Expert Problem Solving in a Visual Medical Domain

Proceedings of the Annual Meeting of the Cognitive Science Society, 20(0)

Azevedo, Roger

1998

Expert Problem Solving in a Visual Medical Domain

Roger Azevedo (azevedo@andrew.cmu.edu)
Department of Psychology; Carnegie Mellon University
Pittsburgh, PA 15213 USA

Abstract

This study examined the problem solving strategies used by staff radiologists and radiology residents during the interpretation of difficult mammograms. Ten radiologists and ten residents diagnosed 10 cases under two experimental conditions (authentic and 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. Verbal protocols were analyzed and revealed that mammography interpretation was characterized by a predominant use of data-driven or mixed-strategies depending on case typicality and clinical experience. Repeated measures ANOVAs revealed that the radiologists scanned the cases significantly faster than the 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 regardless of experimental condition both groups (a) used the same types of operators, control processes, diagnostic plans, (b) committed the same number of errors, and (c) committed case-dependent errors. Overall, 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 visual cognitive task. The results have been applied to the design of a computer-based tutor for training residents to interpret mammograms. Future empirical directions include building a more comprehensive model of the perceptual and cognitive processes underlying mammogram interpretation by converging eye-movement, cortical activation (e.g., fMRI) and verbal protocol data.

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.

This study was designed to investigate the problem solving strategies of staff radiologists and radiology residents during the interpretation of difficult breast disease cases depicted on mammograms. 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.
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 diagnostic accuracy).

The following section presents a brief overview of the major empirical findings from cognitive studies of radiological expertise.

Cognitive Studies of Radiological Expertise

Radiological expertise has been investigated by numerous authors employing disparate theoretical perspectives 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 (Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988; Rogers, 1996) have actually investigated the underlying cognitive and perceptual factors. In fact, these studies have all been conducted in the area of chest radiography. As a result, a fundamental understanding of the nature and acquisition of expertise in other radiological sub-specialties such as mammography has yet to be achieved.

In terms of cognitive research, there have been few explicit accounts of the problem solving strategies of radiology residents and staff radiologists during the diagnostic process. Lesgold and colleagues (1988) 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 participants' 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 determines 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 (WM) interference can prevent the construction of an adequately interconnected representation of the patient's anatomy.

To summarize, Lesgold and colleagues (1988) have characterized expert radiologists as being: (1) able to sustain the looking and reasoning cycle even in the face of considerable complexity, (2) opportunistic planners with very rich recognition and constructive perceptual abilities, (3) able to ignore irrelevant data, (4) more able to take immediate account of relevant data than residents, (5) able to build a thorough representation of a patient's anatomy, (6) able to quickly begin executing pertinent general plans, (7) flexible and able to tune schemata, (8) able to analyze several objects that overlap in a film, and (9) opportunistic in the face of new evidence. Lastly, their schema-driven processing was not found to be consistently successful.

In contrast, the researchers characterize the less-experienced radiologists' diagnostic performance as incomplete 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 (Lesgold et al., 1988).

In a recent study in chest radiography interpretation, Rogers (1996) examined the interaction between perception and problem solving. Verbal protocol data was collected from 8 residents and 2 staff members while they examined seven computer-displayed chest x-rays. Results indicated that accurate perceptual characterization of a finding might 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 and problem solving activities. 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 (1) confirmation of expectation of secondary findings to support diagnostic hypotheses, (2) use of features of primary findings to rule out competing findings and diagnostic hypotheses, and (3) use of features of primary findings to match or contradict expectations.

More recently, Faremo (1997) investigated the problem solving processes used by 8 3rd-year medical students and 8 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 and the analyses revealed that both groups differed in their problem solving behaviours. Furthermore, groups differed significantly in the accuracy of their responses for findings, diagnoses/differential diagnoses, follow-up actions, 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. Detailed analyses were conducted on a subset of the protocols and additional between and within group differences 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 of this study, several recommendations were made concerning how a computer-based instructional 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 (e.g., chess and physics). The few studies that have been reviewed have provided an initial characterization of the diagnosing process, specification of the top-down and bottom-up processes involved in diagnostic reasoning, and the role of perceptual and problem solving processes. This research has focused mainly on the area of chest radiography and in-depth analyses have revealed differences across levels of medical experience. 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 similar investigation is required. More specifically, a cognitive model characterizing the underlying differences in diagnostic problem solving between radiologist professionals with different levels of expertise is needed. Such a model has been developed by studying radiologists' performance during the interpretation of difficult breast diseases exhibited on mammograms using appropriate cognitive science methodologies.

Method

Participants

A total of 20 participants, 10 staff radiologists and 10 radiology residents from McGill University's teaching hospitals participated in this study. The 10 staff radiologists had MD degrees and Board Certification in radiology and were affiliated with one of the teaching hospitals. Their post-residency training ranged from 3.5 to 34 years (mean of 20.3 years), including a range of 5 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 0 to 200 mammograms), and "seen" (i.e., viewed but not diagnosed) an average of 68 mammograms per week (range 5 to 360 mammograms).

The 10 radiology residents had MD degrees and were on rotation at one of McGill's five teaching hospitals. This group was comprised of 2 3rd-year, 1 4th-year, and 7 5th-year residents. All of the residents had completed one mammography rotation. They reported to have 0 to 12 months of mammography training (mean of 6 months). Sixty percent of the residents reported to have "read" between 25 to 100 mammograms, while the other 40% reported to have "read" between 200 to 1,000 mammograms. None of the residents reported that they "read" or "see" mammograms on a weekly basis.

Breast Disease Cases

Ten 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's teaching files. Each case was comprised of a brief clinical history and at least 4 mammograms including the CC (craniocaudal) and MLO (mediolateral) views of the left and right breasts. The cases include 3 benign and 7 malignant cases. The diagnoses for each case were confirmed by pathology reports from specimens extracted from the breast tissue corresponding to each case. The breast disease cases included cases typically encountered in mammography textbooks and clinical research articles, atypical cases infrequently encountered in daily practice, and cases with typical mammographic manifestations encountered in daily practice. These cases include mammographic features that ranged from fairly obvious to detect to cases that require the use of a magnifying glass to detect subtle mammographic features.

Coding Scheme

A coding scheme was constructed based on the content analysis of breast disease and mammography, theoretical and methodological articles in cognitive science and the results of previous studies in various relevant areas such as medical cognition (Hassebrock & Prietula, 1992; Patel, Arocha, & Kaufmann, 1994; Patel & Groen, 1986; Patel & Ramoni, 1997), discourse processing (Bracewell & Breuleux, 1993; Breuleux, 1993; Frederiksen, 1975), syntactical analysis, and artificial neural networks for mammogram interpretation.

Fifty of the 200 protocols collected 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, 1993; 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 (e.g., hypothesis generation) and comprised a total of 30 operators. *Control processes* were comprised of diagnostic planning, goal verbalizations, and meta-reasoning.

Inter-rater reliability was established by recruiting a graduate student with experience in the area of breast disease and mammography and training her to use the coding scheme. She was instructed to independently code 20/200 randomly selected protocols thus yielding a reliability coefficient rating of .92.

Research Design

A mixed factorial design consisting of 10 participants nested across 2 levels of radiological expertise (between-subjects factor) crossed with 2 experimental conditions (within-subjects factor).

Procedure

Participants were tested individually. The experimenter provided the participant with a 1-page handout of instructions for the diagnostic task. 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 10 cases that were counterbalanced across experimental conditions and participants.

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/or 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" (Ericsson & Simon, 1993) throughout the entire diagnostic process. The experimental procedure was repeated for each subject until he/she diagnosed all ten cases under the two experimental conditions (5 authentic and 5 augmented) without any time constraints.

Results

The results of this study are presented in the context of (a) a cognitive 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.

Cognitive Model of Mammogram Interpretation

The cognitive model of diagnostic problem solving in mammogram interpretation was constructed from a cognitive task analysis and content analysis of the domain and refined based on the verbal protocol analyses of 40 randomly selected protocols. 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"* (e.g., from reading the clinical history to specifying subsequent examinations) and/or an *"iterative approach"* in which the results of a step may feed back to previous steps in the model. The linear approach is analogous to the use of a data-driven problem solving strategy whereby a subject reads the clinical history, scans the set of mammograms, and provides a diagnosis. The iterative approach is analogous to the use of a mixed-problem solving strategy (i.e., includes both data-driven and goal-driven problem solving strategies). For example, following the initial scanning and characterization of the mammographic findings the radiologist may postulate a set of differential diagnoses which will lead him/her to inspect particular area(s) of a mammograms.

Problem Solving Strategies

Overall, the in-depth analyses (of the same 40 protocols) indicated that diagnostic problem solving in mammography is characterized by (a) the predominant use of a data-driven diagnostic strategy, (b) the use of a mixed-strategy depending on case typicality and clinical experience, and (c) rapid schema-based problem-solving which facilitates search and the characterization of mammographic features and integration of clinical history cues (followed by accurate diagnosis and subsequent radiological recommendations).

Frequency of Problem Solving Operator Use

The results indicate that residents used more operators than the staff (based on the same 40 protocols used above). 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 4 operators account for 76% of all operators used by all subjects regardless of level of expertise and experimental condition.

Frequency of Control Processes Used

The frequency of control process use (based on the same 40 randomly selected protocols) revealed that the staff used slightly more control processes than the residents (47 as compared to 41). Diagnostic planning was the most often used control process regardless of level of expertise and experimental condition.

Number of Radiological Findings, Observations, and Diagnoses

Repeated measures ANOVAs failed to reveal any significant differences in the mean number of radiological findings, observations and diagnoses between the groups and across experimental conditions. On average, however, participants identified at least 1 radiological finding, made 3 radiological observations, and gave 1 diagnosis per case.

Scanning Time for Data Acquisition

Scanning time was operationally defined as the amount of time (in seconds) it took a participant to attain an initial representation of a breast disease case. A repeated measures ANOVA indicated that the staff radiologists were significantly faster than residents in scanning the mammograms (F [1,18] = 4.89, $p < .05$). On average, a staff radiologist took 46 seconds to scan a case while a resident took 66 seconds.

Reading Time for Diagnosis

Reading time was operationally defined as the total amount of time (in seconds) it took a participant to solve a breast disease case. A repeated measures ANOVA did not reveal a significant main effect for expertise (F [1,18] = 1.5) or condition ($F = 0.11$), and there was no interaction ($F = .0009$). On average, a staff radiologist took 175 seconds to read a case while a resident took 200 seconds.

Overall Diagnostic Accuracy

Overall diagnostic accuracy includes the combination of diagnoses (e.g., malignant) and radiological recommendations (e.g., perform a biopsy). The categories included *accurate*, *indeterminate* and *inaccurate*. For example, a diagnosis of a carcinoma followed-up by an excisional biopsy would constitute an *accurate* overall diagnosis. The percentages for overall diagnostic accuracy provided by both groups across experimental conditions were calculated. A 2X2 $\chi 2$ 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 (χ^2 [1, N = 200] = .57, $p > .05$).

Overall, 25% of the participants (5/20), including 3 staff and 2 residents correctly diagnosed and provided the correct subsequent recommendations for the 10 breast disease cases.

Analysis of Errors Based on Overall Diagnostic Performance

An in-depth analysis of the 34 (of 200 protocols collected) errors committed by participants based on overall diagnostic accuracy across experimental conditions was performed. Analyses of the 34 coded protocols revealed five types of the errors including:

(1) a **perceptual detection error** was a failure to detect a mammographic finding;

(2) a **finding mischaracterization error** was an incorrect characterization of a mammographic finding;

(3) a **no diagnosis error** was the detection, correct identification, and characterization of a mammographic finding but a failure to make a diagnosis;

(4) a **wrong diagnosis error** was the detection, correct identification, and characterization of a mammographic finding but proposing a wrong diagnosis; and,

(5) a **wrong recommendation error** was the correct detection and characterization of a mammographic finding, and proposing a diagnosis at some level of abstraction, but proposing an inappropriate subsequent examination.

The analyses of the frequency and types of errors by level of expertise and experimental was conducted. Overall and in descending order of frequency, the results indicate errors consisted of wrong recommendations (38%), perceptual detection (26%), finding characterization (24%), no diagnosis (6%), and wrong diagnosis (6%).

The analyses revealed that regardless of level of expertise, the commission of errors was case-related. Furthermore, 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.

Conclusions

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. The results indicated that staff radiologists scanned the cases significantly faster than, radiology residents. No differences were found on several aspects of the groups' performance 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 visual cognitive task.

The contributions of this study include: (1) a cognitive model of diagnostic problem solving in the area of mammography, (2) enhanced understanding of the problem solving processes underlying mammography interpretation, and (3) empirical evidence for the design of a computerized tutoring prototype for training radiology residents to diagnose mammograms.

The theoretical implications of the study include the need for further studies to build a more comprehensive model of the perceptual and cognitive processes underlying mammogram interpretation. Secondly, the results are presently being applied in the development of a basic conceptual framework for the design of the RadTutor, a computer-based tutor for training radiology residents to interpret mammograms.

Future Directions

This section presents the empirical and practical implications of this study. These implications are presently being pursued in two different directions: (1) the design of the RadTutor, a computerized tutoring prototype for training residents to diagnose mammograms, and (2) an empirical study that will incorporate eye-movement, cortical activation, and verbal protocol data to build a comprehensive diagnostic model of mammogram interpretation.

Instructional Implications - RadTutor

The second project is the design of the RadTutor (Azevedo, Lajoie, Desaulniers, Fleiszer, & Bret, 1997), a computer-based prototype for training radiology residents to interpret mammograms. The RadTutor presently incorporates the model of mammogram interpretation, problem solving strategies, and error analyses resulting from this study. Furthermore, it is also based on other research in instructional psychology, cognitive science, artificial intelligence and education.

Directions for Future Empirical Research

An empirical 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). 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).

Recent advances in brain imaging have facilitated the construction of comprehensive models of cognitive processes through the convergence of physiological and psychological research. Non-invasive brain imaging techniques such as 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 participant's eye-movements and the activity of his/her cortical regions (e.g., activation in the pre-frontal cortex active during goal-driven problem solving) would also be captured using an eye-tracking system that is embedded in a fMRI scanner which enables whole brain imaging.

Multilevel analyses (verbal, performance, cortical) would be conducted in order 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 subsequent studies in the area of medical visual diagnosis will 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.

Acknowledgements

This research was supported by a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada (SSHRC) and a McGill University - SSHRC grant. The author also wishes to thank Monique Desaulniers, David M. Fleiszer, Susanne P. Lajoie, Sonia Faremo, Patrice M. Bret, and the radiologists and residents for their participation in this study.

References

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

Azevedo, R., Lajoie, S.P., Desaulniers, M., Flesizer, D., & 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.), *Artificial intelligence in education* (pp. 386-393). Amsterdam: IOS Press.

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.

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.

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

Hassebrock, F., & Prietula, M.J. (1992). A protocol-based codong schema for the analysis of medical reasoning. *International Journal of Man-Machine Studies, 37*(5), 613-652.

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.

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: Erlbaum.

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

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., & 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.

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

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.