., & Coulson, R.L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology, 31*(5), 24-33.

Staszewski, J.J. (1988). Skilled memory and expert mental calculation. In M. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 71-128). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load. *Cognition & Instruction, 8*(4), 351-362.

Swenson, R. G. (1980). A two-stage detection model applied to skilled search by radiologists. *Perceptual Psychology, 27*, 11-16.

Swenson, R. G., Hessel, S. J., & Herman, P. G. (1982). Radiological interpretation with and without visual search: Visual search aids the recognition of chest pathology. *Investigative Radiology, 17*, 145-151.

Swets, J.A. (1985). On searching films and begging the question. *Investigative Radiology, 20*, 115-116.

Swets, J.A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Mahwah, NJ: Erlbaum.

Tortorici, M.R., & Apfel, P.J. (1995*). Advanced radiographic and angiographic procedures*. Philadelphia, PA: F.A. Davis Company.

Tubbs, R.M. (1992). The effect of experience on the auditor's organization and amount of knowledge. *The Accounting Review, 67*, 783-801.

Turner, R. M. (1988). Organizing and using schematic knowledge for medical diagnosis. In J. L. Kolodner (Ed.), *Case-based Reasoning: Proceedings of a Workshop on Case-based Reasoning* (pp. 435-446). Clearwater Beach, Florida.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232.

van Dijk, T.A., & Kintsch, W. (1983) *Strategies of discourse comprehension.* New York, NY: Academic Press

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. Posner (Ed.), *Foundations of cognitive science* (pp. 527-579). Cambridge, MA: MIT Press.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*, 513-539.

Vaastra, H.F., Boshuizen, H.P.A., & Schmidt, H.G. (1993, April). *Expertise in auditing: Case representation differences between economy students, novice-auditors, and experts.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Waldrop, M.M. (1984). The necessity of knowledge. *Science, 223*, 1279-1282.

Weinstein, M.C., & Fineberg, H.V. (1980). *Clinical decision analysis.* Philadelphia, PA: W.B. Saunders.

Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*, 73-87.

Winograd, T. (1983). *Language as a cognitive process*. Reading, MA: Addison-Wesley.

Wortman, P.M. (1966). Representation and strategy in diagnostic problem solving. *IEEE Transactions on Human Factors in Electronics, 8*, 48-53.

Wu, Y., Giger, M.L., Doi, K., Vyborny, C.J. Schmidt, R.A., & Metz, C.E. (1993). Artificial neural netwroks in mammography: Application to decision making in the diagnosis of breast disease. *Radiology, 187*, 81-87.

Zeitz, C.M., & Glaser, R. (1996). Expert knowledge and performance. In E. de Corte & F.E. Weinert (Eds.), *International encyclopedia of developmental and instructional psychology* (pp. 505-510). Oxford: Pergamon.

# Appendix A

## Participant Consent Letter

**McGill University**

To: Research Participants

From: Roger Azevedo
McGill University
Dept. of Educational and Counselling Psychology
Applied Cognitive Science Research Group
3700 McTavish St.
Montreal, Quebec, H3A 1Y2

Re: Purpose of Research Study

The Faculty of Medicine at McGill University is cooperating with the Department of Educational and Counselling Psychology to conduct a study of radiological expertise. More specifically, a systematic approach employing a combination of analytical methodologies and perceptual probes will be utilized to clarify the coexisting contributions of cognitive and perceptual factors in the development of radiological expertise. I am interested in understanding how radiology residents and staff radiologists diagnose breast diseases as depicted on mammograms. The types of activities that I would ask participants to assist me with include: (1) completing a participant information sheet, (2) solving 10 breast disease cases, and (3) being filmed and audio taped throughout the experimental session.

I, ______________________________ (print name), agree to participate in a research study conducted by McGill University's departments of Educational and Counselling Psychology and the Faculty of Medicine.

I understand the aforementioned conditions outlined in the covering letter and agree to become an active participant in this study.
I understand that my participation in the study will remain totally anonymous and confidential.
I understand that the data from this study may be published.
I understand my right to withdraw my consent and participation at any time without consequence.
I, ______________________________(signature), give the researcher permission to show segments of the video and audio recordings at research conferences.

**I HAVE CAREFULLY READ AND UNDERSTAND THE ABOVE AGREEMENT CONDITIONS AND WILLFULLY CONSENT TO PARTICIPATE IN THIS RESEARCH STUDY.**

Signature: ________________________ Date: ____________________

Phone Number:________________________

**Thank you for your participation !**

# Appendix B

## **Participant's Brief Questionnaire**

**Date:** _______
**Participant #:** _______

**Participant Information Sheet**

1. Identification # (the last 4 digits of your phone number):_______

2. Gender: (1) Male (2) Female

3. Medical specialty and level: (1) staff radiologist (2) radiology resident

4. Undergraduate, Graduate and Postgraduate Education:

| Period of study | Name and country of institution | Discipline | Degree | Year graduated (or indicate if in progress) |
|---|---|---|---|---|
| to | | | | |
| to | | | | |
| to | | | | |
| to | | | | |
| to | | | | |

5. Number of years as a medical physician _______

6. Number of years as a staff radiologist _______

7. Residency level: R1 R2 R3 R4 R5

8. Number of mammograms SEEN/READ during entire medical experience (approx.):
(1) 0-25 (2) 25-200 (3) 200-1000 (4) if > than 1000, please specify _______

9. Years of experience in mammography _______

10. Approximate number of mammograms SEEN per week (if applicable)_______

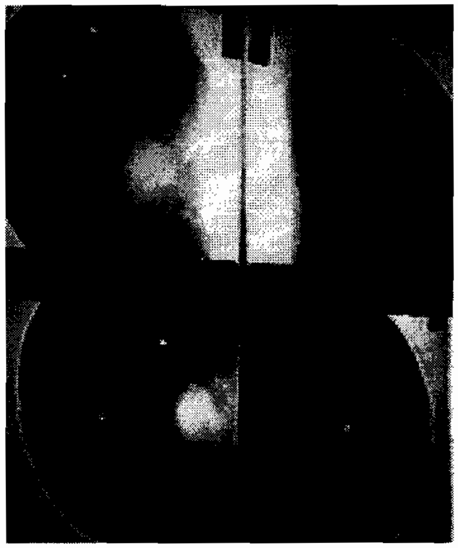
11. Approximate number of mammograms READ per week (if applicable)_______

# Appendix C

## The Clinical History, Mammograms, Consulting Radiologist's Diagnostic Protocol, Radiological Findings, Additional Radiological Observations, Diagnoses, Differential Diagnoses, and Subsequent Radiological Examinations for Case 1

**Clinical History:** A 60 year old woman who presented with a mass in the right breast.

**Mammograms:**



**Expert's Protocol:**
This case is that of a 60 year old woman who we are told presented with a right breast mass. On the mammogram she has dense fibroglandular tissue bilaterally and what catches our eye is a large mass in the right retro aereolar area. Now most of this mass is well-circumscribed but part of the mass, especially at the back is not. I would say, if I was to hazard a guess, that maybe about 30% of the mass is not well-circumscribed. So , in view of that we cannot be sure that the mass is benign but it could still be benign and represent a large fibroadenoma or cyst. It may be malignant and represent a well-circumscribed carcinoma. So we have to recommend a biopsy. Alternatively, we could do an ultrasound, but there is a catch a mucinous carcinoma can look very lucent like a cyst. I was afraid that this mass would be very sonoluscent, would look like a cyst and could mimic a carcinoma. So I recommended a biopsy and the biopsy result was a mucin-producing carcinoma or colloid carcinoma.

**Radiological Finding:**

A. large mass is partly well-circumscribed (i.e., mass is not completely well-circumscribed) (i.e., not entirely [or 75%] circumscribed) found on the right CC and OBL views

**Additional Radiological Observations:**

B. dense fibroglandular tissue bilaterally

C. large mass lies in the right retroaereolar area

**Differential Diagnoses:**

1. large fibroadenoma
2. large cyst
3. well-circumscribed carcinoma

**Diagnosis:**

1. **mucin-producing carcinoma (or colloid carcinoma or mucinous carcinoma)**

**Other Radiological Examinations:**

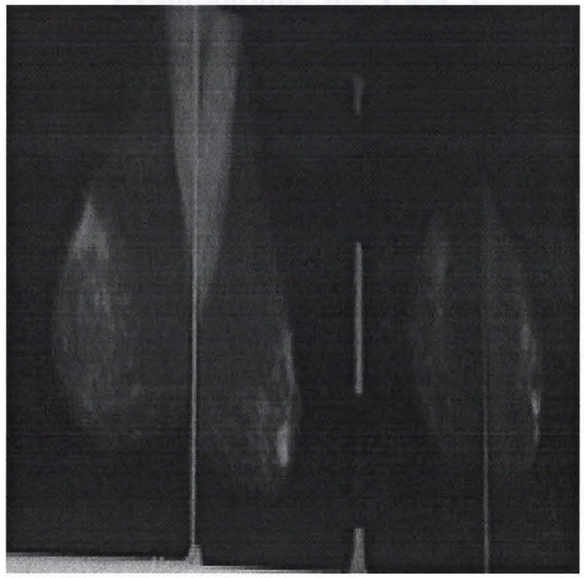
1. ultrasound (could be done) to verify solid nature of lesion
2. excisional biopsy (must be done)

# Appendix D

## The Clinical History, Mammograms, Consulting Radiologist's Diagnostic Protocol, Radiological Findings, Additional Radiological Observations, Diagnoses, Differential Diagnoses, and Subsequent Radiological Examinations for Case 2

**Clinical History:** A 54 year old woman who came for a check-up; there was no palpable mass. No significant past history and no previous mammograms.

**Mammograms:**



**Expert's Protocol:**

A 54 year old woman who came for a check-up and there was no palpable mass. What strikes our eye is that there is rather prominent fibroglandular tissue with several small nodules but two or three small nodules are standing out in the left retroaereolar area. I was not sure what those represented; they appear to be well-circumscribed although part of one of them is not well perfectly circumscribed. So I recommended an ultrasound and those turned out to be cysts. So I recommend a follow-up for smaller nodules found bilaterally but more so on the right. Two years after, a subsequent examination not included here showed the cysts were smaller and the other small nodules which looked benign had not increased in size. Now in the axilla you can see other nodules which have the appearance of small axillary lymph nodes. If a well-circumscribed nodule does not change in appearance or increase size in 2.5 years it is generally considered to be mostly probably benign.

**Radiological Findings:**

A. several small nodules
B. two or three larger nodules are well-circumscribed

**Additional Radiological Observations:**

C. prominent fibroglandular tissue
D. left retro aereolar area is prominent on the left
E. small axillary lymph nodes

**Diagnosis:**

1. **left retroaereolar cysts** (other smaller nodules not diagnosed but unchanged on follow-up and presumable benign.

**Other Radiological Examinations:**

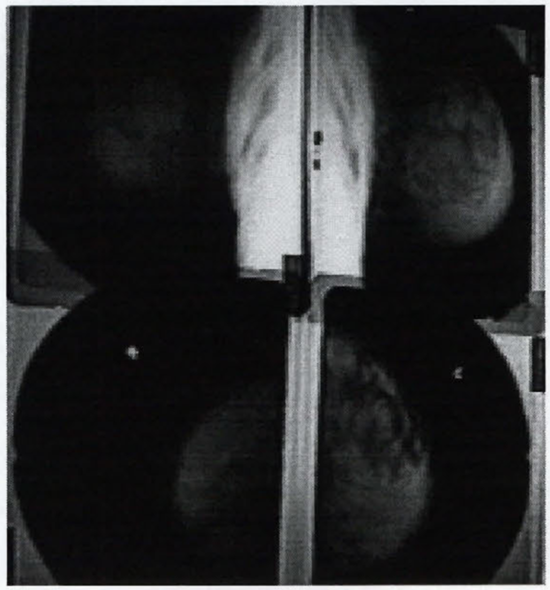
1. recommend follow-up, ultrasound was diagnostic

**Results of Subsequent Management:**

1. nodules in the left retroaereolar area shown to be cysts at ultrasound
2. smaller nodules were followed

**Clinical History:** This is a young patient in her thirties who presented with a large mass in the lateral portion of the left breast.

**Mammograms:**



**Expert's Protocol:**

This is a young patient who presented with a large mass in the lateral portion of the left breast. You can see here that the left breast is larger than the right and this mass is well-circumscribed and consists of fat and islands of fibroglandular tissue in it. Some people have compared this to a slice of sausage because it's so well-circumscribed and containing those islands of fibroglandular tissue in it. This is a hamartoma; and the other name is a fibroadenolipoma it consists of benign tissue which is independent from the remainder of the tissue but does not metastasize. The remainder of the exam shows dense fibroglandular tissue compatible with the patient's age.

**Radiological Findings:**

A. mass which is well-circumscribed, consists of fat, and contains encapsulated islands of fibroglandular tissue

**Additional Radiological Observations:**

A. dense fibroglandular tissue as usual at this patient's age

B. left breast is larger than the right, this asymetry is frequent in the normal population

**Diagnosis:**

1. **hamartoma or fibroadenolipoma** (a benign process)

**Other Radiological Examinations:**

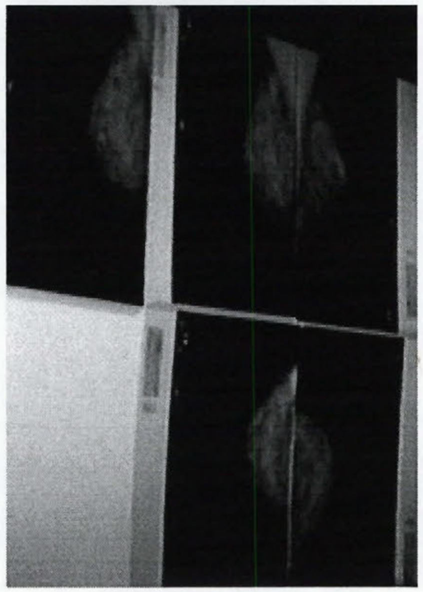
1. none

# Appendix H

## The Clinical History, Mammograms, Consulting Radiologist's Diagnostic Protocol, Radiological Findings, Additional Radiological Observations, Diagnoses, Differential Diagnoses, and Subsequent Radiological Examinations for Case 6

**Clinical History:** A woman in her fifties who presented with a mass in the upper outer quadrant of the right breast.

**Mammograms:**



**Expert's Protocol:**

This lady in her 50's she presented with a mass in the upper outer quadrant of the right breast. She shows dense fibroglandular tissue bilaterally plus a density associated with architectural distortion, and some retraction; this is highly suspicious for malignancy. The patient subsequently underwent a right mastectomy.

**Radiological Findings:**

A. a density associated with a loss of organization, that is architectural distortion and some retraction

**Additional Radiological Observations:**

B. dense fibroglandular tissue bilaterally

**Diagnosis:**

1. **Invasive intraductal carcinoma**

**Recommendations:**

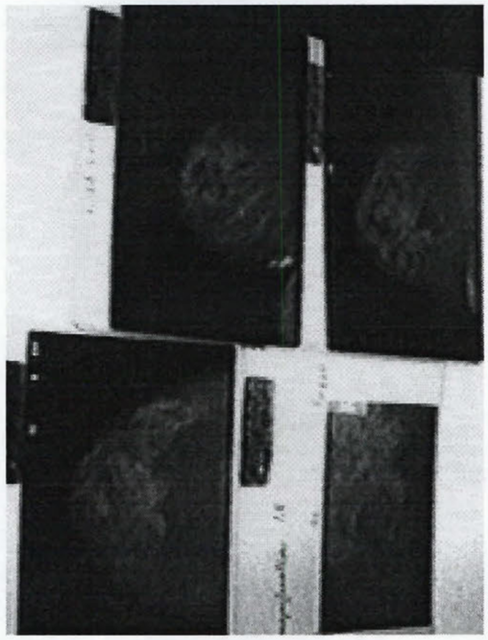
1. open biopsy (core or needle biopsy are acceptable)

# Appendix I

## The Clinical History, Mammograms, Consulting Radiologist's Diagnostic Protocol, Radiological Findings, Additional Radiological Observations, Diagnoses, Differential Diagnoses, and Subsequent Radiological Examinations for Case 7

**Clinical History:** An elderly patient who came for a routine examination. There was no abnormality on the physical examination. No previous mammograms.

**Mammograms:**



**Expert's Protocol:**
An elderly patient who came for a routine examination. There was no abnormality on the physical examination. On the mammogram there is a nodule in the central portion of the right breast and there are microcalcifications associated with this nodule; some are small and some are of different sizes. A magnification view was done. On that view they are been seen, they are pleomorphic microcalcification suspicious for malignancy. The pleomorphic microcalcifications have been compared to crushed stones. This was a carcinoma.

**Radiological Findings:**

A. a nodule in the center of the right breast
B. microcalcifications of different sizes (i.e., pleomorphic microcalcifications)

**Diagnosis:**

1. **carcinoma**

**Other Radiological Recommendations:**

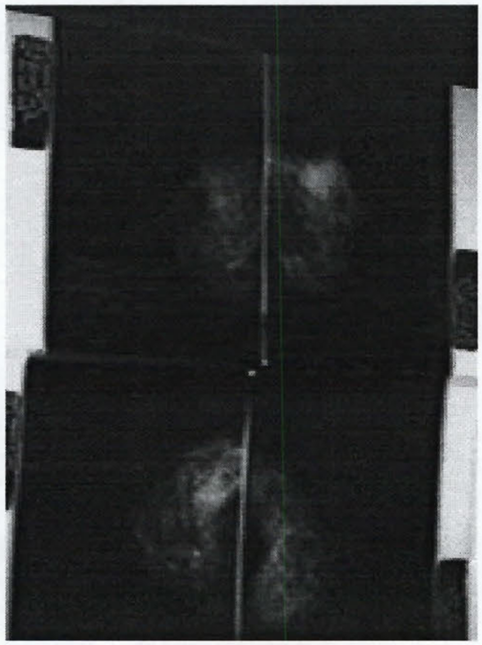
1. magnified view
2. excisional biopsy (or needle or core biopsy is acceptable - In the case of microcalcifications, needle or core biopsy is less accurate than excisional).

# Appendix J

## The Clinical History, Mammograms, Consulting Radiologist's Diagnostic Protocol, Radiological Findings, Additional Radiological Observations, Diagnoses, Differential Diagnoses, and Subsequent Radiological Examinations for Case 8

**Clinical History:** A woman in her early fifties who came without a palpable lesion; a few years previously she had breast surgery on the right but the results of surgery are not known.

**Mammograms:**



**Expert's Protocol:**

A woman in her early fifties who came without a palpable lesion and a few years she had breast surgery on the right breast. The result of surgery are not known. On the mammogram she has dense fibroglandular tissue bilaterally. At a first glance it looks like there is more tissue in the superior region of the left breast, but when we look in the cephalo-caudal this resolves pretty well. Looking more carefully at the right breast, we see small microcalcifications associated with a scar, from previous surgery; this is a very difficult problem. Those microcalcifications are irregular and they are pleomorphic. Sometimes fat necrosis can look like that but carcinoma can also have this appearance and we have to recommend a biopsy. The lesion was malignant and the patient had a mastectomy.

**Radiological Findings:**

A. irregular (pleomorphic) microcalcifications in a scar

**Additional Radiological Observations:**

B. dense fibroglandular tissue bilaterally

C. tissue in the superior region on the left

**Diagnosis:**

1. **carcinoma (on the right)**

**Other Radiological Examinations:**

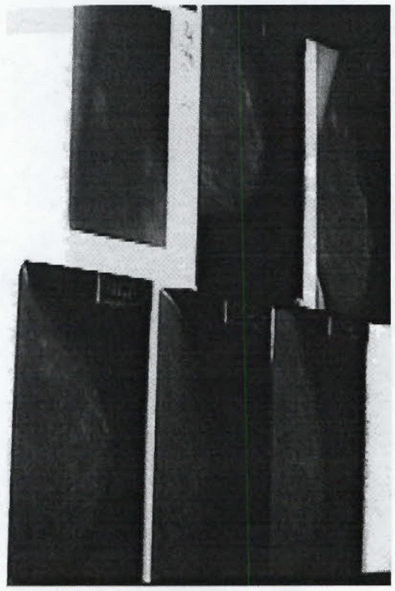
1. excisional biopsy (core biopsy and magnification views also acceptable)

# Appendix K

**The Clinical History, Mammograms, Consulting Radiologist's Diagnostic Protocol, Radiological Findings, Additional Radiological Observations, Diagnoses, Differential Diagnoses, and Subsequent Radiological Examinations for Case 9**

**Clinical History:** A 68 year old woman has a palpable abnormality in the lateral portion of the left breast (upper outer quadrant).

**Mammograms:**



**Expert's Protocol:**
She may have had a palpable abnormality in the lateral portion of the left breast upper outer quadrant. So on the oblique view we see a large spiculated lesion having the appearance of malignancy and we don't see much else at the first glance. But on the cephalo-caudal view we see a second small density which in retrospect could represent a second lesion. The technician took more views, there is a mediolateral view and another oblique view. These show a small area of retraction which appears to be a second carcinoma. She took a spot film which shows the two carcinomas. We know this one is in the upper outer quadrant of the left breast because we see it on two views. The larger one we are not sure more likely is in the axillary tail but we can't be sure of its exact location because we can't see it on the cephalo-caudal view. This another cephalo-caudal view showing the smaller lesion. All of the cephalo-caudal views show just one lesion and the oblique or medial-lateral obliques show the two lesions.

**Radiological Findings:**

A. large spiculated lesion having the appearance of malignancy
B. small retraction may represent a second carcinoma

**Diagnosis:**

**2 carcinomas**

**Other Radiological Examinations:**

1. spot film

# Appendix L

## The Clinical History, Mammograms, Consulting Radiologist's Diagnostic Protocol, Radiological Findings, Additional Radiological Observations, Diagnoses, Differential Diagnoses, and Subsequent Radiological Examinations for Case 10

**Clinical History:** A 60 year old woman who presented with inflammation of her left breast for two weeks.

**Mammograms:**

**Expert's Protocol:**
A woman 60 year old presented with inflammation in her left breast for two weeks. The breast was swollen, painful and red. What we see on the mammogram is some asymmetry; there is more tissue on the left than on the right. We don't see a discrete mass in the left breast, but there is skin thickening around the left breast. This tissue has a reticulated appearance as especially the supportive tissue including the Copper's ligament which is thickened. This can be due to several diseases but in this setup of relatively acute inflammation it could represent inflammatory mastitis which is a benign process or inflammatory carcinoma which is a rapidly progressing carcinoma with a poor prognosis. However, there are other things to be considered in the differential. Sometimes unilateral edema presents like this. For instance, if the patient is in heart failure you would expect edema of both breasts, but often the edema is unilateral. The same applies to renal failure or other causes of generalized edema. Secondly, if the patient had surgery or radiotherapy to the breast this will give the same appearance. More rare diseases like lymphoma of the breast, or metastatic disease of the breast, leukemia or acute trauma to the breast can look like that. This pattern is rarely described In tuberculosis or syphilis of the breast. Biopsy recommended.

**Radiological Findings:**
A. some asymmetry , B. no discrete mass in the left breast, C. skin thickening around the left breast, and D. tissue has a reticulated appearance

**Differential Diagnoses:**
1. inflammatory mastitis, 2. unilateral edema, 3. lymphoma of the breast, 4. metastasis to the breast, 5. leukemia to the breast, 6. post-radiation, 7. post-trauma

**Diagnosis:**
1. **inflammatory carcinoma**

**Other Radiological Examinations:**

1. biopsy recommended

# Appendix M

## Description of the Breast Disease Cases

**Case 1: Mucin-Producing Carcinoma**

This case is that of a 60 year old woman who presented with a mass in the right breast (see Appendix C). The mammograms depict a radiological finding consisting of a large mass in the right retroaereolar area and a radiological observation consisting of dense fibroglandular tissue bilaterally. Most of the mass is well-circumscribed, but approximately 30% of the inferior part is not. This finding is not entirely suggestive that the mass is benign representing a large fibroadenoma or large cyst. Instead, it may be malignant (and represent a well-circumscribed carcinoma). A biopsy is recommended to clarify the nature of the finding. An ultrasound could also have been performed. However, this procedure would not have been very effective given the lucency of the mass. The biopsy results determined that this mass was a mucin-producing carcinoma (also known as colloid carcinoma, or mucinous carcinoma). This is an atypical case and somewhat rare in North America. Knowledge of the clinical mammography literature should facilitate the diagnosis of this case more than clinical experience.

**Case 2: Left Retroaereolar Cysts**

This is the case of a 54 year old woman who came for a check-up and there was no palpable mass (see Appendix D). The mammograms depict two radiological findings each comprised of several small nodules and few larger well-circumscribed ones. In addition, there are three radiological observations: prominent fibroglandular tissue, the left retroaereolar area being more prominent than the right, and small axillary lymph nodes. The expert's recommendations included a follow-up for the smaller nodules found bilaterally. In the axilla there were some other nodules which had the appearance of small axillary lymph nodes. The diagnosis of left retroaereolar cysts was obtained using ultrasound. The diagnosis is frequent but the mammographic presentation is unusual. Knowledge of the clinical mammography literature knowledge and clinical experience should facilitate the diagnosis of this case.

**Case 3: Infiltrating Intraductal Carcinoma**

This is the case of a woman in her late forties who presents with a palpable nodule in the lateral portion of left breast (see Appendix E). The mammograms depict two radiological findings, including an area of retraction with some disorganization in the lateral portion of the left breast and a poorly defined nodule in the center of the retraction and disorganization. In addition, there is a radiological observation in the form of dense fibroglandular tissue. On the magnification view, the poorly defined nodule which is associated with retraction is very suspicious for malignancy. The patient had a biopsy which confirmed a diagnosis of infiltrating intraductal carcinoma. This is a difficult case and illustrates a rare mammographic manifestation of breast cancer. Clinical experience should facilitate the diagnosis of this case.

**Case 4: Calcium in Small Cysts**

This is the case of a woman in her sixties who presented with no palpable mass and no abnormality on physical examination but had a family history of carcinoma (her daughter had breast cancer) (see Appendix F). The mammograms depict one radiological finding, round calcifications on CC view which appear as linear calcifications on MLO view. In addition, there are two radiological observations comprised of several round benign-looking nodules and glandular tissue with a nodular appearance. These represent calcium in tiny cysts, a benign phenomenon. No further examinations were required. The radiological manifestations represent a well-known phenomena that is frequently diagnosed in breast clinics. Knowledge of the clinical mammography literature knowledge and clinical experience should facilitate the diagnosis of this case especially for experienced radiologists.

**Case 5: Hamartoma or Fibroadenolipoma**

This is the case of a young patient who presented with a large mass in the lateral portion of the left breast (see Appendix G). The mammograms depict one radiological finding, a well-circumscribed mass consisting of fat and containing encapsulated islands of fibroglandular tissue. In addition, there are two radiological observations comprised of dense fibroglandular tissue (typical given the patient's age). The left breast is larger than the right, but this asymmetry is frequently encountered in the normal population. This case was diagnosed as a hamartoma or a fibroadenolipoma which consists of benign tissue which is independent from the remainder of the tissue but does not metastasize. The expert did not order any further radiological examinations. This case is rare and therefore knowledge of the clinical mammography literature should facilitate the diagnosis of this case.

**Case 6: Invasive Intraductal Carcinoma**

This is a case of a woman in her fifties who presented with a mass in the upper outer quadrant of the right breast (see Appendix H). The mammograms depict one radiological finding comprised of a density associated with disorganization (i.e., architectural distortion and some retraction). In addition, there is a radiological observation comprised of dense fibroglandular tissue bilaterally. This is highly suspicious for malignancy. An open biopsy revealed a diagnosis of invasive intraductal carcinoma. The patient subsequently underwent a right mastectomy. The manifestations of cancer depicted in this case are routinely encountered in breast clinics.

**Case 7: Carcinoma**

This is the case of an elderly patient who came for a routine examination (see Appendix I). There was no abnormality on the physical examination. The mammograms depict two radiological findings in the center of the right breast, a nodule and microcalcifications of different sizes (i.e., pleomorphic microcalcifications). The pleomorphic microcalcifications are better characterized on the magnification view and they are suspicious for malignancy. This was diagnosed as a carcinoma. Subsequent radiological recommendations would included a magnified view followed by an excisional biopsy. The radiological manifestation is typical and routinely encountered in clinical practice. Clinical experience should facilitate the diagnosis of this case, however, residents may have seen the mammographic features in textbooks.

**Case 8: Carcinoma**

This is the case of a woman in her early fifties presented with no palpable lesion and she had previously had surgery on the right breast. The results of previous surgery were not known (see Appendix J). The mammograms depict one radiological finding comprised of irregular (pleomorphic) microcalcifications in a scar from previous surgery. In addition, there are two radiological observations, dense fibroglandular tissue bilaterally and tissue in the superior region of the left breast. An excisional was performed and the lesion was diagnosed as being malignant. The patient subsequently underwent a mastectomy. The radiological manifestation is typical and routinely encountered during clinical practice. Clinical experience should facilitate the diagnosis of this case and the mammographic features are described in mammography textbooks.

**Case 9: Two Carcinomas**

This is the case of a 68 year old woman who presented with a palpable abnormality in the lateral portion (upper outer quadrant) of the left breast (see Appendix K). The mammograms depict two radiological findings, a large spiculated lesion having the appearance of malignancy and a small retraction which could have represented a second carcinoma. The woman may have had a palpable abnormality in the lateral portion of the left breast (in the upper outer quadrant) but this was not determined clinically. The MLO view depicts a large spiculated lesion having the appearance of malignancy. On the CC view a second small density is visible which could have represented a second lesion. Additional mammograms were taken and showed a small area of retraction which appeared to correspond to the second carcinoma. The two carcinomas were visible on the spot film. In summary, there is one carcinoma is in the upper outer quadrant of the left breast which is visible on two views. The larger carcinoma was likely considered to be located in the axillary tail. However, the exact location is problematic since it did not appear on the CC view. All of the CC views showed just one lesion and the MLO views showed two lesions. A subsequent spot film of the larger carcinoma was taken. The radiological manifestation is typical and routinely encountered during clinical practice. Clinical experience should facilitate the diagnosis of this case.

**Case 10: Inflammatory Carcinoma**

This is a case of a 60 year old woman who had experienced inflammation in her left breast for two weeks (see Appendix L). On clinical examination, the breast was swollen, painful and red. The mammograms depict four radiological findings, some asymmetry, no discrete mass in the left breast, skin thickening around the left breast, and the tissue with a reticulated appearance. This could have represented several diseases but in this case of relatively acute inflammation it could represent inflammatory mastitis which is a benign process. This could have an also represented an inflammatory carcinoma which is a rapidly progressing cancer with a poor prognosis. However, in this case the differential diagnoses may have included unilateral edema, renal, radiotherapy to the breast, lymphoma, metastatic disease, leukemia, acute trauma, tuberculosis or syphilis of the breast. A biopsy was performed and the diagnosis was inflammatory breast carcinoma. Clinical experience and knowledge of the clinical literature in mammography should facilitate the diagnosis of this case.

# Appendix N

## A Detailed Account of the Problem Solving Operators

| Operator | Definition | Example |
|---|---|---|
| **Data Acquisition:** This operator was used to acquire clinical cues from the clinical history and set of mammograms. The participant typically reads the clinical history out loud and inspects the mammograms while placing them on a view box. | | |
| Reads Clinical History | Participant read the clinical history attached to the envelope containing the set of mammograms for each case. | "... case 1A. A 60 year old woman who presented with a mass in the right breast" |
| Places and Inspects Mammograms | Participant removed the set of mammograms from the envelope, placed them on the viewbox and visually inspected them . | |
| **Data Identification:** This operator was used to verify that each mammogram in a set was taken at the same point in time (on the same day). | | |
| Identifies Mammograms as Single Study | Participant verbalized some aspect related to the number of mammograms in the set, including additional views (e.g., a magnification view) and/or the date when the mammograms were taken. | "... okay, so we have bilateral CC and MLO films taken on the same day" |
| **Data Assessment:** These operators were used to comment on the technical quality of the films and the may include a participant's comment regarding the inability to characterize a mammographic finding due to the inferior quality of the films. | | |
| Comments on the Technical Quality of the Mammograms | Participant verbalized related to the technical aspects of the films including poor penetration, positioning, or absence of anatomical features such as the pectoralis muscle. | "... the contrast is these examinations is somewhat low" |
| Inability to Assess Mammographic Cue | Participant verbalized inability to characterize a mammographic observation, finding or anatomical region of the breast due to inferior film technical quality. This operator was also used when a participant had attempted to characterize something but could not resolve the appearance of the mammographic cue due to poor film quality. | "... I'm not sure...It's not presented enough to help me with this lesion" |

| Operator | Definition | Example |
|---|---|---|
| **Data Examination:** These operators were employed to interpret cues presented in a set of mammograms. Generally, these operators are used to determine the significance of a cue in terms of whether it represents a radiological observation or radiological finding. Once a cue has been recognized and interpreted, its significance can be used as an input for subsequent information-processing operations. The third operator is used to determine if a mammographic cue is an actual mammographic cue or a film artifact. | | |
| Identifies Mammographic Cue | The participant's goal when using this operator was to select, examine and characterize mammographic cues and make initial interpretations of mammographic observations and findings. | "... the main abnormality is in the right breast where there is a large well-circumscribed lesion" |
| Identifies Marked Mammographic Cue | The participant actually selected an abnormality in a case presented under the augmented experimental condition which had been highlighted with arrows. | "... I do see the mass which has been highlighted by arrows" |
| Identifies Artifacts the Mammograms | The participant referred to artifacts that were produced either during the process of obtaining the mammograms or from repeated handling. | "... There are dirt specs on these films which are not calcifications" |
| **Data Exploration:** These operators were employed to conduct a more detailed review or further interpretation of cues previously selected from the clinical history and set of mammograms. Typically, it involves further characterizing mammographic cues (observation and findings) according to their location, shape, size, margins, density, number, distribution, description, and/or associated findings. it may involve noting mammographic cues that are not present on the films, and re-reading cues from the clinical history. | | |
| Characterizes Mammographic Cue | The participant characterizes a previously selected mammographic cue. | "... measuring 6 by 4.5 cm in diameter" |
| Characterizes Associated Mammographic Cue | The participant characterizes a mammographic cue that is associated with another cue selected previously. | " ... and the mass .... has a significant fat contents and contains multiple soft tissue nodular densities" |
| Notes-Absent-Data | The participant notes that a particular cue is not in the case data. | "... and there are no calcifications within this mass" |

| Operator | Definition | Example |
|---|---|---|
| Re-Reads Clinical History | The participant re-reads the clinical history or selectively identifies or repeats a specific cue or cues from the clinical history which may facilitate the search for a mammographic cue or put the patient's clinical history is context with the mammographic manifestations. | "... okay so it's in the right upper outer quadrant" |

**Data Comparison:** These operators were employed in the comparison of mammographic cues (observation and findings) to: (1) different views of the same breast (ipsilateral), (2) different views of the other breast (contralateral), or (3) to the rest of the same breast. These comparisons are critical in determining the (1) presence of a mammographic cue, (2) level of abstraction at which a cue will be characterized, and (3) level of abstraction at which the cue will be classified.

| | | |
|---|---|---|
| Compares to Ipsilateral | The participant compares a mammographic cue on one view to an alternative view of the same breast. A cue presented on the CC view of the left breast is compared to the MLO view of the right breast. | "... partially well-defined nodules on the CC view which appear well-defined on the MLO view" |
| Compares to Contralateral | The participant compares a mammographic cue on one view to the same cue on the same view of the other breast. A cue presented on the CC view of the left breast is compared to the CC view of the left breast. | "... it's hard to for me to say with certainty whether there's also microcalcifications on the left side" |
| Compares Mammographic Cue to Additional Views | The participant compares a mammographic cue on one view to other views of the same breast. A cue present in the right breast is compared to a magnified view of the same breast region. | "... the margins appear well-defined on the MLO view, but the mag view shows the mass to have extremely well-defined margins" |
| Compares to Rest of Breast | The participant compares the presence of a mammographic cue to the rest of the breast (on the same view). A cue present on the CC view right breast in compared to the rest of the breast (CC view of the right breast). | "... there is a very large lesion more than 10 cm in diameter in the upper left quadrant of the right breast which is less dense than the rest of the breast tissue" |

| Operator | Definition | Example |
|---|---|---|
| **Data Explanation:** These operators were used to interpret the significance or role of clinical history cue(s)or mammographic cue(s) by explaining the underlying pathophysiological basis of a disease. Often, this type of data explanation is not used to generate a particular hypothesis, but is used to account for the natural history which could have led to the development of data finding through the use of either implicit (conditional) or explicit (causal) semantic relationships. | | |
| Infer-Pathophysiological-Cause (of a Cue) | The participant goes beyond merely noting whether a cue represents an abnormal finding. This operator is used to infer a pathophysiological condition underlying a particular cue. | "... and when you see that kind of abnormal architecture and density in a fatty breast you have to wonder about something going on underneath" |
| Infer-Natural-History (of a Cue) | The participant uses this operator to explain the developmental course of a particular cue with respect to a disease or pathophysiological conditions. | "... density presented here is probably due to scarring from the previous surgery" |
| **Data Classification:** These operators were used to classify a set of mammograms, mammographic cues, or the identification and classification of mammographic cues. This operator is usually used prior to hypothesis generation. | | |
| Classifies Mammographic Cue | The participant classified a mammographic cue. | "... a cluster of microcalcifications which overall have a benign-looking appearance" |
| Identifies & Classifies Mammographic Cue | The participant identifies and classifies a mammographic cue. | "... asymmetry seen in the left upper outer quadrant and medial portions of the breast which represent fibroglandular tissue " |
| Classifies Mammograms | The participant classifies a set of mammograms. | "... left breast is normal" |

| Operator | Definition | Example |
|---|---|---|
| **Hypothesis Generation:** These operators were used to generate diagnostic hypotheses. Specific operators provide different means for generating hypotheses in response to a particular data cue, pattern of cues, data classification, or other diagnostic hypotheses. Is it important to note that cases with multiple "diagnosable" benign and malignant findings require the use of separate operators for each mammographic cue. | | |
| Trigger | This operator generates the first use of a hypothesis. In response to clinical history or mammographic cue(s), a participant may propose either the possibility or unlikelihood of a disease, a pathophysiological a hypothesis, a category hypothesis, a set of differential diagnoses, or a diagnosis. | "invasive ductal carcinoma", "suspicious for malignancy", or "carcinoma" |
| Further-Specification | This operator triggers a hypothesis which is subsumed by a previously generated hypothesis. For example, after generating the hypothesis of carcinoma, the participant at a later point may further specify one or more specific types carcinomas, such as invasive ductal carcinoma. A participant can also use the further-specify operator to generate a specific variant of a previously generated disease hypothesis. | A participant can further-specify a "benign lesion hypothesis" by generating the possibility of a "hamartoma". |
| Generalization | This operator is the inverse of further-specification. If a generated hypothesis is determined to be too specific, it is sometimes necessary to ease the constraints and consider a more general, but still related, type of hypothesis. In this case the data, although suggestive, could not (yet) discriminate between two diseases which are both members of a more general category; therefore, the radiologist "steps back" to a more general hypothesis which can explain the data and support the existing (and competing) hypotheses. | A radiologist could state, "Carcinoma might be relevant, but so might sclerosing adenosis ... the data is ambiguous so far... well, we do know we are dealing with a mass that is very suspicious for malignancy." |

| Operator | Definition | Example |
|---|---|---|
| Unspecified Diagnosis | This operator fails to specify a diagnosis (at any level of abstraction). A participant may have specified a diagnostic previously during the diagnostic task but then fails to specify a diagnosis usually during the last segment of the protocol. | "... it is most likely a benign condition..............that's how far I would go. So in conclusion, I don't know what this is" |

**Hypothesis Evaluation:** These operators were used to evaluate a diagnostic hypothesis by interpreting its ability to account for the presence or absence of clinical history cues, mammographic cues, data interpretations, or its relationship to other diagnostic hypotheses. Hypothesis evaluations usually assume either a confirmatory or disconfirmatory value.

| | | |
|---|---|---|
| Confirmation | A cue is interpreted as being consistent with a hypothesis and provides a confirmatory value for the hypothesis. | "... the calcifications are consistent with malignancy" |
| Disconfirmation | A cue is interpreted as being inconsistent with a hypothesis and provides a disconfirmatory value for the hypothesis. | "... it's not solid enough nor uniform enough to represent a philloides tumor" |
| Causal Relationship | A cue is used when expected or implied causal relationships between two or more diagnostic hypotheses are considered as confirmatory or disconfirmatory evidence. | "... I do not believe this represents a case of lymphomatis mastitis, melanoma, or mastitis because it's surrounded by fat" |

**Summarization:** This operator was used to maintain clinical history and mammography cues in active memory and involves operators such as *repeating significant facts*, acquired from the case.

| | | |
|---|---|---|
| Repeat Data | The participant repeats a cue from the clinical history and/or set of mammogram and subsequently repeats while performing the diagnostic task. | "... to conclude there is a large mixed density mass in the upper outer quadrant of the left breast" |

# Appendix O

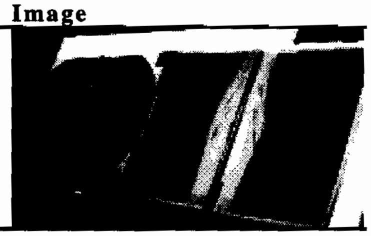
## A Sample of the Coded Protocols

| S1 - 5A | staff radiologist | |
|---|---|---|
| **Seg. #** | **Transcript** | **Image** |
| 1 | Thirty year old woman, large mass in the lateral portion of the left breast.<br>Data Examination ⇒ Reads Clinical History |  |
| 2 | Large mass in the right portion of the left breast.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Size [large]*<br>*Region [right portion]*<br>*BodyLocation [left breast]* | |
| 3 | Well, she's a young patient with dense breasts.<br>Data Examination ⇒ Identifies Mammographic Cue<br>*Radiological Observation #1 [dense breasts]*<br>ASSOC:<br>*[age]* | |
| 4 | The right breast just grossly looks normal.<br>Data Classification ⇒ Classifies Mammographic Cue<br>*Radiological Observation #2 [right breast]*<br>*Degree [grossly]*<br>*Classification [normal]* | |
| 5 | One coarse benign calcification, it's not very dense, otherwise unremarkable.<br>Data Classification ⇒ Identifies & Classifies Mammographic Cue<br>*Radiological Observation #3 [calcification]*<br>*Number [one]*<br>*Description [not very dense]*<br>*Classification [benign]* | |
| 6 | Okay so, the obvious finding is the huge upper outer quadrant mass made-up of mixed elements.<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Size [huge]*<br>*Region [upper outer quadrant]*<br>*Description [comprised of mixed elements]* | |
| 7 | There is some fat in it, and in the fat there is a bunch of solid lobulated nodules.<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Size [huge]*<br>*Region [upper outer quadrant]*<br>*Description [fat and solid lobulated nodules]* | |
| 8 | I don't know what this is.<br>*Hypothesis Generation ⇒ Trigger*<br>*Radiological Finding #1 [mass]*<br>Classification [ ] | |
| 9 | I won't be worried it's a cancer,<br>Hypothesis Generation ⇒ Further Specification<br>*Radiological Finding #1 [mass]*<br>*Classification [not cancer]* | |

| | |
|---|---|
| 10 | but nonetheless it ah probably needs to come out since it's palpable.<br>Diagnostic Plan ⇒ Subsequent Examinations<br>*Radiological Finding #1 [mass]*<br>*Recommendation #1 [excision]*<br>*COND:[palpable]* |
| 11 | The rest of it doesn't disturb me.<br>Data Classification ⇒ Classifies Mammograms<br>*Classification [not suspicious for malignancy]* |
| 12 | Some kind of lipoma, fibroadenolipoma, some bizarre diagnosis I don't know about.<br>Hypothesis Generation ⇒ Generalization<br>*Radiological Finding #1 [mass]*<br>*Classification [lipoma] OR [fibroadnolipoma] OR []* |

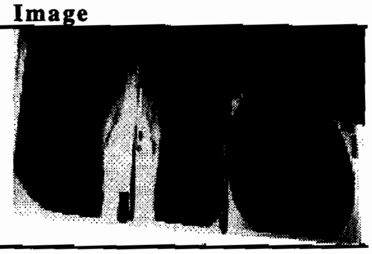
**S5 - 5A staff radiologist**

| Seg. # | Transcript | Image |
|---|---|---|
| 1 | ....................................<br>*Data Acquisition ⇒ Places & Inspects Mammograms* |  |
| 2 | Case 5A. This is a young patient in her 30s who presented with a large mass in the lateral portion of the right breast.<br>*Data Acquisition ⇒ Reads Clinical History* | |
| 3 | I examine both breasts for symmetry<br>GOAL [examine both breasts for asymetry] | |
| 4 | and there's an obvious 13.5 by 10.5 complex non-homogeneous lesion in the left upper outer quadrant.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Finding #1 [lesion]*<br>*Quantity [an]*<br>*Size [13.5 by 10.5 cm]*<br>*Density [complex non-homogeneous]*<br>*BodyLocation [left breast]*<br>*Region [upper outer quadrant]* | |
| 5 | Ahm this lesion has a significant fat content and contains multiple<br>soft tissue ah nodular densities.<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Finding #1 [lesion]*<br>*Degree [significant]*<br>*Density [fat content]*<br>AND<br>*Quantity [multiple]*<br>*Density [soft tissue nodular densities]* | |
| 6 | The possibility of a hamartoma should be considered.<br>*Hypothesis Generation ⇒ Trigger*<br>*Certainty [possibility]*<br>*Classification [hamartoma]* | |
| 7 | Ah .... additional views would not add very much. I would suggest ah a surgical biopsy to clarify diagnosis.<br>*Diagnostic Plan ⇒ Subsequent Radiological Examinations*<br>*Recommendation [surgical biopsy to clarify diagnosis]* | |

| S7 - 5A | staff radiologist | |
|---|---|---|
| **Seg.#** | **Transcript** | **Image** |
| 1 | *Data Acquisition ⇒ Places & Inspects Mammograms* |  |
| 2 | So this is a young patient in her thirties who presented with a large mass in the lateral portion of the left breast.<br>*Data Acquisition ⇒ Reads Clinical History* | |
| 3 | The breast tissues are very dense bilaterally<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #1 [breast tissue density]*<br>*Degree [very]*<br>*BodyLocation [left breast]*<br>AND<br>*Radiological Observation #2 [breast tissue density]*<br>*Degree [very]*<br>*BodyLocation [right breast]* | |
| 4 | due to abundant fibroglandular tissue and dysplasia.<br>*Data Explanation ⇒ Infer-History*<br>*Radiological Observation #3 [fibroglandular tissue]*<br>*Amount [abundant]*<br>AND<br>*Radiological Observation #4 [dysplasia]*<br>CAU:<br>*Radiological Observation #1 [breast tissue density]*<br>AND<br>*Radiological Observation #2 [breast tissue density]* | |
| 5 | There is a very large more than 10 cm in diameter well-defined mass lesion in the upper outer quadrant of the left breast<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Finding #1 [mass lesion]*<br>*Size [very large more than 10 cm in diameter]*<br>*Margins [well-defined]*<br>*Region [upper outer quadrant]*<br>*BodyLocation [left breast]* | |
| 6 | which is less dense than the breast tissue<br>*Data Comparison ⇒ Compare to Rest of Breast*<br>*Radiological Finding #1 [mass lesion]*<br>*Radiological Observation #5 [breast tissue]*<br>*Degree [less dense]*<br>*BodyLocation [left breast]* | |
| 7 | and shows several nodular densities within the lucency.<br>*Data Exploration ⇒ Characterizes Associated Mammographic Cue*<br>*Radiological Observation #6 [nodular densities]*<br>*Quantity [several]*<br>ASSOC:<br>*Radiological Finding #1 [mass lesion]* | |

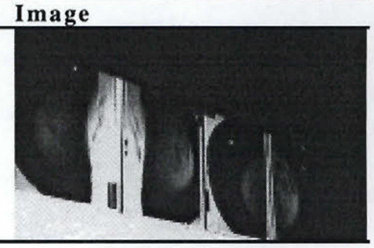
| | |
|---|---|
| 8 | The lucency is suggestive of a fat-containing structure....<br>*Data Classification ⇒ Classifies Mammographic Cue*<br>*Radiological Finding #1 [lucency]*<br>COND:<br>*Classification [fat-containing structure]* |
| 9 | so I must suspect that it is a lipomatos type of mass.....<br>*Hypothesis Generation ⇒ Trigger*<br>*Radiological Finding #1 [lucency]*<br>*Classification [lipomatos-type of mass]* |
| 10 | But the nodular components are somewhat disturbing.<br>*Data Classification ⇒ Classifies Mammographic Cue*<br>*Radiological Observation #5 [nodular densities]*<br>*Classification [somewhat disturbing]* |
| 11 | I would advise to do an ultrasound examination.<br>*Diagnostic Plan ⇒ Subsequent Radiological Examination*<br>*Radiological Finding #1 [mass lesion]*<br>AND<br>*Radiological Observation #5 [nodular densities]*<br>*Recommendation [ultrasound]* |
| 12 | I really don't know what this lesion is......<br>*Hypothesis Generation ⇒ Generalization*<br>*Radiological Finding #1 [mass lesion]*<br>AND<br>*Radiological Observation #5 [nodular densities]*<br>*Classification [ ]* |
| 13 | I must presume that it is a mixed mass lesion with fibroadenoma type of components and fat.<br>*Hypothesis Generation ⇒ Further Specification*<br>*Radiological Finding #1 [mass lesion]*<br>AND<br>*Radiological Observation #5 [nodular densities]*<br>*Certainty [presume]*<br>*Classification [mixed mass lesion with fibroadenoma type of components and fat]* |
| 14 | It is most likely a benign condition.<br>*Hypothesis Generation ⇒ Generalization*<br>*Radiological Finding #1 [mass lesion]*<br>AND<br>*Radiological Observation #5 [nodular densities]*<br>*Certainty [most likely]*<br>*Classification [benign condition]* |
| 15 | That's how far I can go. I don't know what this is!<br>Hypothesis Generation ⇒ *Unspecified Diagnosis*<br>*Radiological Finding #1 [mass lesion]*<br>AND<br>*Radiological Observation #5 [nodular densities]*<br>*Classification [ ]* |

**S9 - 5A staff radiologist**

| Seg. # | Transcript | Image |
|---|---|---|
| 1 | Next patient is a young patient in her 30s who presented a large mass in the lateral portion of the left breast.<br>*Data Acquisition ⇒ Reads Clinical History* |  |
| 2 | This is a very unusual case. I've not seen a case like this before.<br>*Meta-Reasoning ⇒ Experiential-Memory* | |
| 3 | The breasts overall are ah very dense<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #1 [breast tissue density]*<br>*Degree [very]*<br>*BodyLocation [left breast]*<br>AND<br>*Radiological Observation #2 [breast tissue density]*<br>*Degree [very]*<br>*BodyLocation [right breast]* | |
| 4 | but in the left breast we note ah an area of decreased attenuation ah and within it multiple well-defined nodules.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Finding #1 [area of attenuation]*<br>*BodyLocation [left breast]*<br>*Number [an]*<br>*Degree [decreased]*<br>ASSOC:<br>*Radiological Observation #3 [nodules]*<br>*Quantity [multiple]*<br>*Margins [well-defined]* | |
| 5 | I would like to take more history from this patient to see if she's breast feeding, if she has ah fever, chills, any signs that this may be an infectious etiology, whether perhaps she has a malignancy elsewhere.<br>*Diagnostic Plan ⇒ Subsequent Examinations*<br>*Radiological Finding #1 [area of attenuation]*<br>AND<br>*Radiological Observation #3 [nodules]*<br>*Recommendation #1 [take more history]*<br>*ORDER [breast feeding] [sign of fever] [chills]*<br>*[signs of infectious etiology] [malignancy elsewhere]* | |
| 6 | Ahm.. overall though I must say that the nodules themselves have ah well-defined borders,<br>*Data Exploration ⇒ Characterizes Associated Mammographic Cue*<br>*Radiological Observation #3 [nodules]*<br>*Margins [well-defined]* | |
| 7 | and whatever this is, it most likely represents a benign process<br>*Hypothesis Generation ⇒ Trigger*<br>*Radiological Finding #1 [area of attenuation]*<br>AND<br>*Radiological Observation #3 [nodules]*<br>*Certainty [most likely represent]*<br>*Classification [benign process]* | |

| | |
|---|---|
| 8 | but not knowing what the diagnosis is, and not having seen this before ah.. if I were reading this case ah I would ah consult my colleagues or text.<br>*Diagnostic Plan ⇒ Subsequent Examinations*<br>*Classification [ ]*<br>*AND*<br>*Radiological Finding #1 [area of attenuation]*<br>*AND*<br>*Radiological Observation #3 [nodules]*<br>*COND:*<br>*Recommendation #2 [consult colleagues]*<br>*OR-EXCLUSIVE*<br>*Recommendation #3 [consult textbooks]* |
| 9 | If they had no opinion I would at least try to put a needle into one of these nodules and to get some histology or pathology from them.<br>*Diagnostic Plan ⇒ Subsequent Examinations*<br>*Recommendation #1 [consult colleagues]*<br>*IF [no opinion]*<br>*THEN Recommendation #4*<br>*[insert needle in 1 of nodules to get some histology]*<br>*AND [insert needle in 1 of nodules to get some pathology]* |
| 10 | The ah differential diagnosis I suppose could include something malignant like unusual liposarcoma with elements of degeneration within it and fat around it. Ahm a lipoma I suppose with ah hemorrhage within it ah I think is less likely.<br>*Hypothesis Generation ⇒ Further Specification*<br>*Radiological Finding #1 [area of attenuation]*<br>*AND*<br>*Radiological Observation #3 [nodules]*<br>COND:<br>*Certainty [could include]*<br>*Classification [malignant like unusual liposarcoma with elements of degeneration within it and fat around it]*<br>OR-EXCLUSIVE<br>*Certainty [suppose]*<br>*Classification [lipoma with hemorrhage within it]* |
| 11 | I do no believe that this represents a case of lymphomatis, mastastis, ah melanoma, or other mastatises because it's surrounded by an area of fat.<br>*Hypothesis Evaluation ⇒ Causal-Relationship*<br>*Radiological Finding #1 [area of attenuation]*<br>*AND*<br>*Radiological Observation #3 [nodules]*<br>*NEG COND:*<br>*Classification [lymphomatis] [mastastis] [melanoma] [other mastatises]*<br>*[surrounded by an area of fat]* |
| 12 | So in conclusion, I don't know what this is!<br>*Hypothesis Generation ⇒ Unspecified Diagnosis*<br>*Radiological Finding #1 [area of attenuation]*<br>*AND*<br>*Radiological Observation #3 [nodules]*<br>*Classification [ ]* |

| | | |
|---|---|---|
| 13 | but I would probably stick a needle into it if I had no other clinical information.<br>*Diagnostic Plan ⇒ Subsequent Radiological Examinations*<br>    *Recommendation #1 [consult colleagues]*<br>    *IF [Recommendation #1 - clinical info not available]*<br>    *THEN [perform Recommendation #4]* | |

| S13 - 5A | radiology resident | |
|---|---|---|
| **Seg. #** | **Transcript** | **Image** |
| 1 | So, this is a young patient in her 30's who presented with a large mass in the lateral portion of the left breast.<br>*Data Acquisition ⇒ Reads Clinical History* |  |
| 2 | So we are dealing with dense breasts bilaterally.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #1 [breast tissue density]*<br>*BodyLocation [left breast]*<br>AND<br>*Radiological Observation #2 [breast tissue density]*<br>*BodyLocation [right breast]* | |
| 3 | However, the left breast is very nodular<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #3 [nodular breast]*<br>*BodyLocation [left breast]*<br>*Degree [very]* | |
| 4 | and ah presents with very large nodules scattered throughout the breast.<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Observation #3 [nodules]*<br>*Size [very large]*<br>*Quantity [nodules]*<br>*Distribution [scattered throughout]*<br>*BodyLocation [left breast]* | |
| 5 | There is a arch-like, almost looks like a lot of skin thickening although I doubt that's what it is.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #4 [arch-like skin thickening]*<br>*Certainty [although I doubt that's what it is]* | |
| 6 | But there is a circumsferential density all around the breast that is kind of encasing the breast.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #5 [density]*<br>*Quantity [a]*<br>*Description [circumsferential]*<br>*Distribution [all around encasing the breast]*<br>*BodyLocation [left breast]* | |
| 7 | Ahm.. and she's presenting with a large mass in the lateral portion of the breast.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Quantity [a]*<br>*Size [large]*<br>*Region [lateral portion]*<br>*BodyLocation [left breast]* | |

| | |
|---|---|
| 8 | But it's very unusual that density here that I'm seeing<br>*Meta-Reasoning ⇒ Experiential-Memory*<br>*Radiological Observation #5 [density]*<br>*Clinical Appearance [presence of unusual density]* |
| 9 | which I don't see in the right breast.<br>*Data Comparison ⇒ Compare Contralateral*<br>*Radiological Observation #5 [density]*<br>*Appearance [ ]*<br>*BodyLocation [right breast]* |
| 10 | I guess again, what we could do is do an ultrasound.<br>*Diagnostic Plan ⇒ Subsequent Examination*<br>*Recommendation #1 [perform an ultrasound]* |
| 11 | This could represent a fibroadenoma or again just a cyst.<br>*Hypothesis Generation ⇒ Trigger*<br>*Radiological Finding #1 [mass]*<br>COND:<br>*Certainty [could represent]*<br>*Classification [fibroadenoma] OR-EXCLUSIVE [cyst]* |
| 12 | This patient could have just a cystic breast<br>*Hypothesis Generation ⇒ Further Specification*<br>*Radiological Finding #1 [mass]*<br>AND<br>*Radiological Observation #1 [breast tissue density]*<br>AND<br>*Radiological Observation #3 [nodular breast]*<br>AND<br>*Radiological Observation #4 [arch-like skin thickening]*<br>AND<br>*Radiological Observation #5 [density]*<br>COND:<br>*Classification [cystic breast]* |
| 13 | but we could do an ultrasound. I think that would be reasonable, and I would like to know clinically what that area here corresponds to.<br>*Diagnostic Plan ⇒ Subsequent Radiological Examination*<br>ORDER<br>*Recommendation #1 [perform an ultrasound]*<br>AND<br>*Recommendation #2 [clinical examination of the lateral portion of the left breast]* |
| 14 | There is asymmetry in the size of the breast and I would like to know what that looks like clinically.<br>*Diagnostic Plan ⇒ Subsequent Examination*<br>*Recommendation #2 [clinical examination of the lateral portion of the left breast]* |
| 15 | I would just say mass in the lateral portion of the left breast.<br>*Hypothesis Generation ⇒ Generalization*<br>*Radiological Finding #1 [mass]*<br>*Region [lateral portion]*<br>*BodyLocation [left breast]* |
| 16 | But an ultrasound would reveal this I presume. That's it.<br>*Diagnostic Plan ⇒ Subsequent Radiological Examination*<br>*Recommendation #1 [perform an ultrasound]* |

| **S9 - 7B** | **staff radiologist** | |
|---|---|---|
| **Seg. #** | **Transcript** | **Image** |
| 1 | Okay. Second case. Elderly patient. Came in for routine examination. There was no abnormality on the physical examination. No previous mammograms<br>*Data Acquisition ⇒ Reads Clinical History* |  |
| 2 | and we're provided ah with four films which I've really not set in order. There're all of the right breast and I'm going to presume that they're all taken on the same date.<br>*Data Identification ⇒ Identifies Mammograms as Single Study* | |
| 3 | It's kind of hard to read these........<br>*Data Assessment ⇒ Comments on the Quality of the Mammograms* | |
| 4 | We have a cluster of microcalcifications which overall have a benign appearance<br>*Data Classification ⇒ Identifies & Classifies Mammographic Cue*<br>*Radiological Finding #1 [mircocalcifications]*<br>*Distribution [clustered]*<br>*Classification [benign appearance]* | |
| 5 | as seen within the density with ah at least partially well-defined boarders<br>*Data Explanation ⇒ Characterizes Associated Mammographic Cue*<br>*Radiological Finding #2 [density]*<br>*Degree [partially]*<br>*margin [well-defined]*<br>ASSOC:<br>*Radiological Finding #1 [mircocalcifications]* | |
| 6 | And with magnification, as usual, we see the microcalcification better and in this particular case they still remain fairly benign looking.<br>*Hypothesis Generation ⇒ Trigger*<br>*Degree [fairly]*<br>*Classification [benign]* | |
| 7 | It's not always the case and without the ah previous mammograms or ah.. films of the left breast, ah I would think that these are ah benign calcifications probably within ah a fibroadenoma.<br>*Hypothesis Generation ⇒ Further Specification*<br>*Radiological Finding #1 [mircocalcifications]*<br>AND<br>*Radiological Finding #2 [density]*<br>*Certainty [probably]*<br>*Classification [benign calcifications with a fibroadenoma]* | |

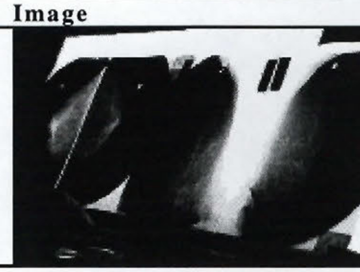
| | |
|---|---|
| 8 | And although I cannot entirely exclude malignancy, I don't think you can ever really entirely exclude malignancy,<br>*Hypothesis Generation ⇒ Generalization*<br>*Radiological Finding #1 [mircocalcifications]*<br>AND<br>*Radiological Finding #2 [density]*<br>*Classification [benign calcifications with a fibroadenoma]*<br>OR-ALTERNATIVE<br>*Classification [malignancy]* |
| 9 | I would ah follow this up in ah 6 months ah and if it is unchanged then I would routinely send the patient back to screening examinations every year.<br>Diagnostic Plan ⇒ *Subsequent Examinations*<br>*Radiological Finding #1 [mircocalcifications]*<br>AND<br>*Radiological Finding #2 [density]*<br>*Recommendation [follow-up in 6 months]*<br>*IF [condition changed] THEN [send patient for further screening mammograms on a yearly basis]* |
| 10 | Just before we conclude, I gonna have a look at the remainder of the breast to make sure I'm not missing any other potentially significant lesions<br>GOAL [inspect remainder of the breast] |
| 11 | and ah I don't think so. So that's that.<br>*Data Exploration ⇒ Notes-Absent-Data*<br>[no significant lesions in the right breast] |

| S9 - 8B | staff radiologist | |
|---|---|---|
| **Seg. #** | **Transcript** | **Image** |
| 1 | Okay. Next case. A woman in her early 50s who came with a palpable lesion. A few years previously she had breast surgery on the right but the result of surgery are not known.<br>*Data Acquisition ⇒ Reads Clinical History* |  |
| 2 | Well, we note asymmetric density seen in the left upper ah and medial portions I suppose of the ah breast, which represents fibroglandular tissue.<br>*Data Classification ⇒ Identifies & Classifies*<br>*Mammographic Cue*<br>*Radiological Observation #1 [asymmetric density]*<br>*Region [upper left] AND [medical portions]*<br>COND:<br>*Classification [fibroglandular tissue]* | |
| 3 | On the ah right side the patient has had surgery so there's less density<br>*Data Explanation ⇒ Infer History*<br>*Radiological Observation #2 [density]*<br>*Region [right side]*<br>*Change [less]*<br>*COND:*<br>*[surgery]* | |
| 4 | there are areas of architectural distortion.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #3 [architectural distortion]*<br>*Distribution [areas]* | |
| 5 | So we might be dealing with some post-operative changes, fat necrosis or something of that sort<br>*Hypothesis Generation ⇒ Trigger*<br>*Radiological Observation #1 [asymmetric density]*<br>AND<br>*Radiological Observation #2 [density]*<br>AND<br>*Radiological Observation #3 [architectural distortion]*<br>COND:<br>*Certainty [might be dealing with]*<br>*Classification [post-operative changes]*<br>OR-ALTERNATIVE<br>*Classification fat necrosis]*<br>OR-ALTERNATIVE<br>*Classification [ ]* | |
| 6 | but the area that somebody has highlighted for us demonstrates a cluster of ah microcalcifications<br>*Data Examination ⇒ Identifies Marked Mammographic Cue*<br>*Radiological Finding #1 [microcalcifications]*<br>*Number [a]*<br>*Distribution [cluster]* | |

| | |
|---|---|
| 7 | which ah overall I think must be characterized as indeterminate because some of them are small, too small to be characterized, whereas others have pointy edges.<br>*Hypothesis Generation ⇒ Trigger*<br>    *Radiological Finding #1 [microcalcifications]*<br>      *Classification [indeterminant]*<br>COND:<br>  INDEFNITE<br>      *Size [small]*<br>      *Margins [point edges]* |
| 8 | This may or may not be logistically feasible because in your practice and how your practice is set up, whether you read on the spot or after the fact. If you read in after the fact, obviously you don't have the luxury of bringing back all the patients that have microcalcifications, ah but in this particular case even if the patient has left I would bring the patient back and dictate the report such a way as to relieve some anxiety and ah say something to the effect that a complete mammographic examination would include ah bringing the patient back for additional view.<br>*Diagnostic Plan ⇒ Subsequent Radiological Examination*<br>    *Radiological Finding #1 [microcalcifications]*<br>*Recommendation [compression magnification views]* |
| 9 | So overall I cannot tell whether these changes are malignant or not, particularly since somebody was worried enough to have taken some tissue from this patient.<br>*Hypothesis Generation ⇒ Generalization*<br>    *Radiological Finding #1 [microcalcifications]*<br>      *Classification [malignant]*<br>      *OR-ALTERNATIVE*<br>*Classification [not malignant]* |
| 10 | We have to give them some benefit of the doubt ah and follow-up on this with an additional view.<br>*Summarization ⇒ Repeat Data*<br>    *Radiological Finding #1 [microcalcifications]*<br>      *Recommendation [compression magnification views]* |

| S6 - 9A | staff radiologist | |
|---|---|---|
| **Seg. #** | **Transcript** | **Image** |
| 1 | Next is a 68 year-old woman who has palpable abnormality in the lateral portion of the left breast, upper outer quadrant.<br>*Data Acquisition ⇒ Reads Clinical History* |  |
| 2 | Mammography of the left breast reveals a spiculated mass in the upper outer quadrant.<br>*Data Examination ⇒ Identifies Mammoghraphic Cue*<br>*Radiological Finding #1 [mass]*<br>*Number [a]*<br>*Margins [spiculated]*<br>*Region [upper outer quadrant]*<br>*BodyLocation [left breast]* | |
| 3 | Very suspicious of ah malignant lesion,<br>*Hypothesis Generation ⇒ Trigger*<br>*Radiological Finding #1 [mass]*<br>*Classification [very suspicious malignant lesion]* | |
| 4 | the spiculated nature of it underlines microcalcifications as well.<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Margins [spiculated]*<br>ASSOC:<br>*Radiological Observation #1 [microcalcifications]*<br>*Quantity [more than one- microcalcifications]* | |
| 5 | We have better exposed film in the magnified view which confirms these findings.<br>*Data Comparison ⇒ Compares Mammographic Cue on Additional Views*<br>*Radiological Finding #1 [mass]*<br>AND<br>*Radiological Observation #1 [microcalcifications]*<br>[present on magnified view] | |
| 6 | The lesion is very irregular in outline with fine strands projecting into the adjacent fatty tissue.<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Degree [very]*<br>*Margins [irregular outline with fine strands projecting into adjacent fatty tissue]* | |
| 7 | It measures about 12 mm.<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Finding #1 [mass]*<br>*Certainty [about]*<br>*Size [12 mm]* | |

| | |
|---|---|
| 8 | The microcalcification is not well seen on the magnified view.<br>*Data Comparison ⇒ Compares Mammographic Cue on Additional Views*<br>*Radiological Observation #1 [microcalcifications]*<br>*[not well seen on the the magnified view]* |
| 9 | however even without calcifications is still consistent with malignancy<br>*Hypothesis Evaluation ⇒ Confirmation*<br>*Radiological Finding #1 [mass]*<br>NEG: COND<br>*Radiological Observation #1 [microcalcifications]*<br>*Classification [malignancy]* |
| 10 | so it has to be removed.<br>*Diagnostic Plan ⇒ Subsequent Examination*<br>*Radiological Finding #1 [mass]*<br>*Recommendation #1 [perform biopsy]* |

| S7 - 10B | staff radiologist | |
|---|---|---|
| **Seg. #** | **Transcript** | **Image** |
| 1 | 60 year-old woman who presented with inflammation of her left breast for two weeks.<br>*Data Acquisition ⇒ Reads Clinical History* |  |
| 2 | I have bilateral mammogram available.<br>*Data Identification ⇒ Identifies Mammograms as Single Study* | |
| 3 | There is small moderate amount of glandular tissue<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #1 [glandular breast tissue]*<br>*Size [small]*<br>*Degree [moderate amount]* | |
| 4 | and there appears to be asymmetry in the density of the breast tissues visible on the mediolateral oblique view as well as on the cranial caudal view.<br>*Data Comparison ⇒ Compares Contralateral*<br>*Radiological Observation #2 [breast tissue asymmetry]*<br>*Certainty there appears]*<br>*BodyLocation [left breast]*<br>AND<br>*Radiological Observation #3 [breast tissue asymmetry]*<br>*Certainty there appears]*<br>*BodyLocation [right breast]* | |
| 5 | There is a diffusely increased haziness.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Observation #4 [density]*<br>*Distribution [diffusely]*<br>*Change [increased]* | |
| 6 | There is skin thickening noted especially around the aereolar region.<br>*Data Examination ⇒ Identifies Mammographic Cue*<br>*Radiological Finding #1 [skin thickening]*<br>*Region [aereolar region]* | |
| 7 | I don't see a definite mass lesion or suspicious microcalcification.<br>*Data Exploration ⇒ Notes-Absent-Data*<br>*[definitive mass lesion]*<br>OR-EXCLUSIVE<br>*[suspicious microcalcifications]* | |

| | |
|---|---|
| 8 | In conclusion, these changes could be compatible with Paget's disease or a nonspecific mastitis.<br>*Hypothesis Generation ⇒ Trigger*<br>*Radiological Observation #1 [glandular breast tissue]*<br>AND<br>*Radiological Observation #2 [breast tissue asymmetry]*<br>AND<br>*Radiological Observation #3 [breast tissue asymmetry]*<br>AND<br>*Radiological Observation #4 [density]*<br>AND<br>*Radiological Finding #1 [skin thickening]*<br>COND:<br>Classification [Paget's disease]<br>OR-ALTERNATIVE<br>Classification [non-specific mastitis] |

| S13 - 10B | radiology resident | |
|---|---|---|
| **Seg. #** | **Transcript** | **Image** |
| 1 | A 60 year old woman who presented with inflammation of the left breast for two weeks.<br>*Data Acquisition ⇒ Reads Clinical History* |  |
| 2 | So we have a mediolateral oblique view and a cranial caudal view of both breasts.<br>*Data Identification ⇒ Identifies Mammograms as Single Study* | |
| 3 | Ah, the abnormalities have already been marked up on the left breast.<br>*Data Examination ⇒ Identifies Marked Mammographic Cue*<br>*Radiological Finding #1 [skin thickening]*<br>*BodyLocation [left breast]*<br>AND<br>*Radiological Observation #1 [lesion]*<br>*BodyLocation [left breast]*<br>AND<br>*Radiological Observation #2 [calcification]*<br>*BodyLocation [left breast]* | |
| 4 | In the right breast, I see a benign looking calcification<br>*Data Classification ⇒ Identifies & Classifies Mammographic Cue*<br>*Radiological Observation #3 [calcification]*<br>*Number [a]*<br>*BodyLocation [right breast]*<br>*Classification [benign-looking]* | |
| 5 | but otherwise I don't see anything suspicious there. On the right.<br>*Data Exploration ⇒ Notes-Absent-Data*<br>*[noting suspicious]*<br>*BodyLocation [right breast]* | |
| 6 | On the left, there is thickening of the skin ahm in the central portion here which is probably inferior central and also on the superior aspect.<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Finding #1 [skin thickening]*<br>*Region [central portion] AND [inferior lateral] AND [superior aspect]*<br>*BodyLocation [left breast]* | |
| 7 | There is asymmetry in density of the breast here we see in the retroareolar area,<br>*Data Exploration ⇒ Characterizes Mammographic Cue*<br>*Radiological Finding #2 [breast tissue density]*<br>*Region [retroaereolar region]*<br>*BodyLocation [left breast]* | |

| | |
|---|---|
| 8 | seen on the cranial caudal view abnormally dense lesion with some spiculation to it<br>*Data Comparison ⇒ Compare Ipsilateral*<br>    *Radiological Finding #2 [breast tissue density]*<br>      *Degree [abnormally dense]*<br>      *Description [spiculated]* |
| 9 | and we see a small punctate calcification in the center.<br>*Data Exploration ⇒ Characterizes Associated Mammographic Cue*<br>    *Radiological Finding #2 [breast tissue density]*<br>      *Degree [abnormally dense]*<br>      *Description [spiculated]*<br>      *ASSOC:*<br>    *Radiological Finding #3 [calcification]*<br>      *Number [a]*<br>      *Size [small]*<br>      *Description [punctate]* |
| 10 | Now in the clinical context here, we're talking about signs of inflammation in the left breast. The first thing we need to rule out in a 60 year old woman like this is inflammatory breast cancer, ahm especially with the skin thickening that we see here is worrisome.<br>*Hypothesis Generation ⇒ Trigger*<br>    *Clinical History [age, inflammation left breast]*<br>AND<br>    *Radiological Finding #1 [skin thickening]*<br>COND:<br>*Classification [inflammatory breast cancer]* |
| 11 | I still think this needs to be followed-up or very closely both clinically or by mammography. I'm afraid this would have to be biopsied as well.<br>*Diagnostic Plan ⇒ Subsequent Examination*<br>    *Recommendation #1 [follow-up clinically]*<br>*AND*<br>    *Recommendation #2 [follow-up mammographically]*<br>*AND*<br>    *Recommendation #3 [biopsy]* |

# Appendix P

## A Description of All Diagnoses and Recommendation for the Ten Breast Diseases by Case Number, Experimental Condition, Level of Expertise & Participant Number

## Case 1A

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 1 | suspicious, incompletely well-circumscribed mass | • follow-up by (1) comparing to previous films, and (2) ultrasound (if cyst then stop) (if not solid conduct fine needle aspiration under ultrasound) |
| S 3 | probably not a benign mass | • biopsy before excision |
| S 5 | could represent a fibroadenoma, cyst or carcinoma | • ultrasound to confirm if it is a cyst (if not a cyst follow the patient or have a stereotaxic biopsy) |
| S 7 | suspicious although it could be benign (cyst) | • ultrasound to exclude the presence of a large cyst but it's solid lesion then removal is advised |
| S 9 | suspicious in view of its size (this is controversial) | • ultrasound to verify it's a cyst or solid mass and obtain fluid and/or tissue for diagnosis |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 11 | carcinoma | • needle biopsy |
| S 13 | cystosarcoma philloides or cyst | • ultrasound (if solid then excisional biopsy) |
| S 15 | cyst, philloides, giant fibroadenoma or fibrous tumor | • if solid on ultrasound then excisional biopsy |
| S 17 | suspicious for malignancy | • spot films then biopsied or excised |
| S 19 | cancer | • ultrasound and (if not completely cystic then biopsy) |

## Case 1B

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 2 | large cyst, giant papilloma or carcinoma | • ultrasound followed by biopsy and aspiration |
| S 4 | simple cyst or carcninoma | • ultrasound followed by biopsy |
| S 6 | cyst, cystosarcoma philloides or medullary carcinoma | • ultrasound followed by biopsy |
| S 8 | giant fibroadenoma or medullary carcinoma | • biopsy |
| S 10 | suspicious mass | • ultrasound and since it's solid then do surgical biopsy under ultrasound control |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 12 | malignant process or cystic lesion | • ultrasound (if solid then bypass biopsy and send patient to surgery) |
| S 14 | medullary carcinoma, cystosarcoma philloides, cyst or hamartoma | • ultrasound (if either solid or cystic) would affect excision |
| S 16 | suspicious mass, could be benign cyst | • ultrasound (if cystic then stop) (if solid then biopsy) |
| S 18 | could be cystic or malignant | • could be biopsied |
| S 20 | philloides tumor or large abscess or large fungating carcinoma | • surgery (with needle localization if necessary |

## Case 2A

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| **S 1** | probably cysts on the left and well-circumscribed nodules on the right | • ultrasound and comparison to previous (retrospective follow-up, if available) |
| **S 3** | benign or malignant (but more likely to be benign) | • ultrasound the 2 nodules and do compression view of the increased density area |
| **S 5** | benign process could represent a cyst or fibroadenoma (probably benign) | • ultrasound (if not cystic do stereotaxic biopsy) |
| **S 7** | two more than 1 cm well-defined nodular lesions in the left retroareolar region and a group of benign nodular regions in the right breast | • ultrasound to confirm if solid or cystic<br>• follow-up in 6 months |
| **S 9** | probably benign such as a small cyst or fibroadenoma | • ultrasound the left breast and if any are greater than 1 cm conduct follow-up |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| **S 11** | 2 mass lesions in the left breast probably benign and cysts | • ultrasound to confirm if they are cysts and biopsy the right breast |
| **S 13** | probably benign cysts | • ultrasound the two nodules |
| **S 15** | probably cysts | • ultrasound reveal cysts then leave them and have routine be follow-up in 1 year (if not perfectly cystic then biopsy) |
| **S 17** | probably cysts | • ultrasound to prove they are cysts.<br>(if they are not cysts then biopsy or aspirate them) |
| **S 19** | multiple masses are benign and other masses are indeterminate and probably cysts | • do magnified compression view to see if margins are spiculated.<br>(if smooth then ultrasound to see if solid or cystic)<br>(if spiculated on mag view or solid on ultrasound then biopsy) |

## Case 2B

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| **S 2** | probably benign cysts or fibroadenomas | • confirm with ultrasound |
| **S 4** | benign-looking lesions | • ultrasound (if cystic then stop) (if solid then biopsy) |
| **S 6** | two probably benign cysts in the left breast | • ultrasound |
| **S 8** | probably benign cysts | • ultrasound |
| **S 10** | probably cysts | • ultrasound (if cystic then stop) (if solid then do needle biopsy) |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| **S 12** | probably benign cysts | • ultrasound to confirm they're benign |
| **S 14** | probably cysts | • ultrasound (if cyst then stop examination)<br>(if solid then biopsy and remove) |
| **S 16** | (could be) probably cysts | • ultrasound (if cyst then stop) (if solid then biopsy) |
| **S 18** | probably cysts | • ultrasound (if cyst then stop) (if solid then local biopsy and aspirate) |
| **S 20** | probably cystic lesions | • ultrasound (if solid then excised) (if cystic then follow and make sure it doesn't have a solid component such as a papillary carcinoma) |

## Case 3A

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 1 | at least mildly suspicious for a carcinoma | • ultrasound and follow-up with biopsy |
| S 3 | suspicious for malignancy | • excisional biopsy |
| S 5 | neoplastic lesion | • stereotaxic biopsy |
| S 7 | malignant lesion | • ultrasound examination followed by biopsy |
| S 9 | very suspicious for carcinoma on the left | • compression spot views followed by biopsy |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 11 | suspicious for malignancy lesion | • needle biopsy |
| S 13 | suspicious lesion with spiculation | • perform true lateral or compression view to better view the lesion |
| S 15 | worrisome and bothersome lesion | • surgical excision |
| S 17 | moderately suspicious for malignancy | • biopsy |
| S 19 | lesion has moderate chance of being malignant (e.g., sclerosing adenosis) | • biopsy |

## Case 3B

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 2 | | • ultrasound |
| S 4 | carcinoma | • exaggerated lateral and stereotaxic biopsy |
| S 6 | carcinoma | • biopsy (upper outer quadrant of the left breast) |
| S 8 | very suspicious for malignancy | • biopsy |
| S 10 | very suspicious lesion | • biopsy |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 12 | very suspicious for a malignancy | • biopsy |
| S 14 | benign process (e.g., sclerosing adenosis) | • biopsy after getting previous mammograms, ultrasound report, history of trauma in the area, history of nodule |
| S 16 | suspicious for malignancy | • biopsy |
| S 18 | malignant | • excisional biopsy |
| S 20 | highly suspicious for malignancy | • needle localization with surgical excision |

## Case 4A

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 1 | most likely benign | • follow-up with mammogram |
| S 3 | milk of calcium in small cysts | • ultrasound of other nodules |
| S 5 | cystic lesion of the breast (milk of calcium) | • ultrasound (if not cystic then stereotaxic biopsy) (if cystic then stop) |
| S 7 | benign type of microcalcifications (cup-like) | • ultrasound |
| S 9 | calcification due to dysplasia | • screening mammogram and follow-up in 6 months for the other nodules |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 11 | suspicious for malignancy lesion | • perform CC view with spot compression and true lateral and MLO view to further localize lesion and delineate margins of mass lesion |
| S 13 | simple cyst and probably pleomorphic microcalcifications | • magnification views and ultrasound the round lesion |
| S 15 | simple cyst | • ultrasound<br>(if solid then biopsy)<br>(if cyst then stop) |
| S 17 | milk of calcium in cysts | • ultrasound |
| S 19 | milk of calcium in small cysts | • magnified view to assess margins (speculated or not) or if not helpful then ultrasound (cystic or solid)<br>(if solid then biopsy) |

## Case 4B

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 2 | benign calcifications | • follow-up mammogram for the other nodules |
| S 4 | milk of calcium in small cysts | • |
| S 6 | benign calcifications but suspect malignancy | • repeat examinations with compression views and ultrasound the mass in the right breast which is benign |
| S 8 | milk of calcium | • thorough clinical examination |
| S 10 | milk of calcium | • get a good true lateral film |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 12 | milk of calcium cysts | • none |
| S 14 | milk of calcium cysts | • none |
| S 16 | milk of calcium in small cysts | • follow-up in 1 year and if patient present then do compression view or ultrasound to verify if solid or cystic<br>(if cystic then stop) |
| S 18 | benign calcifications in ducts | • none |
| S 20 | very suspicious calcification in ducts (early DCIS) | • ultrasound determine if cystic or solid<br>(if solid then biopsy to exclude carcinoma) |

### Case 5A

#### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 1 | fibroadenolipoma | • probably needs to come out |
| S 3 | fibroadenolipoma | • remove for cosmetic reasons |
| S 5 | hamartoma | • biopsy |
| S 7 | hamartoma | • |
| S 9 | "I don't know what this is" | • put a needle into 1 of these nodules and get some histology or pathology for them |

#### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 11 | hamartoma | • none |
| S 13 | fibroadenoma or cyst | • ultrasound |
| S 15 | hamartoma or fibroadenoma | • none |
| S 17 | hamartoma of the breast | • none |
| S 19 | hamartoma | • none |

### Case 5B

#### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 2 | fatty tumor | • consult with other radiologists |
| S 4 | large benign lesion | • removed for cosmetic reasons |
| S 6 | cytosarcoma | • none |
| S 8 | fibroadenolipoma | • none |
| S 10 | hamartoma | • leave mass if it's not bothering the patient |

#### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 12 | hamartoma of the breast | • none |
| S 14 | giant hamartoma of the left breast | • none |
| S 16 | hamartoma | • none |
| S 18 | breast hamartoma | • none |
| S 20 | fat necrosis with multiple regions | • find out if she has history of trauma and chronicity of event (if no trauma then ultrasound and consider giant fibroadenoma or walled-off abscess) |

## Case 6A

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 2 | cancer or hamartoma | • needle biopsy |
| S 4 | carcinoma, cystosarcoma or lymphomas | • needle biopsy |
| S 6 | suspicious lesion in the upper outer quadrant | • ultrasound and then biopsy |
| S 8 | malignant | • biopsy |
| S 10 | suspicious mass | • ultrasound (if cystic then stop) (if not cystic then biopsy under ultrasound or open biopsy) |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 12 | malignant process | • biopsy |
| S 14 | carcinoma | • get old mammograms, history of trauma, and conduct stereotaxic biopsy |
| S 16 | suspicious for carcinoma | • ultrasound and if one can't see it then conduct stereotaxic biopsy |
| S 18 | suspicious for carcinoma | • ultrasound to determine if cystic or not (if solid biopsy) |
| S 20 | carcinoma or DCIS | • biopsy |

## Case 6B

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 1 | suspicious (enough to warrant further examination) | • ultrasound to see if cystic (if not cystic then biopsy) |
| S 3 | suspicious for carcinoma | • excisional biopsy |
| S 5 | malignant | • Cleopatra projection and stereotaxic breast biopsy |
| S 7 | asymmetrical nodular lesion present in the axiallary tail of the right breast | • further investigation including compression spot films in magnification views and probably ultrasound |
| S 9 | suspicious for malignancy | • needs further work-up such as a compression spot view to see the tissue |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 11 | strongly suspicious lesion | • biopsy and additional views |
| S 13 | very suspicious | • do a compression view of the area (biopsy) |
| S 15 | carcinoma | • core biopsy or direct excision |
| S 17 | highly suspicious category | • do spot magnification films and then biopsy |
| S 19 | carcinoma or asymmetrical tissue | • do spot compression to see if tissue spreads out and becomes a mass (biopsy) |

### Case 7A

#### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 2 | very fine microcalcifications | • biopsy |
| S 4 | somewhat suspicious | • pre-op needle localization before surgery |
| S 6 | suspicious for malignancy | • ultrasound to outline nature of the region followed by excision |
| S 8 | malignant microcalcifications | • has to be removed |
| S 10 | very high suspicion | • stereotaxic or open biopsy |

#### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 12 | malignant process | • biopsy the lesion |
| S 14 | malignant category | • biopsy under needle localization |
| S 16 | suspicious for malignancy process | • go to surgery |
| S 18 | malignant | • needle localization for excisional biopsy |
| S 20 | carcinoma | • either perform a fine needle aspiration or do a core biopsy<br>also removal of axillary node (with open biopsy) |

### Case 7B

#### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 1 | ductal carcinoma in situ | • biopsy |
| S 3 | suspicious but not very specific microcalcifications | • excise due to their focal nature |
| S 5 | considered neoplastic | • stereotaxic or surgical biopsy |
| S 7 | (lesion is) indeterminate slightly suspicious but could be fibroadenoma | • surgical consultation or biopsy |
| S 9 | benign calcifications (probably within a fibroadenoma), however, cannot entirely exclude malignancy | • follow-up in 6 months |

#### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 11 | quite suspicious | • excisional biopsy with needle localization prior to surgical excision because it's not a palpable lesion |
| S 13 | lesion with microcalcifications within | • biopsy |
| S 15 | carcinoma | • biopsy or excision (depending on surgeon) |
| S 17 | invasive type of ductal carcinoma | • excisional biopsy |
| S 19 | ductal carcinoma | • biopsy |

## Case 8A

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| **S 2** | few indeterminate calcifications | • compare to previous films |
| **S 4** | recurring malignant microcalcifications or carcinoma | • excision of microcalcification cluster |
| **S 6** | carcinoma | • core biopsy in the area of the microcalcifications with a 16 or 14 gauge needle |
| **S 8** | indeterminate microcalcifications | • follow-up |
| **S 10** | very suspicious | • biopsy |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| **S 12** | malignant process | • biopsy |
| **S 14** | malignant | • stereotaxic needle biopsy |
| **S 16** | malignancy cannot be excluded | • stereotaxic biopsy |
| **S 18** | suspicious for malignant lesion area | • needle localization for excisional afterwards |
| **S 20** | recurring carcinoma or residual carcinoma (that did not resolve) | • biopsy |

## Case 8B

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| **S 1** | (can't rule out) ductal carcinoma in situ or a low-grade malignancy (in the upper outer quadrant) | • biopsy |
| **S 3** | indeterminate, it could be malignant | • excisional biopsy or biopsy with a large core needle |
| **S 5** | suspicious for malignancy | • surgical biopsy |
| **S 7** | suspicious for neoplastic disease | • supplemental studies |
| **S 9** | cannot tell whether these changes are malignant or not | • follow-up with additional views |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| **S 11** | suspicious for carcinoma | • stereotaxic biopsy |
| **S 13** | pleomorphic calcifications that are worrisome | • biopsy |
| **S 15** | calcifications in the right breast are probably benign | • stereotaxic biopsy |
| **S 17** | highly suspicious for malignancy | • compare to previous, magnified spots films of area of calcification and then biopsy |
| **S 19** | cancer (probably ductal carcinoma in situ) | • excisional biopsy |

## Case 9A

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 2 | cancer | • biopsy |
| S 4 | 2 carcinomas | • biopsy both |
| S 6 | malignant | • excision |
| S 8 | very suspicious | • biopsy |
| S 10 | 2 lesions are very very high suspicion for malignancy | • biopsy both lesions |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 12 | suggestive of malignancy | • biopsy 1 lesion |
| S 14 | ominous looking lesion | • biopsy |
| S 16 | very high chance of being a malignancy | • stereotaxic biopsy |
| S 18 | suspicious for malignancy | • excisional biopsy |
| S 20 | carcinoma | • needle localization, surgical excision with axillary node resection for staging |

## Case 9B

### STAFF RADIOLOGISTS

| | | |
|---|---|---|
| S 1 | suspicious for carcinoma | • biopsy both lesions |
| S 3 | one lesion that is definitively suspicious | • excision and investigate second lesion with compression views |
| S 5 | neoplasm of the upper pole of the left breast | • stereotaxic biopsy or surgical biopsy |
| S 7 | one lesion suspicious for malignancy, possible multi-focal carcinoma | • follow second lesion with magnification views |
| S 9 | 2 carcinomas | • excision |

### RADIOLOGY RESIDENTS

| | | |
|---|---|---|
| S 11 | both lesions are highly suspicious for carcinoma | • Cleopatra view and excision |
| S 13 | both lesions are highly suspicious for carcinoma | • both lesions need to be biopsied |
| S 15 | both lesions are suspicious for carcinoma | • need tissue diagnosis (either biopsy or excision) |
| S 17 | one lesion is highly suspicious<br>second lesion not clear on all views | • first lesion - biopsy<br>second lesion - do a cone down mag view of both MLO and CC |
| S 19 | multi-centric ductal carcinoma | • biopsy |

**Case 10A**

**STAFF RADIOLOGISTS**

| | | |
|---|---|---|
| S 2 | mastitis or inflammatory carcinoma | • |
| S 4 | acute cellulitis or inflammatory carcinoma | • biopsy |
| S 6 | inflammatory carcinoma | • |
| S 8 | carcinoma or mastitis | • course of antibiotics and follow-up clinical exam |
| S 10 | inflammatory carcinoma | • antibiotics, repeat films then perform broad biopsy in the retroareolar area |

**RADIOLOGY RESIDENTS**

| | | |
|---|---|---|
| S 12 | mastitis or inflammatory carcinoma | • antibiotic treatment and biopsy left breast |
| S 14 | nothing that looks malignant | • compare to previous, examine patient, follow-up in 3-4 months |
| S 16 | inflammatory carcinoma | • skin biopsy |
| S 18 | inflammatory carcinoma or acute inflammation | • treat clinically and perform biopsy |
| S 20 | inflammatory breast carcinoma | • antibiotic treatment need more clinical history (infection, redness, pustular or not) following results of antibiotic treatment then biopsy or consider further imaging |

**Case 10B**

**STAFF RADIOLOGISTS**

| | | |
|---|---|---|
| S 1 | does not look specific for carcinoma | • need more clinical history (history of radiation, does she have mastitis, does she appear clinically sick) and do spot compression of the region |
| S 3 | can't complete rule out carcinoma either infectious inflammatory process or neoplastic inflammatory process | • need more clinical examination and follow-up treatment |
| S 5 | Paget's disease, inflammatory carcinoma or infectious cellulitis | • surgical biopsy to rule out possibility of neoplasm |
| S 7 | Paget's disease or non-specific mastitis | • |
| S 9 | cellulitis or mastitis | • ultrasound |

**RADIOLOGY RESIDENTS**

| | | |
|---|---|---|
| S 11 | infectious mastitis or inflammatory carcinoma | • breast biopsy |
| S 13 | inflammatory breast cancer | • follow-up clinically and mammographically and biopsy |
| S 15 | mastitis or inflammatory breast cancer | • find out more patient history (fever, tenderness, heat) |
| S 17 | inflammatory breast cancer | • ultrasound and mastectomy |
| S 19 | inflammation due to infection or carcinoma | • lesion may require biopsy |

# Appendix Q

## Ethics Form

# MCGILL UNIVERSITY
# FACULTY OF EDUCATION

## CERTIFICATE OF ETHICAL ACCEPTABILITY FOR RESEARCH INVOLVING HUMAN SUBJECTS

A review committee consisting of **three** of the following members:

| | |
|---|---|
| 1. Prof. J. Derevensky | 1. Prof. M. Maguire |
| 2. Prof. R. Ghosh | 2. Prof. N. Jackson |
| 3. Prof. M. Downey | 3. Prof. H. Perrault |

has examined the application for certification of the ethical acceptability of the project entitled:

**The Study of Radiological Expertise and its Implications for the Design of Computer-Based Radiology Learning Environment**

as proposed by: **Roger Azevedo**

Applicant's Name **Roger Azevedo** Supervisor's Name **Dr. S.P. Lajoie**

Applicant's Signature Roger Azevedo Supervisor's Signature Suzanne P. Lajoie

Degree Program **Ph.D. Educational Psychology**

Granting Agency **FCAR: Fonds pour la Formation de Chercheurs et l'Aide a la Recherche**

The review committee considers the research procedures as explained by the applicant in this application, to be acceptable on ethical grounds.

(Signatures)

a) Dr. J. Derevensky Date June 7/95

b) Ratna Ghosh Date June 12, '95

c) [illegible] Date 13 June 95

Associate Dean (Academic) [illegible] Date [illegible]

March 1995

# Appendix R

## Conference Paper & Book Chapter Based on the Dissertation

# Radiological Expertise and the Effects of Perceptual Scaffolding on the Diagnosis of Mammograms

**Roger Azevedo** and **Susanne P. Lajoie**
McGill University
Applied Cognitive Science Research Group
3700 McTavish Street, B199
Montreal, Quebec, H3A 1Y2, Canada
{cxaz,cysl}@musica.mcgill.ca

**Monique Desaulniers** and **David M. Fleiszer**
Montreal General Hospital & McGill University
1650 Cedar Avenue
Montreal, Quebec, H3G 1A4, Canada
dfleisze@is.mgh.mcgill.ca

## Introduction

This study investigated the differences in the underlying cognitive processes used by staff radiologists and radiology residents while diagnosing breast diseases as depicted on mammograms. The research: (1) characterized the cognitive processes of both novice and expert radiologists by conducting in-depth analyses of verbal protocols, (2) incorporated an *augmented* experimental condition (to test the hypothesis that highlighted critical findings would facilitate the attainment of a diagnostic schema and increase diagnostic accuracy), (3) analyzed the film tracings made by both residents and radiologists and their relationship to diagnostic accuracy, and (4) rectified certain methodological and analytical inadequacies of previous cognitive studies in visual domains.

## Method

A total of twenty ($N$=20) subjects, 10 staff radiologists and 10 senior radiology residents from McGill University's five teaching hospitals participated in this study. The staff radiologists had an average of 14 years of mammography experience (diagnosed an average of 30,000 mammograms, and diagnosed an average of 66 cases per week). The residents had an average of 6 months of mammography experience (had diagnosed 25 to 1000 mammograms, and did not diagnose mammograms on a weekly basis).

Ten relatively difficult breast disease cases were selected for the study. Each case was comprised of a brief clinical history and a set of 4 mammograms. Each subject solved a total of ten breast disease cases (5 authentic and 5 augmented). In the augmented condition, the critical mammogram findings relevant to the diagnosis were highlighted. In the authentic condition, the critical mammogram findings were not highlighted. The cases were counterbalanced across conditions and subjects.

The experimental procedure involved: (1) instructing the subject to "think out loud" while he/she diagnosed each case (and solved a practice case), (2) presenting the clinical history to the subject, (3) displaying the mammogram set on a view-box and (4) instructing the subject to point to the critical film findings while diagnosing each case. Each session was recorded on audio and video.

## Results

The verbal data (in conjunction with the video data) were subjected to protocol analysis techniques (Ericsson & Simon, 1993; Patel & Groen, 1986). Expert diagnostic reasoning in visual diagnosis of mammograms was characterized by: (1) top-down and bottom-up cognitive and perceptual processes, (2) schema-based problem-solving which facilitates accurate characterization of film features, integration of clinical history cues followed by rapid and accurate diagnosis, and (3) use of forward reasoning processes during the diagnostic process. Residents' diagnostic reasoning process was characterized by: (1) mostly bottom-up reasoning involving the characterization and subsequent integration of film features, (2) use of abductive reasoning in generating diagnostic hypotheses and eliminating them based on the presence or absence of clinical cues and/or film features, and (3) inferior characterization of film features which was directly related to diagnostic inaccuracy.

Two separate repeated measures ANOVAs yielded no significant differences between the mean number of clinical findings and clinical observations by levels of expertise and experimental conditions. In addition, residents (1) provided slightly more differential diagnoses, and (2) were less accurate in their diagnoses than staff radiologists.

## Discussion

The contributions of this study include: (1) a comprehensive cognitive model of diagnostic reasoning in radiology, (2) enhanced understanding of the perceptual and cognitive processes underlying radiological diagnosis, (3) an initial theory of learning in ill-structured domains, and (4) empirical evidence for the design of a computerized system for training radiology residents to diagnose mammograms.

## References

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Patel, V. L., & Groen, G. J. (1986a). Knowledge-based solution strategies in medical reasoning. *Cognitive Science*, *10*, 911-116.

# Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society

Edited by Michael G. Shafto and Pat Langley

August 7–10, 1997
Stanford University

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1997 Mahwah, New Jersey London

*Artificial Intelligence in Education*
*B. du Boulay and R. Mizoguchi (Eds.)*
*IOS Press, 1997*

# RadTutor: The Theoretical and Empirical Basis for the Design of a Mammography Interpretation Tutor

Roger Azevedo[1], Susanne Lajoie[1], Monique Desaulniers[2], David Fleiszer[3], Patrice Bret[2]
[1]*Applied Cognitive Science Research Group, McGill University, 3700 McTavish Street, B199 Montreal, Quebec, Canada H3A 1Y2, {cxaz, cysl}@musica.mcgill.ca*
[2]*Department of Diagnistic Radiology,* [3]*Department of General Surgery, Montreal General Hospital*
*1650 Cedar Avenue, Montreal, Quebec, Canada H3G 1A4*

The purpose of this paper is to present a cognitively-based and empirically-derived approach for the design of a prototype computerized tutor to train radiology residents in diagnosing mammograms exhibiting breast diseases. A plethora of computer-based radiology training environments have recently been developed with the objective of supporting the acquisition of radiological expertise. In general, these systems have failed to reach this objective since they: (1) lack a theoretical framework incorporating the empirical evidence on medical cognition and radiological expertise, (2) fail to adopt an adequate model of instruction, and (3) consist of comparison studies (between several CBI typologies and traditional teaching) which often lack methodological rigor. This paper outlines the conceptual framework for the development of the prototype which includes: (1) a review and critique of existing computer-based radiology training environments, (2) a review of previous studies on radiological expertise, (3) a description of an empirical study aimed at attaining a cognitive model of diagnostic reasoning in mammography, (4) a description of the results of analyses of authentic radiology resident teaching rounds, (5) deriving instructional principles for the design of the mammography tutor, and (6) a description of the mammography prototype tutor.

## 1. Introduction

Radiological expertise is complex, involving several years of acquiring formalized medical knowledge as well as many years of clinical experience. It involves the integration of several distinct bodies of knowledge with separate organizing principles, including physiology, anatomy, pathophysiology, and projective geometry of radiography. Various theoretical frameworks postulate that the attainment of accurate visual diagnostic reasoning abilities involves the interaction of cognitive and perceptual factors. However, a systematic effort employing a combination of analytical methodologies and perceptual probes is required to clarify the coexisting contributions of cognitive and perceptual factors in the development of radiological expertise. In order to adequately understand the diagnostic process, a more detailed investigation is required. Specifically, a cognitive model characterizing underlying differences between radiology residents and staff radiologists should be elicited using appropriate cognitive science methodologies. The purpose of this paper is to provide a cognitively-based and empirically-derived approach for the design of an ITS for training radiology residents and radiologists in diagnosing mammograms exhibiting breast diseases. This paper will: (1) present a critique of existing computer-based radiology training environments, (2) provide an overview of the previous cognitive studies on radiological expertise, (3) report the results of an empirical study aimed at attaining a cognitive model of diagnostic reasoning in mammography interpretation, (4) delineate teaching strategies based on analyses of radiology teaching rounds, (5) propose instructional principles for the design of a mammography tutor, and (6) present a description of the computer-based mammography tutor.

## 2. Computer-Based Environments for Radiology Training

The problem of inconsistency in radiology residency training programs has recently been addressed by the widespread proliferation and dissemination of computer-based training programs. In general, these systems have failed to provide a viable solution since they: (1) lack a theoretical framework incorporating the empirical evidence on novice-expert differences, medical cognition and radiological expertise; (2) fail to adopt an adequate model of instruction; (3) consist of comparison studies (between CBI and traditional lectures or several CBI typologies) which lack both methodological and statistical rigor; (4) represent technology-driven projects employing various CBI typologies (e.g., linear CAI, branching CAI, CBT, multimedia, hypermedia, and WWW) which include hundreds of cases on a CD-ROM without a principled approach as how to best sequence the cases and present radiological images of inferior quality; (5) typically adopt a behavioral approach to teaching (e.g., present instructional content in small decontextualized units) and testing (e.g., multiple-choice with canned feedback messages) which are at odds with recent theoretical and empirical advances in cognitive science, learning and assessment in ill-structured domains; (6) fail to adopt instructional objectives stated in residency programs, and; (7) are not integrated as an integral part of the curriculum/residency training program and therefore become another source of supplementary instructional resources.

There has been a general increase in interest in the application of ITSs in the area of medical education in recent years. The rationale for building computerized tutors is based on the assumption that the user's cognitive processes can be modeled, traced, and corrected in the context of problem-solving [1]. In recent years, several ITSs have been developed for radiology training. For example, the CT Brain Tutor for training radiology residents to diagnose brain tumors from CT scans [2], and a tutor for training radiology residents to diagnose neurological MRI images [3]. The ITSs developed by these authors are based on cognitive science principles of expertise development and incorporate tutoring interventions and tutorial dialogues that are based on analyses of human interactions.

The extensive work of Sharples and colleagues [4] in developing the CT and MRI tutors is concerned with accounts of professional practice and skill development and how these issues influence the design of their tutors. They have used statistically-based principles and a structured image description language for teaching radiological image interpretation and diagnosis of cerebral diseases. Their approach to visual concept tutoring is based on grouping exemplars. The tutors provide a sequence of matched pairs of exemplar and non-exemplar images, with the non-exemplars differing from the exemplars in a minimum number of attributes at a time. This tutoring approach facilitates the novice to expert transition by assisting the residents in the progression from visual to structural schemas (facilitating rapid pattern matching) and therefore ensuring transfer of skills and learning [5]. Lastly, their tutors aim at training radiologists to view and describe images in a systematic manner.

As such, an ITS approach would offer consistency and standardization in the training of mammography interpretation. Therefore, this paper will present a cognitively-based and empirically-derived approach for the design of a computerized tutor to train radiology residents in diagnosing mammograms exhibiting breast diseases. The design is based on cognitive studies on radiological expertise, authentic analyses of radiology teaching rounds, and instructional principles.

## 3. Theoretical Framework: The Acquisition of Radiological Expertise

Radiological expertise has been investigated by numerous authors employing disparate theoretical and empirical paradigms. Three basic "paradigms" that have been applied widely are: (1) search studies which investigate eye movement patterns while experts and novices read x-ray films, (2) signal-detection studies which investigate the ability of novices and experts to detect normal and abnormal film findings, and (3) cognitive research aimed at eliciting the underlying cognitive and perceptual factors involved in radiological expertise. Relatively few studies [6,7,8] have actually investigated the underlying cognitive and perceptual factors involved in radiological diagnosis. As a result, a fundamental understanding of the constitution and acquisition of expertise in other radiological sub-specialties such as mammography has yet to be determined. The following section presents a brief overview of cognitive studies of radiological expertise and describes the major findings in chest radiography and preliminary findings of a recent study in mammogram interpretation.

### *3.1 Cognitive Studies of Radiological Expertise*

In terms of cognitive research, there have been few explicit accounts of the contributions made by radiology residents and staff radiologists during the detection/diagnostic process. Lesgold and colleagues [7,8] conducted two studies investigating the constitution and acquisition of radiological expertise in chest x-ray interpretation. Analytical techniques included perceptual probes and in-depth analyses of subjects' verbal protocols. Their research findings indicate that experts build schemas of patient anatomy, evoke pertinent schemas quickly and exhibit flexibility in tuning their schemas. Secondly, the assignment of x-ray features of normal anatomy schemata determine which features are "left over" and hence show signs of abnormality.

Lastly, normal anatomy schemata might contain attached procedures or localization rules for determining where the abnormalities reside. The expert's flexibility in tuning schemata, in the case of a dominant hypothesis and a more remote possibility stemming from inconsistencies presented in the film, depends upon the availability of mental processing capacity. For example, if sub-processes such as localization are not automated and require conscious processing, working memory interference can prevent the construction of an adequately interconnected representation of the patient's anatomy.

To summarize, Lesgold and colleagues [7,8] have characterized expert radiologists as: (1) having the ability to sustain the looking and reasoning cycle even in the face of considerable complexity, (2) being opportunistic planners with very rich recognition and constructive perceptual abilities, (3) being able to ignore irrelevant data, (4) being more able to take immediate account of relevant data, (5) build a thorough representation of the patient's anatomy, (6) quickly begin executing pertinent general plans, (7) exhibit flexibility and tuning of schemata, (8) can analyze several objects that overlap in a film, and (9) are opportunistic in the face of new evidence. Lastly, their schema-driven processing was not found to be consistently successful.

In contrast, the researchers describe the less-experienced radiologist's diagnostic skill performance as characterized by incompleteness in three respects: (1) the confirming or refuting tests are not applied to the invoked schema, (2) a generally appropriate schema is not triggered efficiently enough, and/or (3) the details of the differential diagnosis process are incomplete [8,9,10].

### *3.2 A Cognitive Model of Diagnostic Reasoning in Mammography Interpretation*

The development of a mammography tutor would necessitate an empirical study aimed at extracting the underlying diagnostic reasoning of radiologists in an authentic situation with varying levels of expertise. As such, a study was designed to empirically investigate radiological expertise and the effects of perceptual scaffolding on the diagnosis of difficult mammograms [9]. Unlike previous cognitive studies in radiological expertise, this study: (1) consisted of an ecologically-valid experiment where the subjects diagnosed mammograms without being repeatedly probed (during the diagnostic process); (2) incorporated two experimental conditions (the *authentic* condition involved a natural problem-solving diagnostic task whereby subjects were instructed to read a type-written clinical scenario and interpret a set of corresponding mammograms while the *augmented* condition was identical to the authentic condition except that the critical mammogram features were highlighted) to test the hypothesis that highlighted mammogram findings (critical to the diagnosis) would facilitate the attainment of a diagnostic schema and thus increase diagnostic accuracy, and (3) included methodological techniques allowing for the characterization of the underlying perceptual and cognitive factors by conducting in-depth analyses of the verbal protocols and film tracings of subjects with varying levels of expertise (radiology residents and staff radiologists). The findings comprise one source of empirical evidence from which to base the design of a tutoring environment for training radiology residents in the interpretation of mammograms. Multi-level analyses of video and audio data, including quantitative, qualitative and protocol analyses were conducted to construct a cognitive model of the diagnostic reasoning process of residents and staff radiologists.

Preliminary results based on the analyses of ten staff radiologist (experts) and ten radiology residents (novices) exposed to the two experimental conditions indicated that the diagnostic reasoning process of both residents and staff radiologists is characterized by both forward (data-driven) and backward (hypothesis-driven) chaining. Forward reasoning occurred during the initial diagnostic reasoning task when subjects were extracting the mammogram findings and integrating them into a coherent diagnostic schema leading to a diagnosis (i.e., from data to hypothesis). Backward reasoning was exhibited when subjects attempted to integrate "loose ends" (e.g., irrelevant mammographic and clinical history findings) that did not fit a general diagnostic schema (i.e., from hypothesis to data). The hypothesis that highlighted mammogram findings (augment condition) would facilitate the attainment of a diagnostic schema and therefore increase diagnostic accuracy was supported. In the majority of the cases, residents were able to accurately describe the relevant (highlighted) findings and provide a somewhat accurate diagnosis. However, the quantitative analyses indicated that they also described significantly more non-critical findings than the staff. As expected residents has the most difficulty with eliciting the critical film findings and diagnostic accuracy. The protocols also revealed many individual differences amongst subjects. For example, the expert with the most mammography experience solved all the cases using a purely forward reasoning approach by eliciting all critical mammogram findings and stating the correct diagnosis. In contrast, another expert (with significantly less mammography experience) engaged more in a novice-type strategy by solving most cases using abductive (including both forward and backward) reasoning. This finding is correlated to the number of mammograms interpreted throughout each subject's medical career and related to the fact that few of the subjects specialized in mammography.

## 4. Analyses of the Radiology Teaching Rounds: Eliciting Tutoring Strategies

The tutoring strategies and levels of instructional scaffolding adopted in the prototype are based on the authentic analyses of radiology teaching rounds [10]. These analyses examined the diagnostic reasoning process and teaching methodologies employed by a staff radiologist and six residents during two one-hour mammography rounds. During rounds, the radiologist assigned a teaching file to a resident and he/she was instructed to diagnose the case. The resident placed the four mammograms (craniocaudal and mediolateral oblique views of the left and right breasts) on the viewbox and "talked out loud" while diagnosing the assigned case.

### *4.1 Characterization of the Diagnostic Process*

The diagnostic reasoning process for mammography interpretation involved: (1) reading a brief clinical history, (2) identifying the four mammograms in terms of the technical positioning of each image (craniocaudal and mediolateral oblique views of the left and right breasts), (3) identifying all film findings (locating both normal and pathological findings on each set of mammograms), (4) characterizing each film finding (providing the morphological characteristics of each finding such as a *punctate* calcification), (5) localizing anatomical features and relating them to film findings (e.g., skin thickening in the *left retroaereolar region*), (6) providing a diagnosis or several differential diagnoses, and (7) discussing patient management issues. In cases where differential diagnoses were provided, the residents requested additional diagnostic tests (e.g., ultrasound, fine-needle aspiration, biopsy) to constrain the number of differential diagnoses, and discussed how the results of each test would be used to further constrain the number of differential diagnoses.

### *4.2 Teaching Methodologies*

The teaching methods used by the staff radiologist during the breast disease rounds included modeling, coaching, scaffolding and fading, and articulation (similar to the teaching methods advocated by proponents of the cognitive apprenticeship [11]). The modeling of the diagnostic task was observed when junior residents, in particular, could not pass beyond the listing of some irrelevant findings exhibited on the mammograms. In these cases, the radiologist "walked" the resident through the entire diagnostic process. Occasionally, the radiologist would also provide coaching and support. In the case of intermediate residents, the radiologist would provide scaffolding during the diagnostic process in the form of hints, redirecting their viewing process, and subsequently fading all instructional support when the resident demonstrated the capability to pursue the task on his/her own. The finest illustrations of articulation were observed when the radiologist externalized her reasoning process beginning with the assignment of probabilities to pathological features, followed by the systematic elimination of competing differential diagnoses until the definitive diagnosis was achieved. This teaching method was of superior quality since all residents may potentially benefit from the externalization of the expert's diagnostic thinking.

## 5. Bridging the Gap: Deriving Instructional Principles from Empirical Studies in Authentic Settings

The following section will delineate instructional principles derived by integrating the empirical research on medical cognition, radiological expertise, mammogram interpretation, and analyses of radiology rounds. The successful integration of these several sources of evidence are critical to the design of a theoretically-based and empirically-derived prototype for training both residents and staff radiologists in the interpretation of mammograms. Each of the four instructional principles posited will be supported by existing theoretical and empirical evidence described previously in this paper. The instructional principles incorporated in the mammogram interpretation prototype include: (1) the principle of multiplicity, (2) the principle of activeness, (3) the principle of accommodation and adaptation, and (4) the principle of authenticity (for an extensive overview of some of these principles for supporting computer-supported problem-based learning see [12]).

The **principle of multiplicity** is based on the concept that knowledge is complex, context-sensitive, inter-related and thus instruction should promote multiple perspectives, representations and strategies. This principle is based on the theory of cognitive flexibility [13] in medicine which emphasizes the use of multiple knowledge representations and repeated exposure to instructional content. According to this principle, single mental representations and unitary learning approaches are insufficient for (1) capturing the nature of complex instructional materials and (2) knowledge application in ill-structured domains (such as radiology). A recent study examining the effectiveness of hypermedia versus traditional lectures in radiology has demonstrated the use of multiple knowledge representations through the use of text, digitized video clips and animations of radiology physics concepts do facilitate learning and knowledge application [14]. In other words, the tutor should provide the resident with a stock of breast disease cases that can be accessed in a structured manner according to diagnoses and specific mammographic findings.

The **principle of activeness** is based on the concept that learning is an active process, requiring mental construction and manipulation of the subsymbolic (e.g., gray-scale densities exhibited on mammograms) and symbolic representations (e.g., clinical findings exhibited on mammograms and relevant clinical history findings) that comprise the task environment. Therefore, instruction should foster knowledge construction through problem-solving activities which lead to the development of skill acquisition. This principle reflects the nature of learning through active construction of knowledge facilitated by problem-solving activities. Effective instructional methods should promote planning, reasoning, goal-directed problem-solving, and reflection. This principle reflects the empirical findings in the areas of cognitive skill acquisition [15] and the development of expertise [16].

The **principle of accommodation and adaptation** is based on the concept that the learning process is to a large degree affected by the extent of the learner's existing knowledge. As such, instruction should facilitate adaptability by building upon the learner's existing knowledge, monitoring learner progress and rectifying misconceptions when they arise, and fostering the development of metacognitive skills.

The **principle of authenticity** is based on the concept that learning is sensitive to contextual factors which determine the usability of what is learned and the extent of skill transfer. Therefore, instruction should provide learning activities that are required in the domain, that are valued in the real-world context, and that emulate the real-world environment as much as possible. This principle reflects the recent claims by advocates of situated cognition regarding the need to study the contextual and situational aspects of the cognitive phenomena being studied. In other words, the problem-solving activities provided by the tutor should reflect the authenticity encountered in the resident's work environment. For example, a tutor should provide the tools typically used to solve mammogram cases (such as a magnifying glass) and provide residents with learning opportunities that reflect their daily medical practice.

## 6. A Description of the Mammography Prototype: RadTutor

The mammography prototype offers instructional approaches such as observation, modelling, coaching, scaffolding, fading of assistance, shared problem solving between tutor and student, and situated learning in the context of subsequent knowledge use. Residents are engaged in explicitly justifying hypotheses based on the evidence gathered and rating the medical evidence. The tutor traces the user's problem-solving activities and provides micro and macroadaptive interventions based on the (1) collection of pertinent data by observation, (2) selection of medical evidence, (3) correctness in interpreting the medical evidence, (4) construction of a hypothesis based on the integration of the data , and (5) diagnostic accuracy. The tutoring interventions are based on (1) the concept of creating an effective problem space (EPS) reifying the empirical results of the empirical studies on resident-staff differences, (2) cognitive task analyses obtained from consulting radiologists and residents having varying levels of expertise, and (3) the analyses of authentic radiology rounds between staff radiologists and residents with varying levels of expertise. Residents will be forced to select their hypotheses from a list of options (e.g., carcinoma, fibroadenoma, etc.) and estimate their level of confidence in those hypotheses. The system will monitor all of these actions and determine what advice, scaffolding and hints to generate in response to the user's problem-solving activities. Hyperlinks will be built into the system allowing the user access to detailed explanations regarding pertinent aspects of each case. Finally, an important component of the system will be the ability to provide the user with a graphical representation of the expert's reasoning process in solving each case. The ITS architecture will have hypermedia capabilities such as the hyperlinks, allow access to digitized video clips and audio segments, and display digitized radiological assets (e.g., mammograms) retrieved from CD technology.

The following section illustrates the RadTutor's overall framework. It also provides a detailed description of each of the system's instructional features including (1) present the patient history, (2) show a set of mammograms, (3) list of all findings depicted on the each mammogram, (4) highlight all the findings on each of the mammograms, (5) what is the diagnosis, and (6) variability within the same disease category (illness scripts).

Figure 1 presents the interface of the mammography prototype. The interface is divided into six separate components: (1) the tutor's dialogue box (upper left corner), (2) the resident's dialogue box (lower left corner), (3) the mammogram display area (middle of the interface), (4) a list of mammogram findings (upper right corner), (5) a list of diagnoses (lower right corner), and (6) a floating palette (lower right corner). Solving a typical breast disease case would involve the following steps. The tutor displays the mammograms in a random order and instructs the resident to circle and identify the critical findings and to select a diagnosis. The tutor then provides the clinical history *(the case of a 60 year old woman presented with a mass in the right breast)*. The resident can re-position the images in the mammogram display area.

The tutor provides the resident with feedback concerning the placement of the mammograms and prompts the resident to look for findings. The resident then has the option to either select the critical findings from a list of features or go directly to the images and highlight the critical features. As seen in Figure 1, the resident has selected *large fibroadenoma* from the feature identification list. The tutor has responded with a simple feedback message stating that the selected finding is critical to the diagnosis but incorrectly labelled. The resident has selected a working hypothesis (i.e., diagnosis) of *benign* and placed it in the active diagnosis box. The resident has highlighted a *partly well-circumscribed mass* in the right breast with the aid of the tools provided in the floating palette. This is indicated by the white irregular circle on the right medial mammogram. He/she has also typed a characterization of the mammographic finding (*I think the mass is well-circumscribed*). The tutor will subsequently intervene by asking the resident to rectify his characterization since the mass is ***partly** well-circumscribed* and therefore should yield a correct working hypothesis of *suspicious for malignancy*.

The resident can select as many features from the features identification list and can highlight as many mammographic regions as he/she wants. At a micro level, tutoring interventions, instructional scaffolding, and resident queries are based on one of the features identified. However, at the global level, the scaffolding and tutoring interventions are based on the number of findings selected and highlighted, the accuracy of their characterization and differential diagnoses selected. The dialogue between tutor and resident is presented in dialogue boxes and the termination of a dialogue sequence is represented by the dotted lines.

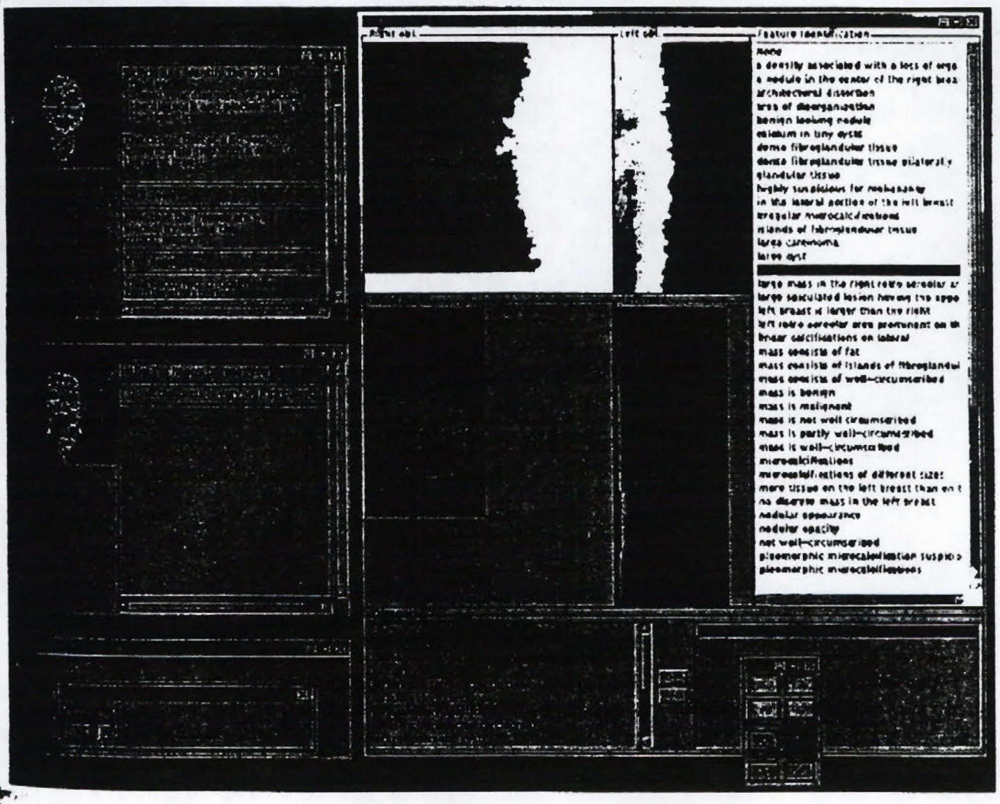



Figure 1. The interface of the mammography prototype.

• **Show Patient History**
First, the user is presented with the patient history. For example, a 60 year old woman presented with a mass in the right breast.
• **Show The Set of Mammograms**
Secondly, the user is presented with the patient's mammograms (typically comprised of left and right mediolateral oblique and craniocaudal views). The environment allows the user to place the mammograms in any order and position by selecting and dragging the image.
• **List and Highlight all of the Findings Depicted on each Mammogram**
Thirdly, the user traces the findings exhibited on each mammogram and/ or selects them from a long list of possible findings. For example, the critical findings related to case #1 include: *dense fibroglandular tissue bilaterally, large mass in the right retroaereolar area, mass is partly well-circumscribed, mass is not well-circumscribe*. If the findings identified are correct, confirmatory feedback will be provided and the findings will be highlighted on each of the mammograms. However, if the findings listed are incorrect the tutor will provide different levels of scaffolding for each depending on the correctness and type of error. The levels of instructional scaffolding and content of feedback messages are based on the results of the authentic analyses of the teaching rounds mentioned earlier in the paper and cognitive task analyses conducted with the consulting staff radiologist. At the macro level, the tutor scaffolds learning by asking a question regarding the region of interest (ROI) where the finding may be located. At the micro level, the tutor asks a specific question (such as "Is there fibroglandular tissue in the upper outer quadrant of the left breast?"). After three attempts the tutor highlights the finding on the mammogram. The use of feedback messages and hints are based on a recent meta-analysis on the effects of feedback in computer-based instruction [17]. This cycle continues until the resident identifies and lists all of the findings exhibited in each case.
• **What is the Diagnosis?**
Once a resident identifies and selects all of the relevant findings, he/she is asked to provide a diagnosis. If the user provides the correct diagnosis he/she will move on to the next step. For example, the correct diagnosis for case #1 is a *mucin-producing carcinoma* (colloid carcinoma or mucinous carcinoma). However, if the user provides an incorrect diagnosis (e.g., *fibroadenoma, cyst, metastatic disease*), different levels of scaffolding are available depending on the correctness of the diagnosis. The first level of intervention consists of reviewing the patient's history and the critical mammography findings. This level of scaffolding is aimed at establishing a diagnostic schema by reviewing the results of the previous step (List and Highlight all of the Findings Depicted on each Mammogram).

The second level of intervention displays the digitized clip illustrating the consulting radiologist's "externalization" of the clinical reasoning involved in correctly diagnosing the mammograms. This approach includes an extensive elaboration of the radiologist's case resolution by establishing the critical findings on the mammogram, selecting the findings from the list of possible findings, and then reasoning the case through. Presently, we are incorporating other features into the tutor such as (1) a glossary of breast disease terms, (2) digitized images and video clips demonstrating how to place mammograms on a viewbox, (3) a scratch pad so that residents can "cut and paste" screens from the tutor and either send them to print or save them to disk, (4) a list of radiology and medical references including abstracts, and (5) network access to other teaching files and on-line resources such as electronic radiology journals.

### *6.1 Variability within the Same Disease Category: The Role of Illness Scripts*

The concept of diagnostic schemas and their role in diagnostic reasoning and integration of medical knowledge has recently been consolidated by [18] in terms of postulating a developmental theory of medical expertise based on the concept of an illness script. An illness script is an elaboration of the more general idea that medical diagnostic and clinical knowledge is organized in a schema. These schemas are disease oriented and contain three components including information about patient background factors, complaints, signs, symptoms, diagnostic procedures, courses of disease progression, and treatments. The first component is entitled the **Enabling Conditions** which includes factors that influence the probability that someone will get a particular disease. For example, age, sex, physical appearance, heredity dispositions, occupations, living environment, medical history, previous and current medication, and risk behavior. The second component is entitled **The Fault** which includes the pathophysiological disturbances in the body, couched in a biomedical model. These are usually subsumed under a diagnostic label. The third component is entitled the **Consequences** and includes complaints, signs, and symptoms. The tutor provides a systematic approach for presenting the variability of each disease category by exposing the resident to multiple cases with varying enabling conditions, faults and consequences. The approach facilitates the development of mammography expertise by allowing the user to acquire an illness script of each disease category comprising all of the variability within that particular category.

## 7. Conclusion

In conclusion, this paper presented a cognitively-based and empirically-derived approach for the design of a prototype computerized tutor to train radiology residents in diagnosing mammograms. The approach included: (1) presents a critique of existing computer-based radiology training environments, (2) provides an overview of the previous cognitive studies on radiological expertise, (3) reports the results of an empirical study aimed at attaining a cognitive model of diagnostic reasoning in mammogram interpretation, (4) delineates teaching strategies based on analyses of radiology teaching rounds, (5) proposes instructional principles for the design of the mammography tutor, and (6) provides a description of the system components of a computer-based mammography interpretation tutor.

## References


[1] S.P. Lajoie, Computers Environments as Cognitive Tools for Enhancing Learning. In: S.P. Lajoie and S.J. Derry (eds.), Computers as Cognitive Tools. Erlbaum, Hillsdale, NJ, 1993, pp. 261-288.

[2] R.T. Macura et al., Computerized Case-Based Instructional System for Computed Tomography and Magnetic Resonance Imaging of Brain Tumors, Investigative Radiology 29(4) (1994) 497-506.

[3] M. Sharples et al., The MR Tutor: Computer-Based Training and Professional Practice. In: J. Greer (ed.), Proceedings of AI-ED'95. ACCE , Charlottesville, VA, 1995, pp. 429-436.

[4] M. Sharples et al., The Cognitive Basis for an MRI Tutor. In: Proceedings of the East-West Conference on Computer Technologies in Education (part 1). Crimea, Ukraine, pp. 214-219).

[5] M. Sharples, Computer-Based Tutoring of Visual Concepts: From Novice to Expert, Computer Assisted Learning 7 (1991) 123-132.

[6] A. Lesgold, Problem Solving. In: R.J. Sternberg and E.E. Smith (eds.), The Psychology of Human Thought. Cambridge University Press, Cambridge, 1988, pp. 188-213.

[7] A. Lesgold et al., The Acquisition of Perceptual Diagnostic Skill in Radiology (Tech. Rep. No. PDS-1). University of Pittsburgh, Learning Research and Development Center, Pittsburgh, 1981.

[8] A. Lesgold et al., Expertise in a Complex Skill: Diagnosing X-Ray Pictures. In: M. Chi, R. Glaser and M. J. Farr (eds.), The Nature of Expertise. Erlbaum, . Hillsdale, N.J., 1988, pp. 311-342).

[9] R. Azevedo, Radiological Expertise and the Effects of Perceptual Scaffolding on the Diagnosis of Mammograms. Unpublished doctoral dissertation, McGill University, Montreal, In prep.

[10] R. Azevedo et al., Tutoring Complex Visual Concepts in Radiology. Paper presented at the XXVI International Congress of Psychology, Montreal, Quebec, August, 1995.

[11] J.S. Brown et al., Situated Cognition and the Culture of Learning, Educational Researcher 18 (1) (1989) 32-42.

[12] T.D. Koschmann et al., Computer-Supported Problem-Based Learning: A Principled Approach to the Use of Computers in Collaborative Learning. In: T.D. Koschmann (ed.), CSCL: Theory and Practice of an Emerging Paradigm. Erlbaum, Hillsdale, NJ, 1996, pp. 83-124.

[13] R.J. Spiro et al., Cognitive Flexibility Theory, Constructivism, and Hypertext: Random Access Instruction for Advanced Knowledge Acquisition in Ill-Structured Domains, Educational Technology 31(5) (1991) 24-33.

[14] S. Shaw, et al., The Effectiveness of Hypermedia Instructional Modules for Radiology Residents, Canadian Journal of Educational Communication 24(3) (1995) 245-268.

[15] K. VanLehn, Cognitive Skill Acquisition, Annual Review of Psychology 47 (1996) 513-539.

[16] K.A. Ericsson and A.C. Lehmann, Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints, Annual Review of Psychology 47 (1996) 273-305.

[17] R. Azevedo et al., A Meta-Analysis of the Effect of Feedback in Computer-Based Instruction, Journal of Educational Computing Research 13(2) (1995b) 109-125.

[18] H.G. Schmidt et al., A Cognitive Perspective on Medical Expertise: Theory and Implications, Academic Medicine 65(10) (1990) 611-621.

gence

'a and W. Swartout

gence
Bases VIII
:tificial Systems
applications
es VII

Artificial Intelligence 95

Bases VI
tificial Life

:ion and Modelling

r Expertise Modelling
ing
ises V
Artificial Intelligence—93

Bases IV

:s III
-91
ansfer
ng and Knowledge Bases

n
Bases
-Based Environments

Artificial Intelligence—89

:ion
e (SCAI '88)

# Artificial Intelligence in Education

## Knowledge and Media in Learning Systems

Edited by

B. du Boulay
*School of Cognitive and Computing Sciences*
*University of Sussex, Brighton, UK*

and

R. Mizoguchi
*Institute of Scientific and Industrial Research,*
*Osaka University, Osaka, Japan*

Proceedings of AI-ED 97 World Conference
on Artificial Intelligence in Education
Kobe, Japan

IOS
Press

Ohmsha

Amsterdam • Berlin • Oxford • Tokyo • Washington, DC