A Cognitive Theory of Visual Interaction

Erika Rogers
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280

Introduction

The impact of visual computer displays on different types of scientific problem solving and the improvement of human performance in these tasks form the unifying theme of the research to be presented. One goal of this research is to design a computer system which allows us to explore human performance in a visual reasoning task, and ultimately, which facilitates and stimulates human reasoning capabilities by providing intelligent assistance. Such intelligence consists of knowing what type of assistance is needed, and when it may be cognitively effective to afford it.

Since it is important to develop an understanding of human cognitive processes such as perception and problem solving before we can hope to assist those processes, our general approach has been the following: 1) use cognitive science techniques to collect and analyze experimental data in a particular task domain; 2) utilize experimental results in the development of a cognitive model of the process under consideration; and 3) embed this model in a computer system designed to probe and test the model in order to refine it.

This paper is meant to provide an overview of the research progress to date on this topic. The collection and analysis of experimental data have been completed, and a first version of the Cognitive Model of Visual Interaction is presented. The details of this work can be found in [11, 12, 13]. Current work in progress is the design and implementation of the intelligent computer system, and our final effort will involve the testing and evaluation of the system, once again using subjects from our task domain.

Experimental Data Collection and Analysis

Our preliminary work has focused on the task of diagnosing chest x-rays, and is linked with a joint research project between Georgia Institute of Technology and Emory University School of Medicine. A number of subjects representing a range of experience were recruited from the radiology program at Emory. The collection of data was organized into several phases, and included informal observations of radiologists during their routine evaluations in the chest x-ray reading room, a transitional phase which compared performance between film and computer media, and finally, an extensive phase of talking-aloud reporting using computer-displayed images only.

The raw data for this part of the study consisted of videotapes and transcripts of each subject's experimental session. A three-part methodology for analyzing the data was developed and appears in Figure 1. The first step, called Task Analysis, began with a simple information-processing characterization of the task. This was expanded into a more detailed task taxonomy that combined results from Lesgold's work with radiologists [3], initial experimental results from our work [13], and more general problem solving results from the related literature (e.g., [1, 8]). This taxonomy provided an outline of the types of general concepts needed to encode the verbal protocols. However, examination of a subset of the actual protocols revealed that a greater level of detail was needed in order to account for the different types of statements made by the subjects.

In order to develop an encoding scheme that might be generalizable, three concept classes were identified and elaborated: Medical Concepts, Cognitive Concepts, and Descriptive Concepts. The Medical Concepts consisted of Anatomy, Finding, and Diagnosis categories, and were designed to be used as a preliminary encoding of the protocol statements which would extract specific medical terminology. Thus the concepts identified as Descriptive or Cognitive could be considered relatively domain-independent, and are illustrated in Figure 2.

This complete encoding scheme was then applied to the actual talking aloud reports produced by the subjects in the Full Protocol Analysis. Each protocol was parsed into simple phrases or statements, and then the words that could be classified as either Anatomy, Finding, or Diagnosis were extracted and listed in the appropriate columns next to the statement. A preliminary reliability study was conducted with an independent observer, and the remaining protocols were then encoded according to the concepts described above.

In the Contextual Analysis phase, the encoded data were re-examined with a view to discovering larger units of behavior over time, incorporating both perceptual and problem solving components. Three types of context were identified: task-related, time-related and experience-related, and initial consideration was given primarily to the first. Task-related types of activities seemed to fall into the following categories: 1) those related to the primary abnormality, 2) those related to secondary abnormalities, and 3) those related to remaining anatomical
objects and general comments. Under these three headings, each subject's actions were listed in terms of the concept codes, in the same order as they appeared in the verbal report. An example of this is demonstrated in Figure 3.

**Overview of Results**

A number of interesting results emerged from the Contextual Analysis which were found to be common to many of the subjects:

- A number of features are associated with anatomical objects and abnormalities (e.g., size, shape, edges and texture are used to describe a mass). This provides some indications of the types of perceptual information which may be important, and the dimensionality of some of the features involved.

- Two focus-of-attention mechanisms are evident: the first is called "Immediate Visual Capture" and describes the situation where certain findings in the image attract immediate attention of the subject; the second, called "Deliberate Landmark Search" occurred when subjects methodically examined landmarks in the chest and classified them as "normal" or not. This latter appeared to be a plan or script-like activity, where classification was fairly rapid, and the components were common to most subjects.

- Results that indicated integration of perception and problem solving included a bottom-up type of behavior, where descriptive features were used as evidence to trigger candidate hypotheses, and a converse top-down activity of hypothesis-controlled feature acquisition (e.g., "I think it's cancer, therefore I'll look for bony abnormalities").

- Use of non-perceptual knowledge such as case-history, previous experience, textbook knowledge and prototype solutions did occur, but not as frequently as might have been expected.
In the next section, these experimental results are combined with a theoretical framework to formulate the initial cognitive model of visual interaction.

Description of Model
In the course of normal human activities, it can be shown that there must be a relationship between perception and problem solving such that perception “delivers” information about the environment to the problem solving process, and, conversely, the problem solving process communicates “directions” to the perceptual process (e.g., I need this type of information rather than that type). Moreover, we are interested in the class of problems that requires extensive interaction between perception and problem solving, where the problem input is in visual format and the task is to interpret this input in a meaningful way.

Based on current models of perception (e.g., [2, 5]) and problem solving (e.g., [4, 7]), the mechanisms for such two-way communication are already potentially in place, and are conceptually illustrated in Figure 4. Constraints placed on the components labelled Model and Plans are critical to our understanding of this interaction. For example, the Model should be able to accommodate knowledge from both sides: visual information delivered by the perceptual process (e.g., percepts that describe findings in the image), and decision-related knowledge based on the current state of the problem solving process (e.g., what hypotheses are active, what kinds of information do they need for evidence, etc.). In addition, there should be a way to account for different levels of Plans. For example, a plan to pursue hypothesis-directed search versus data-driven search is at a different level of abstraction than the detailed plan for gathering the specific perceptual evidence required by a particular hypothesis.

We therefore hypothesize that there is a mediating process, called the Visual Interaction Process, which oversees the transfer of information and instructions between the perceptual cycle and the problem solver (which will henceforth be called the Perceptual Process and the Problem Solving Process, respectively). This mediating process is primarily responsible for maintaining and managing the information flow that passes through Working Memory during the problem solving task in a number of ways, including:

- retrieving associated semantic and hypothesis information from Long Term Memory;
- organizing perceptual information into evidence for/against current candidate hypotheses;
- providing instructions to the focus-of-attention mechanism regarding what to look at or look for in order to satisfy expectations.

Thus the cognitive model consists of three major processes and the knowledge in both Working Memory and Long Term Memory on which they operate.
Long Term Memory is characterized as containing not only general knowledge about the world, but in particular, domain-specific knowledge that is relevant to the problem solving task. This knowledge persists over long periods of time (e.g., the career of an active radiologist), and includes information about anatomical landmarks, types of image abnormalities (called “Findings”) and their associated features (e.g., size, shape, texture), diseases and their associated findings, previous theoretical knowledge and previous experiences.

On the other hand, Working Memory is characterized as a mental workspace that is easily accessible, holds current information about the task (in particular, the “prototype” or “Mental Model” acquired from Long Term Memory), and is quickly updated as more or new information becomes available. It is divided into three main components, namely, a Perceptual Buffer, a Problem-Solving Buffer, and an Internal Representation or Mental Model. The entire model is illustrated in Figure 5, and a more detailed description follows.
Buffers  In describing the requirements for a cognitive architecture, Newell et al [6] discuss the need for “interfaces that connect the sensory and motor devices to the symbol system”. They state that “the external world and internal symbolic world proceed asynchronously and therefore there must be a buffering of information between the two in both directions”. Thus the structures in Working Memory are grouped together to reflect their relationship to the three processes of the model: the Perceptual Process and the Visual Interaction Process communicate via the Perceptual Buffer, and likewise, the Visual Interaction Process and the Problem Solving Process communicate via the Problem Solving Buffer. Each Buffer contains both “data” components (i.e., Percepts and Hypothesis Report, respectively) and “control” components (i.e., Detailed Plan and High Level Plan, respectively), and allow asynchronous communication between processes. The following is a description of each of these components together with that of the Mental Model.

Percepts  We recall from our data the evidence for “immediate visual capture”. This early information may be used as an initial cue to obtain relevant domain-specific information from Long Term Memory. This is supported by findings in the literature as well. For example, Reiman and Chi mention the use of such a cue to “trigger a particular [problem] schema” [10], and, more generally, Newell et al [6] point out the need for access to distal memory structures. Later, as more information is added, or different objects (such as findings or landmarks) capture attention, there is also a need for a structure which allows transfer of this information from the Perceptual Process to the Visual Interaction Process. We say that this structure consists of one or more Percepts which constitute the output of perceptual processing, and can be described as labelled objects with some number of features (e.g., size, shape) attached. The purpose of this structure is to supply the perceptual cue(s) used by the Visual Interaction Process to initialize and update the current information that is maintained in the Mental Model during the problem solving task.

Hypothesis Report  The process of making a decision and completing the task frequently involves the evaluation of the current status of the candidate hypotheses. Frequently the subjects generate a list of candidate hypotheses, and rank them at various times in the problem solving task. Sometimes an initial ranking is expressed, which might later change according to the evidence gathered. At some point the subjects decide whether they have acquired enough information to make a final decision. This decision may be of the form of a definite, single diagnosis, a ranked list of differential diagnoses, an unranked list of differential diagnoses, or simply a list of findings. This type of reasoning does not seem to involve all the information currently gathered in the Mental Model, but rather appears to focus on the hypothesis-related components. This suggests that from the above Mental Model is extracted a substructure that contains a summary or report of the current status of the candidate hypotheses. This report reflects the strength of the evidence available at that stage, thus allowing the Problem Solving Process to produce a ranking, and to make a decision about whether to continue seeking further evidence, or to stop. We therefore call this the Hypothesis Report, which is delivered by the Visual Interaction Process to the Problem Solving Process.

High Level Plan  The Problem Solving Process, having received the Hypothesis Report from the Visual Interaction Process, must evaluate this information and determine whether there is enough to make a commitment to a decision. If not, then it must communicate directions to acquire whatever further information may be necessary to get closer to the goal. The Problem Solver therefore formulates a High-Level Plan that reflects the next strategy to be used. For example, if the hypotheses can be ranked, then the plan may indicate that evidence for the first-ranked hypothesis should be acquired first, followed by that for the second, etc. On the other hand, if the hypotheses cannot be ranked due to insufficient information, the plan may indicate that more features of the primary finding should be collected, and may order some of these features according to their ability to distinguish between hypotheses. It is also possible that the Problem Solving Process may “discover” a new hypothesis (or set of hypotheses) that it wants to explore. This new requirement is reflected in the plan, together with instructions on whether to completely abandon the old hypotheses, or perhaps to maintain them in a lower priority.

Detailed Plan  When the Visual Interaction Process receives the High-Level Plan from the Problem Solving Process, it first uses the plan to reorganize the Mental Model to reflect the current priorities - either rankings of hypotheses, or features, or perhaps the addition or deletion of hypotheses. Then, in order to meet the goals of the High-Level Plan, detailed instructions must be passed on to the Perceptual Process. This is done in the form of a Detailed Plan, which is formulated by the Visual Interaction Process using the information in the Mental Model. This plan may contain goals related to particular features of a finding (e.g., size, shape, texture, etc.), different types of findings (e.g., fluid, collapse, etc.) or different landmarks (e.g., heart, bones, etc.), and directions such as "LOOK-AT(lungs)", "LOOK-FOR(fluid)", etc. This is in keeping with the earlier description of the perceptual cycle, which uses the model to formulate plans for action, which then interact with the outside world.
Mental Model  Another observation that results from the data is that during the problem solving task, the types of knowledge available to the subjects include: a candidate list of descriptive features associated with the finding, candidate anatomical locations, candidate diagnostic hypotheses, candidate secondary findings, and remaining anatomical landmarks. Descriptive features (or combinations of features) and anatomical locations are utilized as evidence for particular diagnostic hypotheses. On the other hand, sometimes the hypotheses are directly used to control the acquisition of particular feature/finding information. This implies a tight coupling between the perceptual type of information and the candidate hypotheses. Our data shows that both bottom-up and top-down processing occur at various times in the subjects’ protocols, and furthermore, that precedence can alternate (e.g., sometimes combinations of features can override incorrect hypotheses). This is consistent with the results of Reiman and Chi [10] and Patel et al [9].

Therefore, the Mental Model (or Internal Representation) maintained in working memory contains a combination of relevant finding and hypothesis information, and reflects the relationship between current finding/feature information, and the evidence obtained and/or needed to distinguish between current candidate hypotheses. It is the task of the Visual Interaction Process to initialize and update this working representation based on the Percepts delivered by the Perceptual Process, and also on information dictated by the Problem Solving Process.

Computational System Design
Work is currently underway to design and implement an intelligent computer system which will be used to test and validate aspects of the model. The control and data flow of the computer system act as overlays to the cognitive model of the user, allowing us to query the user for specific information, and then injecting alterations such as image enhancements (on the perceptual side), or plan-like instructions (on the problem solving side). A trace of the user actions facilitates the evaluation of the validity of the model’s predictions - not only how visual interaction works, but also how task performance can be affected by appropriate (or inappropriate) automatic assistance.

References