A Study of Visual Reasoning in Medical Diagnosis

Erika Rogers

Dept. of Computer & Information Science
Clark Atlanta University
Atlanta, GA 30314
erika@pravda.gatech.edu

Abstract

The purpose of this paper is to describe experimental work conducted in the area of diagnostic radiology, with an emphasis on how perception and problem solving interact in this type of task. This work was part of a larger project whose goals included the development of an information-processing model of visual interaction, and the subsequent design of an intelligent cooperative assistant for this domain.

Verbal protocol data was collected from eight radiologists (six residents and two experts) while they examined seven different computer-displayed chest x-rays. A brief overview of the methodology and analysis techniques is presented, together with specific results from one x-ray case. More general results are then discussed in the framework of issues important to the later modeling effort.

Introduction

The current trend towards telemedicine diagnosis has created a need for enhanced software capabilities which not only enable high-speed image transmission, but which can also support the visual reasoning needed both at the local and remote sites. The cognitive load still remains squarely on the human expert's shoulders who must examine these images and make diagnostic evaluations, often without the benefit of any further tools, and with the disadvantage that these images may be both spatially- and contrast-degraded from the original film. The importance of AI tools to support such activities is therefore increasing. However, in order to build these tools, human capabilities in these tasks must be better understood, and this requires extensive cognitive study and engineering.

The study described in this paper was part of a larger project which was conducted in collaboration with the Radiology Department of Emory University Hospital. This project had three major goals: 1) to explore the nature of radiological diagnosis with a view to understanding how perception and problem solving exchange information in this type of visual reasoning task; 2) to incorporate this knowledge into an information-processing type of model of visual interaction; and 3) to utilize this cognitive foundation for the design of an intelligent cooperative assistant. The purpose of this paper is to describe some of the initial experimental work, and to show how it led to insights about visual reasoning activities.

Studies of the radiological process have varied from eye movement studies (Kundel & Nodine, 1975), to task-related hand movements (McNeill et al., 1988), to cognitive modeling studies of expertise involving think-aloud verbal protocols (Lesgold et al., 1981). While the latter provided the strongest

methodological guidelines for the study described in this paper, more emphasis was placed on the range of novice-expert performance rather than on the differences. Also, the main part of this study utilized digitized computer-displayed x-ray images rather than films, since the ultimate goal of the project was to design a computer-based assistant. A preliminary study on the effects of the laboratory environment and the computerized images showed no noticeable impact on the subjects' ability to perform diagnosis (Rogers, 1992), and therefore the next phase proceeded with more extensive think-aloud protocol collection.

Data Collection and Analysis

The goal of the verbal protocol data collection was to obtain in-depth, detailed data from the subjects while they were interpreting and diagnosing a selection of computer-displayed chest x-rays. Eight participants were recruited from the Emory University Hospital radiology program, and included two experts, plus two each from second- to fourth-year residency. Seven cases were selected for display and included one normal chest x-ray as well as the following abnormalities: lung lesion due to bronchogenic carcinoma, hilar adenopathy, tuberculosis, lung mass with appearance of elevated diaphragm, and mitral stenosis. Three of the seven images had a brief case history associated with them, and the images were presented in the Imaging Sciences laboratory at the hospital.

For each case, the subjects viewed the image as long as desired, while concurrently articulating their thoughts. When examination of the image was completed, a formal diagnostic report was produced, and each case was concluded by a set of clarifying questions, subject ratings for image quality, confidence in diagnosis and case difficulty, as well as some retrospective questions for archival purposes. All seven cases were completed in one session, which was videotaped and subsequently transcribed.

In the analysis of this data, the focus was primarily on the functions or characteristic actions that could be attributed to perception, problem solving, and the interaction between the two. Therefore, the first goal was to develop an appropriate encoding scheme which would enable these kinds of concepts to be extracted from the verbal reports.

A preliminary task analysis combined some of the earlier experimental results with Lesgold's task description of radiological diagnosis (Lesgold et al., 1981) and other problem solving concepts in medical diagnosis. A subset of the actual protocols (three of the seven cases) was then used to refine this general concept list into a more comprehensive encoding

	Primary Abnormality	Secondary Abnormality	Other Concepts
↓ ↓ T	LOOK-AT(ANATOMY) LOOK-AT(FINDING1.1) DESCRIBE(SIZE) DESCRIBE(EDGE) DESCRIBE(SHAPE)...		
I M			LOOK-AT(ANATOMY) CLASSIFY(NORMAL)
E ↓ ↓		LOOK-AT(ANATOMY) SEEK(FINDING2.1) NOT-FOUND(SEEK)	

Figure 1: Concept Categories of Encoding Scheme.

Medical Concepts — Anatomy | Finding | Diagnosis

Descriptive Concepts — Spatial References | Features | Classification

Cognitive Concepts — Visual/Physical | Mental | Additional

Figure 2: Example of Contextual Encoding of Statements.

scheme, consisting of three major categories and a number of minor categories as shown in Figure 1. This encoding scheme was then applied to all of the remaining thirty-two verbal protocol reports[1]. The fully encoded protocols were then analyzed with respect to task-related and time-related patterns, as shown in Figure 2. (Experience-related differences were not studied at this time.)

It was found that similar clusters of concepts did occur at corresponding times for particular cases, that the groupings within the clusters or patterns showed tendencies towards perceptual or problem solving compositions, and that there appeared to be an ordering of activities in the movement towards a diagnostic solution.

A summary of the bronchogenic carcinoma case is now presented together with the expert's diagnostic report in Figure 3 to illustrate some of the features of the analysis process. This case provided the richest source of information due to the perceptual features of the image as well as the problem solving complexities of the diagnosis itself.

Figure 4 shows the subjects' ratings of this case according to image quality, overall confidence in diagnosis, and case difficulty. The image quality was considered to be adequate or better by the majority of the subjects, and therefore this did not seem to impair the examination of the case. This case was considered to be challenging by five of the subjects, even though the overall average assessment of difficulty fell somewhere between easy and challenging (2.4/5.0). Overall confidence tended to be fairly high (3.8/5.0) with only two subjects registering medium or lower confidence in their diagnoses. (In this case it was interesting to note that the subject with the lowest confidence also had the worst performance

with respect to noticing the critical secondary abnormality and producing a correct diagnosis).

Almost all of the subjects noticed and labelled the abnormality in the lung immediately upon viewing the x-ray image. This abnormality was then described by different descriptive and spatial features including (in order of frequency): size, horizontal localization, edge, vertical localization, shape, texture, out-of-plane localization and configuration. On the average, the subjects mentioned six of the above categories, which made this type of finding the most detailed in description from all of the cases examined.

One particularly noteworthy aspect of this case concerned the use of secondary findings in the diagnostic process. The bone lesion in the left posterior fifth rib was an important piece of evidence, as explained in the expert's report. Of the eight subjects, five saw the bony abnormality, three of these correctly identified it as a bone lesion, and two actually used the correct identification in the formulation of the correct diagnostic hypothesis.

The activity of localizing the primary finding also appeared to be very important, although it was perhaps made somewhat more difficult in the test situation because only one view of the patient was presented. It is more common to have two views (frontal and lateral) available, and this provides more information, particularly for anterior-posterior localization.

Results

The following discussion presents an overview of some of the results of the protocol analysis with respect to the roles of description, levels of abstraction, context, attention and expectation. These were considered to be key issues which would impact the subsequent development of the information-processing model.

[1] An independent observer was recruited to participate in a reliability study. Details can be found in (Rogers, 1992)

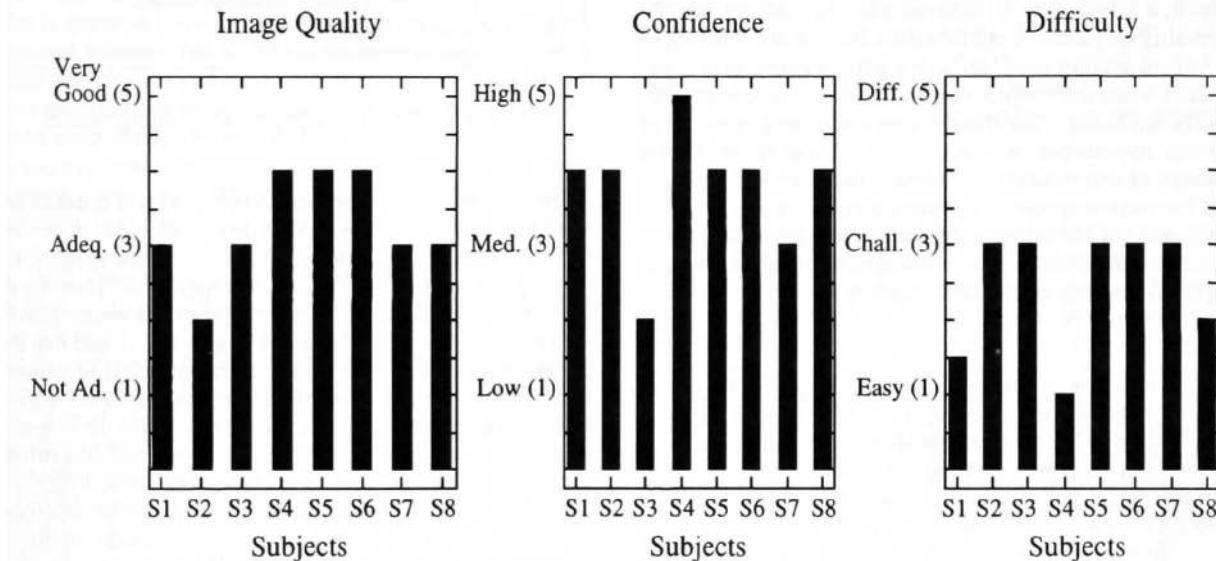Figure 3: Expert's Report for Case 3 - Bronchogenic Carcinoma.

Figure 4: Subject Ratings for Case 3.

Description

For all the cases considered, the characterization of abnormalities in terms of distinct descriptive features was accomplished using a relatively small number of categories. Size, shape, edge, texture, and quantity descriptions were used in varying degrees, depending upon the type of abnormality presented. Sometimes features of the primary finding figured notably in the diagnostic process, while at other times, the findings were described in less detail, and other information such as case history figured more prominently in eliciting accurate diagnostic hypotheses. This suggested that the importance of particular features may vary depending upon the type of abnormality under consideration, and an accurate perceptual characterization of a finding may still not be sufficient to produce a distinct disease diagnosis.

Abnormalities were also described in terms of spatial relations, with references to horizontal and vertical directions most prevalent. The out-of-plane third dimension appeared to be important in cases where there was a mass that appeared to span more than one anatomical landmark. In these cases, shadows on the frontal two-dimensional view could give some clues to location, if the film was of relatively good quality, but often, a lateral view would have served to disambiguate these cases.

Levels of Abstraction

The characterization of findings (i.e., abnormalities in the image) at different levels of abstraction provides an important transition between low-level perceptual detection and higher level problem solving activities. For example, the statement "there is a density" really only means "I have detected a light area", without conveying any further semantic information regarding what this object might represent. On the other hand, a term such as "there is a mass" implies certain associated features such as size, shape and edge, while a statement about a "malignant tumor" suggests not only specific features, but specific diagnostic hypotheses as well.

215

15.	Now...let's see, the heart's normal.
16.	The hilar structures are normal.
17.	The aorta's normal.
18.	The trachea's normal.
19.	The right para...stripe, tracheal wall seems to be normal.
20.	The bony structures are normal.

Figure 5: Landmark Search.

Not only is the type of labeling activity important, but the manner in which the labeled percepts are used can seriously affect the outcome of the diagnostic process. Different levels of oversights ocurred in the transition between perception and problem solving. At the perceptual level, a *detection* oversight occurred when the subject did not notice or see the abnormal object or feature at all. On the other hand, at the identification level, a *labeling* error occurred when the subject saw the abnormality in question, but labeled it incorrectly. Finally, at the problem solving level, difficulty with *integration* was encountered when the subject saw and labeled the abnormality correctly, but failed to use this information in the generation of diagnostic hypotheses. An implication of these results is that the design of computerized assistance really should span the visual interaction spectrum between perception and problem solving, and not just address the poles of the problem, for example, through either image enhancements to aid perception or expert systems to aid decision-making.

Context

When the radiologist begins a diagnostic session, the information available prior to viewing any of the images includes the knowledge of the particular anatomical region under consideration and the particular imaging modality used to collect the image data. The anatomical region normally implies a certain set of anatomical objects in a particular configuration (e.g., a chest normally contains two lungs, a heart on the lower left side, etc.), while the imaging modality calls into play knowledge of the kinds of perceptual cues that are to be expected.

This knowledge is typically organized into a kind of checklist that leads to a relatively orderly plan for examination. An example of this is seen in Figure 5, which shows a sequential set of statements by a subject looking at the normal chest case.

Unlike the landmark search, which can be said to always be of a top-down nature, the search for secondary abnormalities may be either bottom-up or top-down. If no diagnostic hypothesis has been invoked, and the subject is still gathering data, then the search for such abnormalities may appear as a subplan within the landmark search. Figure 6 illustrates such a bottom-up search in the sarcoidosis case. On the other hand, top-down reasoning implies that one or more diagnostic hypotheses are currently active, and that particular secondary findings might support these hypotheses. An example of this is seen in Figure 7, which is an excerpt from the mitral stenosis case.

Further suggestions of plan-like activity can be found in the description of a primary finding in terms of its features and location. For some types of findings, a large number of characteristic features are expressed and utilized in the diag-

6.	Looking at the rest of the lungs, I don't see any areas of consolidation or evidence of pneumonia there.
7.	I don't see any effusions.
8.	I'm going to look at the remainder of the superior mediastinum
9.	to see if I see any other adenopathy
10.	and sort of in the left heart border I see a kind of a double bulge

Figure 6: Bottom-Up Secondary Finding Search.

7.	but the first thing I saw was the heart,
8.	and I, it's just a little bit big.
9.	Now patients with heart failure can get something like a cough with a little bit of pink frothy sputum.
10.	It's not really sputum, it's just sometimes they can cough up some pink stuff.
11.	So I'm going to think about heart failure,
12.	and look for pulmonary edema,
13.	or signs of heart failure.

Figure 7: Top-Down Secondary Finding Search.

nostic process. An example is shown in Figure 8, taken from a subject's report on the tuberculosis case. In other cases, features were sought which led to either more specific finding hypotheses, or even diagnostic hypotheses. Thus the data supports the claim that the "direction of reasoning provides the procedural context" (Evans & Gadd, 1989), and that there are different types of plans, and different levels of planning. Furthermore, the direction of reasoning affected the different types of perceptual information used to support the diagnostic stages. Bottom-up or data-driven reasoning was supported by use of secondary *findings* to generate diagnostic hypotheses, use of *features* of primary findings to specialize labeling of primary findings, and use of *features* of primary findings to generate diagnostic hypotheses. On the other hand, top-down or expectation-driven reasoning involved confirmation of expectation of secondary *findings* to support diagnostic hypotheses, use of *features* of primary findings to rule out competing findings and diagnostic hypotheses, and use of *features* of primary findings to match or contradict expectations.

Attention

In the cases reviewed, at least two different types of attention activity were noted. The first, characterized by a relatively

5.	I'm kind of focussing in on the main abnormality, which is in the right upper lobe,
6.	and I see a cavitary lesion in the right upper lobe, with some atelectatic changes.
...	
13.	The lesion is poorly defined, and has a cavitary area in the center, and the wall is probably about 3 to 4 millimeters thick, and it's irregular.
14.	There are some linear densities leading from the hilar region to the mass.

Figure 8: Features of Primary Finding.

Figure 9: Expectations.

fast noticing and labeling of an abnormality as soon as the x-ray image appeared was called "immediate visual capture", and it was often coupled with a brief description of the abnormality in question (for example, size and shape). This type of attentional behavior has also been described by Kundel and Nodine who showed that even under restricted conditions, selective attention was drawn to the area of the chest having the greatest abnormality (Kundel & Nodine, 1975).

In more general experiments on attention, Treisman showed that simple property differences (e.g., color, brightness, or line orientation) would be seized upon by the initial stage of visual processing so that they appear to "pop out" of a scene (Treisman, 1982). Immediate visual capture appears to be consistent with these findings: if an object composed primarily of water (such as a tumor) overlies an object composed primarily of air (such as lung), then the first object will appear brighter. If the same object overlies bone, it will appear darker. Thus certain types of brightness features in the x-ray image may be more conducive to such preattentive visual pop-out.

Treisman also required subjects to find a target distinguished by the *lack* of a feature present in the distractor. She found that pop-out occurred when the target had the feature, and serial search occurred when the target lacked the feature. In our data analysis, attention was focused purposefully and serially in the activities of deliberate landmark search and serial search for secondary abnormalities which might or might not be in the image. Although Treisman's work emphasizes simple objects and features, it is possible that similar behavior may occur if someone has been trained to recognize particular objects and features in a manner that is meaningful to a problem solving task such as diagnostic radiology.

Expectation

Throughout the reports of the subjects, there was evidence that prior experience and medical knowledge were often used in conjunction with current observations to produce anticipation of particular kinds of related information from the image. These anticipations include presence or absence of features related to a finding, findings related to a diagnosis, and findings or diagnoses related to case history. These are called expectations, and Figure 9 shows examples of excerpts from the raw data of a number of subjects that lend support to this idea.

Expectations may be used by the problem solver to optimize plans for the gathering of information that will converge on a solution. For example, if a diagnostic hypothesis is cur-

rently active, it may be more efficient to explore findings and features that are usually expected to be associated with that hypothesis first, rather than just gather unstructured perceptual information in the hope that some of it may be useful. Expectation-driven exploration is the hallmark of top-down processing. When observations match expectations, confidence in the originating hypothesis should increase. On the other hand, when observations fail to meet expectations, a decision must be made as to whether this information can be overlooked, or whether it signals that the hypothesis should be abandoned. In the task of diagnostic radiology, it appears that expectations are largely perceptual in nature, especially in the absence of other information about the patient, such as test results, or physical examination. That is, most expectations have to do with anticipations about what can be seen in the image. In this sense, the expectations generated in this task appear to be dual in nature, in that they can originate with an abstract statement of intent, such as "I'm going to look for pleural effusions", that may be part of a larger plan to distinguish between some hypotheses, but they result in an act of looking: "I don't see any." Thus, expectation may be one of the mechanisms that bridges the gap between perception and problem solving.

Discussion

All of the above issues are closely coupled in the visual reasoning task of chest x-ray diagnosis. Context sets the scene for a particular collection of declarative and procedural knowledge to be retrieved from memory and brought to bear on the problem. This knowledge creates expectations of what the practitioner is likely to see, and plans to explore these expectations emerge, that then guide the attention process in deliberate search. However, there are often unexpected phenomena in the image, which seem to capture attention immediately, and cause currently active plans to be interrupted or abandoned in favor of new exploratory activity[2]. Descriptive features are used to characterize findings, which, in turn, are labelled at different levels of abstraction.

In the interplay between these issues, a pattern of interaction between perception and problem solving begins to emerge. Descriptive features can be said to lie closer to the perceptual side, while context seems to originate with more abstract thought related to problem solving. Expectations appear to lie between these two poles, originating with problem solving, but resulting in the activation of perceptual schemas through focus of attention, which direct acts of looking. These schemas allow perceptual information to be delivered back, and the levels of abstraction mentioned above provide a way to transform the information between expectation and perceptual schema, so that it can be used by the process concerned with achieving a solution to the problem.

These results may also provide a perceptual foundation for the diagnostic strategies employed by the radiologists, and an indication of why such strategies might succeed or fail. For example, in the case presented earlier, one of the subjects exhibited a very typical hypothetico-deductive strategy, where immediate visual capture provided initial data, some search of

---

[2]This is also consistent with Lesgold et al's (1988) contention that expert radiologists are opportunistic planners. However, we saw this kind of activity in less experienced radiologists as well.

the primary finding provided more evidence, and a diagnostic hypothesis was then produced. However, particular evidence was then obtained which contradicted the ensuing expectation, causing the original hypothesis to be abandoned in favor of a new (more accurate) one. Based on this new hypothesis, the subject then looked for supporting evidence, found it in the image, and thereby increased confidence. On the other hand, a second (more experienced) subject examined the same case, but spent most of the diagnostic period gathering perceptual evidence about the primary finding, and never really generated a diagnostic hypothesis at all.

A third example involved a case with an ambiguous primary finding [3]. For this case, some subjects who generated diagnostic hypotheses based on the incorrect finding, never did recover from the original error, while other subjects who appeared to engage in a more deliberate prolonged collection of perceptual information (not specifically related to a diagnostic hypothesis), eventually "saw" a clue, which led to the correct labelling of the finding. This case is interesting from the point of view that there is no really accurate *diagnosis* associated with the x-ray. However, the correct labelling of the finding is the best solution that can be attained, and is important to direct the next step in the patient care. It also suggests that a "not-enough-information" strategy might have been more appropriate in this case, rather than the hypothetico-deductive reasoning which generated early, but inaccurate diagnostic hypotheses.

In the absence of immediate visual capture, deliberate search of landmarks appears to be the common strategy employed. It is often repeated at least once if the patient appears to have a normal chest. It may also be a useful strategy when there is ambiguity in the x-ray image that might be removed by a more thorough investigation of the landmarks. For example, in the previously described ambiguous case, the visual clue concerned a part of the lung which was hard to see, but visible nonetheless. The subjects who looked specifically at that landmark area saw the clue.

Conclusion

The results from this work have been incorporated into a model of visual interaction between perception and problem solving, described in (Rogers, 1995a). This model has been further used as the basis for the design of a blackboard-based computer system called VIA (Visual Interaction Assistant), which incorporates the user, the image display and the program modules into a cooperative, problem solving system. The design has been instantiated into a prototype system for diagnostic radiology called VIA-RAD, and was tested in a small observational study with radiologist subjects (Rogers, 1995b). It is felt that the promising results of this work are due largely to the in-depth study of how practitioners actually perform their task. This approach provides not only further insight into visual diagnostic reasoning but also establishes a firm cognitive foundation for the development of intelligent computerized assistants.

---

[3]The subjects commonly mislabelled the finding as an elevated diaphragm, when it was really a mass inside the lung.

References

Evans, D. A. & 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* (pp. 211–255). Cambridge, MA: MIT Press.

Kundel, H. & Nodine, C. (1975). Interpreting chest radiographs without visual search. *Radiology*, 116, 527–532.

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

Lesgold, A., Feltovich, P. F., Glaser, R., & Wang, Y. (1981). The acquisition of perceptual diagnostic skill in radiology. Tech. Rep. PDS-1, LRDC.

McNeill, K. M., Seeley, G. W., Maloney, K., Fajardo, L., & Kozik, M. (1988). Comparison of digital workstations and conventional reading for evaluation of user interfaces in digital radiology. In *SPIE Medical Imaging II*, 914, 872–876.

Rogers, E. (1992). *Visual Interaction: A Link Between Perception and Problem-Solving*. (Tech. Rep. GIT-CC-92/59). Doctoral dissertation. Atlanta: Georgia Institute of Technology, College of Computing.

Rogers, E. (1995a). A cognitive theory of visual interaction. In B. Chandrasekaran, J. Glasgow, & N. H. Narayanan (Eds.), *Diagrammatic Reasoning: Computational and Cognitive Perspectives* (pp. 481–500). Menlo Park, CA: AAAI/MIT Press.

Rogers, E. (1995b). Via-rad: A blackboard-based system for diagnostic radiology. *Artificial Intelligence in Medicine*, 7, 343–360.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology, Human Perceptual Performance*, 8(2), 194–214.