Modeling the Visual Search of Displays:
A Revised ACT-R/PM Model of Icon Search
Based on Eye Tracking Data.

Michael D. Fleetwood and Michael D. Byrne
Rice University

Please address correspondence to:

Michael Fleetwood
Rice University
Department of Psychology
6100 Main Street MS-25
Houston, TX 77005 USA
(713) 348-2140
fleet@rice.edu

Abstract

Because of the visual nature of computer use, researchers and designers of computer systems would like to gain some insight into the visual search strategies of computer users. Icons, a common component of graphical user interfaces, serve as the focus for a set of studies aimed at (1) developing a detailed understanding of how people search for an icon in a typically crowded screen of other icons that vary in similarity to the target, and (2) building a cognitively plausible model that simulates the processes inferred in the human search process. An eye-tracking study of the task showed that participants rarely refixated icons that they had previously examined, and that participants used an efficient search strategy of examining distractor icons nearest to their current point of gaze. These findings were integrated into an ACT-R model of the task using EMMA and a "nearest" strategy. The model fit the response time data of participants as well as a previous model of the task, but was a much better fit to the eye movement data.

1. Introduction

In graphical user interfaces, icons (small graphical images, often accompanied by a label, which represent a file or command) are becoming increasingly prevalent. Still common on desktop computers, the technology is popping up in a variety of new locations, including mobile telephones, automobile navigation systems, kiosks, handheld computers, etc. An understanding of how users search for icons would be of use to system designers and researchers. Further, the evaluation of future systems would benefit from a model capable of making *a priori* predictions of user performance in icon-based displays. The research presented here focuses on two issues: (1) developing a detailed understanding of how people search for an icon in a typically crowded screen of other icons that vary in similarity to the target, and (2) building a cognitively plausible model that simulates the processes inferred in the human search process.

Much of the early success of cognitive engineering in the HCI field was in examining the efficacy of different designs by using cognitive models to predict task performance times (e.g. Gray, John, & Atwood, 1993). In this respect, laboratory research and industry benefited from the Model Human Processor, the Keystroke Level Model, and the GOMS family of techniques (Card, Moran, & Newell, 1983; John & Kieras, 1996). One deficiency of such models has been their inability to take into account the triad of elements involved in a human-computer interaction (HCI) task. As noted by Gray and Altmann (2001) the study of human-computer interaction should ideally include the study of a triad of elements—the user, the task at hand, and the artifact employed in the task. To this end, the HCI field has seen the development of modeling architectures capable of incorporating the complete triad of elements. EPIC (Kieras &

Meyer, 1997) and ACT-R/PM (Byrne & Anderson, 1998; now integrated into ACT-R 5.0; Anderson, Bothell, Byrne, Douglass et al., in press) were developed to include the cognitive, perceptual, and motor aspects of the user as they interact with the task environment. Additional strides have been made in allowing the modeling architectures to interact with the same software as human users (Ritter, Baxter, Jones, & Young, 2000), further integrating the task and artifact elements of the triad with cognitive models.

Now that all three elements of the triad can be studied in the context of cognitive modeling, we must ensure that the models interact with the environment in a human-like way. This has long been a barrier to the acceptance of cognitive engineering techniques by the wider HCI community. The traditional measures of response time and accuracy, though valuable, are only two metrics of interaction between human and computer.

The studies presented here are aimed at ultimately enabling the development of a simulated human user (Ritter et al., 2000; Young, Green, & Simon, 1989) capable of interacting with graphical user interfaces in a cognitively plausible manner. Specifically, the focus is on how the design of interface objects, icons in this case, affect low-level processes governing visual attention, which in turn affect what are typically considered to be higher-level processes of search strategy. By choosing a relatively complex visual environment to study, we hope to bring to bear some of the established research in the field of visual search on an HCI problem and also on the modeling architecture employed, ACT-R 5.0.

This paper is divided into six sections. First, we discuss some of the relevant research in visual search as it relates to an HCI context. Second, we will provide a brief overview of the ACT-R 5.0 cognitive architecture, giving particular weight to the aspects of the system involved in simulating human vision. Third, we will describe the general methodology used in our experiments. Section four provides a brief summary of a previous set of experiments and ACT-R models of the task. Section five presents an eye-tracking study of the task, and section six describes a new model based on the results of the eye-tracking study.

1.1. Relevant Visual Search Literature

The typical graphical user interface represents a complex visual environment relative to what has typically been examined in visual search studies. Nonetheless, many of the basic findings in human visual search are applicable to the domain.

1.1.1. Paradigm

In a standard visual search experiment, the observer is looking for a target item in a display containing some number of distracting items (Wolfe, 2000). Participants are typically asked to determine if a target object is present or absent on the display. Efficiency of a visual search can be assessed by looking at changes in response time (RT) or accuracy as a function of changes in the "set size," the number of items in the display. The search paradigm is valuable because performance on these tasks varies in a systematic manner with the nature of search stimuli. For example, search for a red object among a set of green objects is fast and accurate regardless of the number of green objects. The slope of the RT x Set Size function will be near zero. For other tasks, in which the target is not so easy to discriminate from the distractors, RT is roughly a linear function of set size.

This paradigm is attractive in the context of studying icons in part because it brings a common human-computer interaction (HCI) experience into the laboratory (McDougall, De Bruijn, & Curry, 2000). Computer users frequently must search for a desired object in a GUI. Examples might include the search for a particular icon on a toolbar or formatting palette, locating a button in a web page or application, finding an icon representing a command in a menu, or the search for a particular file or application in a directory containing other files or applications.

1.1.2. Preattentive Search Effects

In an efficient search, such as a search for a red item among green items, the subjective experience is that the target effectively pops out from its surroundings. Searches where targets can be distinguished from distractors on the basis of a single, basic feature, such as color, motion, or orientation are characterized as efficient or parallel searches. These efficient searches are also known as preattentive searches, as the information gleaned from the environment before visual attention has been directed to areas in the visual scene (preattentive information) is sufficient to guide search. As the target becomes less disriminable from the distractors, the search becomes less efficient and more serial in nature. On the inefficient end of the continuum, items in the distractor set must be examined individually in order to locate the target. An intermediate level of search efficiency would require only a subset of items to be examined in order to locate the target.

An important class of search tasks producing searches of intermediate efficiency is conjunction searches, where features and targets are only distinguishable on the basis of a conjunction of several different features. For example, in a search for a red “X” among green “X”s and red “O”s the target is only distinguishable by a conjunction of color and form. Neither color nor form alone defines the target. Conjunction searches were originally thought to lie towards the extremely inefficient end of searches, where all items that shared any features with the target must be examined in a serial, self-terminating fashion (Treisman & Gelade, 1980). It appears that this claim is too strong (Wolfe, 1994, 2000). As mentioned, studies have shown that search could be restricted to subsets of the items—subsets defined by features like color (Egeth, Virzi, & Garbart, 1984). Other studies showed that more than one feature at a time could contribute to the guidance of conjunction search (e.g. Alkhateeb, Morland, Ruddock, & Savage, 1990; McLeod, Driver, Dienes, & Crisp, 1991; Nakayama & Silverman, 1986; Treisman & Sato, 1990; Wolfe, 1992).

In searches where subsets of items, not just a single item, may be preattentively identified and selectively searched, the search time may be a function of the number of items in the subset. For instance, in a search for a green “T” amidst green “L” and red “T” distractors, the search may be a function of the number of green items on the display. In this case, the entire subset of green items can be selectively searched. The evidence supporting this assumption comes from a number of visual search studies (Carter, 1982; Green & Anderson, 1956; Smith, 1962) and a theoretical model of how such searches might occur, the Guided Search (GS) model (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989).

When a subset of items may be preattentively identified and selectively searched a pattern of results known as the distractor ratio effect is revealed (Bacon & Egeth, 1997;

Poisson & Wilkinson, 1992; Shen, Reingold, & Pomplun, 2000; Zohary & Hochstein, 1989). The distractor ratio effect describes when the ratio between different types of distractors in a conjunctive search task strongly influences the response times to detect a target item. For instance, participants were asked to decide whether a conjunctively defined target was present or absent among distractors sharing color or shape. When the total number of items presented in a display was kept constant, response times varied as a function of the ratio between the two types of distractors, those sharing color or those sharing shape with the target. More specifically, response was faster when either type of distractor was rare than when both types of distractors were equally represented. For example, if the target was a red **X**, response was fastest when the distractors were primarily composed of red **O**s or green **X**s and slowest when there was an equal amount of red **O**s and green **X**s. More explicitly, if the distractors were primarily green **X**s, participants could restrict their search to the red items in the display. In addition, the saccadic selectivity of participants was greatest at these extreme distractor ratios—i.e., participants' searches were guided by the feature of the target (color or shape) that was common with the fewest number of distractors. This indicates that detecting a conjunctively defined target does not necessarily require a serial item-by-item search. A serial item-by-item search would not produce the distractor ratio effect. Shen et al. (2000) found that the observed changes in RT due to the distractor ratio were echoed by eye-movement measures, such as the number of fixations per trial and latency to move.

The distractor ratio effect is predicted by the Guided Search (GS) model (Wolfe, 1994; Wolfe et al., 1989; Wolfe & Gancarz, 1996), which argues that information extracted preattentively can guide shifts of attention during the search process. According to the GS model, the preattentive information encompasses both bottom-up (extrinsic, driven by the environment) and top-down (intrinsic, driven by the perceiver) activations. These sources of information are combined to form an "activation map," which contains peaks of activity at likely target locations (Wolfe, 1994). The focus of attention is directed serially to the locations with highest activation until the target is found or the criterion to make a negative response is reached. When participants are allowed to move their eyes, a "saccade map" is similarly created to guide the movements of the eyes (Wolfe & Gancarz, 1996). Every 200–250 ms the eyes are moved to the point of highest activation in the saccade map.

The guided search model is based on research in a number of controlled laboratory visual search experiments. However, although similar in many respects, the visual environment people normally interact with in HCI contexts is more complex. There is a long history in HCI of extending well-researched paradigms and theories to slightly more complex environments in an effort to generalize the theories and extend their domains. Thus the theories developed under carefully controlled conditions are incrementally evaluated in slightly more complex task environments. The research presented here extends some of the predictions of the guided search model, specifically that of being able to locate multiple stimuli preattentively, to a slightly more complex environment that is closer to the environment experienced by everyday computer users.

Research on icons confirms that the aforementioned research on visual search applies to more complex environments and stimuli. McDougall, de Bruijn, and Curry (2000) found three characteristics to be of primary importance in the measurement of symbols and icons: concreteness, distinctiveness, and complexity. Of the three, distinctiveness and complexity are most relevant to the visual search literature. Distinctiveness cannot be assessed in isolation; it is contingent upon the nature of the visual display in which an icon is located. Generally, the more distinct an icon from its surroundings the quicker it can be located. With regard to complexity, the amount of detail or intricacy within an icon was found to influence the rate at which it could be detected, with very simple or very abstract icons being detected faster. With respect to locating icons on a computer display, Byrne (1993) found that when users were asked to carry out a search task, they were able to locate simple icons faster than more complex icons. This was attributed to the concept that simple icons were discriminable on the basis of only a few features relative to more complex icons, and this ease of discriminability aided users in their search.

1.2. ACT-R 5.0

A cognitive architecture is both a theory of human cognition and performance and a framework for developing computational models of behavior. Cognitive architectures have been used widely to model human behavior (Anderson & Lebiere, 1998; Newell, 1990) and more specifically, human-computer interaction (e.g. Altmann, 2001; Anderson, Matessa, & Lebiere, 1997; Byrne, 2001; Kieras, Wood, & Meyer, 1997; Kitajima & Polson, 1997; Ritter et al., 2000). Cognitive architectures provide at least two major benefits for the purposes of the proposed approach. First, architectures incorporate well-tested parameters and constraints on cognitive and perceptual-motor processes, and any model developed in an architecture necessarily inherits these parameters and constraints. This allows architectural models to generate *a priori* predictions about behavior and performance (Salvucci & Macuga, 2002). Second, these predictions are inherent in the model, yet separate from the modeler. The advantage here is that any analyst can run the model with the same outcome (Gray & Altman, 2001). Thus, the model is not limited to a particular researcher or project.

The system that was used to model the experiments was ACT-R 5.0. The ACT-R architecture has been used to successfully model a variety of behavioral phenomena and has proven particularly successful at modeling tasks with a demanding cognitive component (Anderson & Lebiere, 1998). (A thorough description and discussion of the ACT-R framework is given in Anderson and Lebiere, 1998). In ACT-R 5.0 (Anderson, Bothell, Byrne, Douglass et al., in press), among other changes to the architecture, the original system has been combined with modules for perceptual and motor actions (vision, audition, motor, and speech; see Chap 6 of Byrne & Anderson, 1998 for a discussion of the functioning of the different modules; also Byrne, 2001 and Byrne & Anderson, 2001). Because icon search is relatively light on the cognitive demands of the user, it is a task that must be modeled using an architecture that accounts for the perceptual and motor components inherent in the task—i.e. directing visual attention in a relatively complex visual scene. Other researchers have employed cognitive architectures, including ACT-R (Byrne, 2001) and EPIC (Hornof, 2001; Hornof & Kieras, 1997) as part of their investigation of the visual search of menus. This research extends the methodology to a more complex visual environment.

1.2.1. ACT-R System Configuration

ACT-R is a computational theory of cognition and human performance. The system is organized as a set of modules that interact with two types of memory, procedural memory and declarative memory (see Figure 1). The declarative memory contains chunks of things remembered or perceived. These chunks can be facts like "2 + 1 = 3", intentions or goals, or as is the case of the icon search models presented here, a collection of information about the visual environment. There is also a production (or procedural) memory that contains the procedures and skills necessary to achieve a given goal. The units of procedural memory are production rules, IF-THEN condition-action mappings that "fire" when the conditions are satisfied and execute the specified actions. The conditions are matched against a set of buffers whose content is determined by a series of modules. The perceptual-motor system is made up of modules that handle various aspects of perception (visual and auditory) and action (motor and speech). There is also a module devoted to retrieving information from declarative memory.

Communication between central cognition and the modules takes two forms. Each module has one or two buffers that may contain one chunk. The production system can (1) recognize patterns in these buffers and (2) indirectly make changes to these buffers—by requesting that the module perform an action, such as shifting visual attention, making a key press, or requesting the retrieval of a chunk from declarative memory.

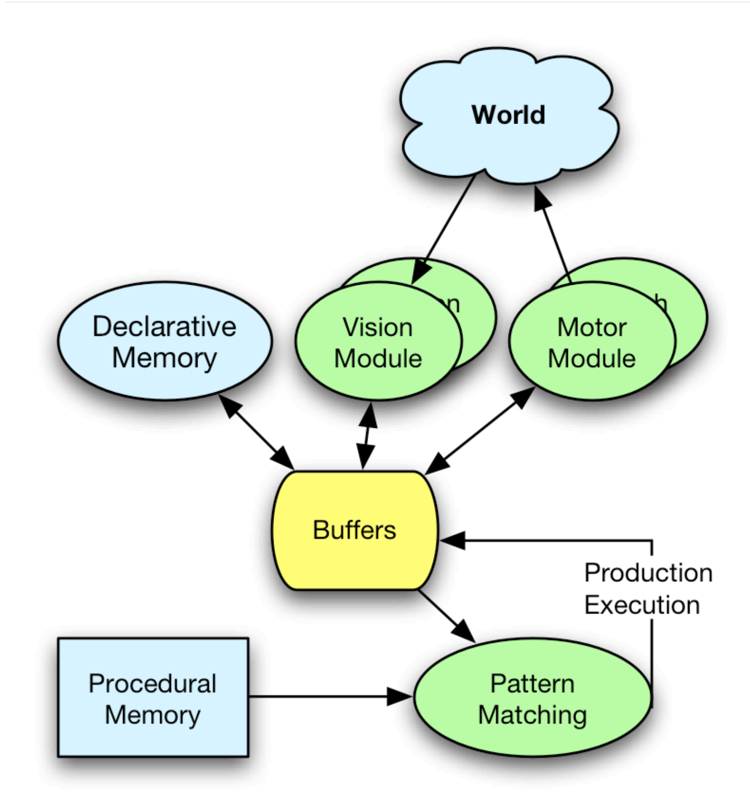


Figure 1. ACT-R 5.0 system diagram.

The basic computational increment is the production cycle, which consists in matching productions against memory, selecting a production to fire and then executing the THEN side of the selected production. Production rules' IF sides are matched against the contents of declarative memory and the contents of the buffers. One of the productions that has its conditions matched is selected to fire. Only one production rule may fire per cycle.

1.2.2. The Vision Module

Given the visual nature of graphical user interfaces, the Vision Module is of key importance in modeling many HCI tasks. As one might expect, the Vision Module is used to determine what ACT-R "sees." Each object on the display is represented by one or more features in the Vision Module. These features are simply a symbolic list of attribute pairs that represent the visual attributes of the objects on the display, such as "red circle." The modeler carries out the parsing of the display into objects and the creation of the list of attribute pairs representing each object. The Vision Module creates chunks from these features that provide declarative memory representations of the visual scene, which can then be matched by productions. The Vision Module is organized around two subsystems, a "where" system and a "what" system.

When a production makes a request of the "where" system, the production specifies a series of constraints, and the visual location buffer returns a chunk representing a location meeting those constraints. Constraints are attribute-value pairs, which can restrict the search based on visual properties of the object (such as "color: red") or the spatial location of the object (such as "screen-y greater-than 153"). This is akin to so-called "pre-attentive" visual processing (Treisman & Gelade, 1980) and supports visual pop-out effects. We take advantage of this capability in modeling the different levels of icon quality observed in the experimental data, particularly with respect to the good quality icons.

A request to the "what" system entails providing a chunk representing a visual location, which will cause the "what" system to shift visual attention to that location and process the object located there, i.e. deposit a chunk representing the visual object at that location into the visual object buffer. In the base system of ACT-R this shift of visual attention takes 135 ms of simulated time, 50 ms for a production to fire and 85 ms to make the shift of visual attention and process the object.

It is important to note that ACT-R 5.0 does not make any predictions regarding eye movements. The system may be used to predict shifts of visual attention, but it has been well-established that there is not a direct correspondence between unobservable attention shifts and observable eye movements (Henderson, 1992; Rayner, 1995).

1.2.3 The Motor Module

Other than visually locating objects, our models must also account for selecting the icons with the mouse. In executing a movement, it must first be prepared by the motor module (unless the movement is a replication of the prior movement, in which case there is no preparation time). The time to prepare the movement is at least 50 ms and ranges upward depending on the movement (button press or mouse movement). Once the movement has been prepared, the amount of time that a movement takes to execute depends on the type and possibly the size of the target object and distance the movement will traverse. Simple movements have a minimum execution time (also 50 ms, called the "burst time") and more complex movements (such as pointing with the mouse) have a longer execution time based on Fitts' Law.

2. General Procedures

The experiments presented here are a replication of a previous set of experiments reported in Fleetwood & Byrne (2002). The experimental paradigm is nearly identical for all of the experiments discussed in this paper. A general methods section is provided here, and any deviations from this general template are specifically noted in the discussion of the individual experiments.

2.1. Design

Three independent variables were manipulated, all of which were within-subject factors. The first of these factors, set size, had four levels, 6, 12, 18, or 24 icons. The second within-subjects factor, icon quality, had three levels. Icons were designed that varied in their levels of distinctiveness and complexity. On one end of the spectrum were icons of "good" quality. These icons were designed to be easily distinguishable from other icons based on the primitive visual (i.e., preattentively discriminable) features of color and shape (specifically curvature). Icons in the good quality set were one of six colors (red, blue, green, yellow, brown, or black) and one of two shapes (circle or triangle). Examples are shown in Figure 2. On the other end of the quality spectrum were icons that were not easily distinguishable (referred to as "poor" quality icons). They were designed to be discriminable through relatively careful inspection when explicitly paired, but relatively indistinguishable in a large distractor set. These poor quality icons were all of the same basic shape and did not include color (other than white, black and shades of gray). The "fair" quality icons were designed to be representative of the area in between these two ends of the spectrum. They were generally of a distinct shape, although more complex than the simple circles and triangles in the good quality icons, and none of them contained any color outside of the spectrum of grayscale colors.

Icon quality, as it is defined here, encompasses a number of separate, dissociable attributes that contribute to the ability of participants to locate an icon quickly among a set of similar icons. The attributes identified by McDougall, Curry, & De Bruijn (1999) and McDougall, De Bruijn, & Curry (2000) that apply to the icons used here are distinctiveness and complexity. In order to quantify the levels of complexity and distinctiveness of our icons, a separate study was conducted in which 22 participants rated each of the icons on these two attributes. Regarding complexity, participants were asked to rate the amount of detail or intricacy of line in the image on a 5-point scale where 1 indicated "very simple" and 5 indicated "very complex". Regarding distinctiveness, participants were asked to rate each icon on the basis of how easily it could be distinguished from all the other icons in a group of icons (on a 5-point scale, 1 represented "not distinct" and 5 represented "very distinctive").

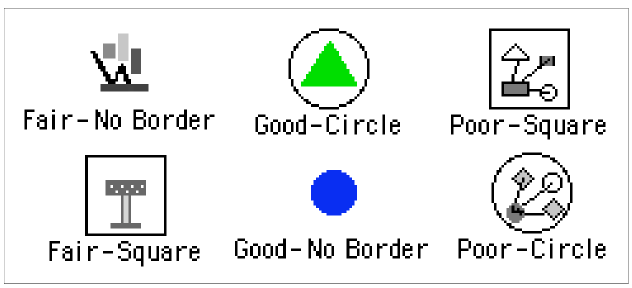


Figure 2. Examples of icons of good, fair, and poor quality used in the experiment. The good quality icons were each a single solid color, whereas the fair and poor quality icons were drawn in grayscale. In accordance with a separate line of research, the icons were shown with different types of borders; however, there was no evidence that the borders had any effect on search performance (Fleetwood, 2001).

Participant ratings corresponded with the authors' classification of the icons into three levels of quality. (The average ratings for each level of icon quality are presented in Table 1.) Participants rated the good quality icons as being the least complex and most distinctive. Poor quality icons were rated as being the most complex and least distinctive, Fair quality icons were rated as being relatively distinctbut of moderate complexity, Individual ratings for each icon are presented in Appendix A.

	Complexity	Distinctiveness
Good	1.1	3.9
Fair	3.2	3.9
Poor	4.1	1.9

Table 1. Mean ratings of complexity and distinctiveness for each level of icon quality.

Additionally, the level of complexity of each icon was calculated using an automated analysis program (Forsythe, Sheehy, & Sawey, 2003). Again, the relative mean automated complexity ratings correspond to the three levels of icon quality. Good quality icons had the lowest level of complexity, 51.5, followed by fair quality icons, 151, and poor quality icons, 187. The automated rating were highly correlated with the participant ratings, $r = 0.89$. (Automated ratings are based on the PerimeterX4 metric. See Forsythe, Sheehy, & Sawey, 2003, for more information.)

A final within-subjects factor, icon border, had three levels. The target icon to be searched for could be presented without a border, with a circle as a border, or with a box as a border. Refer to Figure 2 for examples of each border type. Several previous studies replicated the finding that the type of border did not affect search performance (Everett & Byrne, 2004; Fleetwood, 2001; Fleetwood & Byrne, 2002). The variable will not be considered further here and was only mentioned for the sake of completeness.

Each block in the experiment consisted of 36 trials. Each independent variable was examined at each level of the other independent variables (4 x 3 x 3 = 36). The order of presentation was randomized within each block.

The dependent variable being measured was the response time of the users—specifically, the time from when they clicked on a "Ready" button to indicate that they were finished examining the target icon to when they clicked on the target icon among the set of distractor icons.

One potential independent variable that was held constant was the number of icons matching the target in the search display. On each trial one-third of the icons in the search display had the same pictorial icon and matching border as the target icon (referred to as target-matching or TM icons). For example, in a set size of six icons, one icon would be the target, one icon would be a target-matching icon, and four icons would be non-target-matching icons. This was done in order to more closely approximate a “real world” task in which a user must differentiate among similar or identical icons (such as searching for a document in a directory with multiple documents created by the same word processing program). Ultimately the user was forced to differentiate among the icons by reading the file name.

2.1.1 Materials

The icons used in the experiment were standard sized icons (32 pixels by 32 pixels). Participants were seated approximately 20 inches from the computer screen (800 x 600 pixel resolution). At that distance each icon subtended 1.15 degrees of visual angle. The icons were separated by approximately 1.2 degrees of visual angle horizontally depending on the shape of the icon. Immediately below each icon was the filename corresponding to that icon. The distance from the bottom of a filename to the top of an icon below it subtended approximately 0.4 degrees of visual angle.

Twelve different icons were created to represent each level of icon quality, for a total of 36 distinct icons (3 levels of quality X 12 icons per level). (An image of each icon is provided in Appendix A.)

2.2. Procedures

Users were instructed on how to perform the task, then were given one block of practice trials to develop some familiarity with the task and with the mouse used to point and click on the target icon.

To begin each trial, participants were presented with a target icon and a corresponding file name. After 1500 milliseconds, a button labeled “Ready” appeared in the lower right corner of the screen. Participants would click the Ready button when they felt they had sufficiently examined the target icon and were ready to move on to the next stage of the trial.

Immediately after clicking on the Ready button, the participants were presented with a screen that contained a set of icons (6, 12, 18, or 24), one of which was the target icon. The user’s task was to identify the target icon and click on it as quickly as possible. Clicking on an icon brought them to the first stage of the succeeding trial—the presentation of a new target icon.

Response time was measured from the time they clicked on the “Ready” button to the time they clicked on an icon in the distractor set.

The icons for each trial were placed in a window in the upper left quadrant of the computer screen. The position of the icons within the window was fixed such that icons were placed in the same positions on each trial. For instance, in a set size of six icons, the six icons were always placed in the same locations on the screen, but which six icons were present varied from trial to trial. From the available positions for each trial, the position of target icon was randomly selected. Likewise, the positions of the remaining icons were determined randomly from those available for each trial.

Also randomly selected were the file names for the icons. The distractor file names and the target file names were randomly selected without replacement from a list of 750 names until the list was exhausted; at which time, the list was recycled. All of the file names were two or three syllable English words six to 10 letters in length.

Each user completed four blocks of trials in addition to the practice block for a total of 180 trials.

3. Computational Modeling of the Experiment

A model was constructed in ACT-R 5.0 that interacted with the same software as participants. This section discusses the relevant details of the original model presented in Fleetwood and Byrne (2002).

In the initial, or precue, stage of a trial, the model must remember the target icon and its corresponding filename. As mentioned, each icon is "seen" by ACT-R's Vision Module as a list of attribute pairs. (The complete list of attribute pairs representing each icon is reproduced in Appendix A.) For the "good" quality icons, a single attribute pair represents each icon (e.g. "red circle"). In contrast, more complex icons will have a number of attribute pairs associated with them, gray triangles, white circles, etc. What makes these more complex icons "poor" icons in the experiment is not the number of attribute pairs the icon has per se, but rather the number of attribute pairs the icon shares with other icons in the distractor set. For example, the set of attribute pairs representing many of the icons in the poor quality set include gray triangles and white circles. (See Appendix A for a list of the attribute pairs representing each icon.) The model only stores one attribute pair of the target icon in order to identify the target icon in the distractor set. (The attribute pair that the model stores is randomly selected from the list of attribute pairs representing the icon.) As a result, the model will often examine icons that do not match the target icon exactly, but rather only share one particular attribute pair with the target icon. It is this overlap of attributes, or "similarity," that makes such icons poor icons in this context. In contrast, the good quality icons have no attribute overlap with other good quality icons, and thus, the model examines only icons exactly matching the target icon. Hence, the efficiency of the model's search is a product of the simplicity of the target icon (number of attribute pairs representing it in ACT-R) and the relative similarity of the target to the distractors (number of other icons that are represented by at least one of the same attribute pairs).

The exact nature and number of the attribute pairs used to represent each icon in the "fair" and "poor" conditions are free parameters in the models; however, the set designed for the original models (Fleetwood & Byrne, 2002) was not altered for the current modeling effort.

In the initial, precue, stage of a trial, the model attends the target icon and selects at random one attribute pair (e.g. "gray rectangle") from the list of attribute pairs representing the target icon and stores this attribute pair. The filename is also noted and stored. The model uses the stored feature and filename to identify target-matching icons among the distractor set. Before moving on to the second stage of the trial, the search stage on the distractor screen, the model locates and clicks the "Ready" button. This series of events is completed in ACT-R through seven productions (2 to locate the target icon and store, "remember," an attribute pair, 3 to locate the filename and store it, and 2 to locate and click on the Ready button).

On the second stage of a trial, the model must find the target icon among the distractors. The search process is accomplished through four productions in ACT-R. First, a random icon is found which contains the feature of the target icon stored by the model in the initial stage of the trial (1 production). Next, visual attention is shifted to the filename below the newly located icon (2 productions). Finally, if the filename below the new icon matches the filename stored by the model, then visual attention is shifted up to the icon so that it can be clicked on with the mouse (1 production). If the filename does not match the target icon, then another icon with the stored feature is located and the search progresses. This sequence of events corresponds to 285 ms of simulated time (4 productions X 50ms each + 85ms for one shift of visual attention).

The model output is a trace of what actions took place and when they occurred. The simulated time for the model to complete a trial is a summation of the number of productions that fired (50ms per production), the number of shifts of visual attention (85ms each), and the motor movements made to point and click with the mouse (movement times are based on Fitts' law) subtracting for any instances when actions occurred in parallel. Only one production may fire at a time, but the different ACT-R modules (visual and motor, in this case) may operate in parallel. We also recorded any shifts of visual attention made by the model (when and where they occurred) for comparison with the eye-tracking data.

3.1. Results

The fit of the model to the data (reported in Fleetwood & Byrne, 2002) was quite good (See Figure 3). Most importantly, the model captured each of the pronounced effects that were seen in the data—those of set size and icon quality.

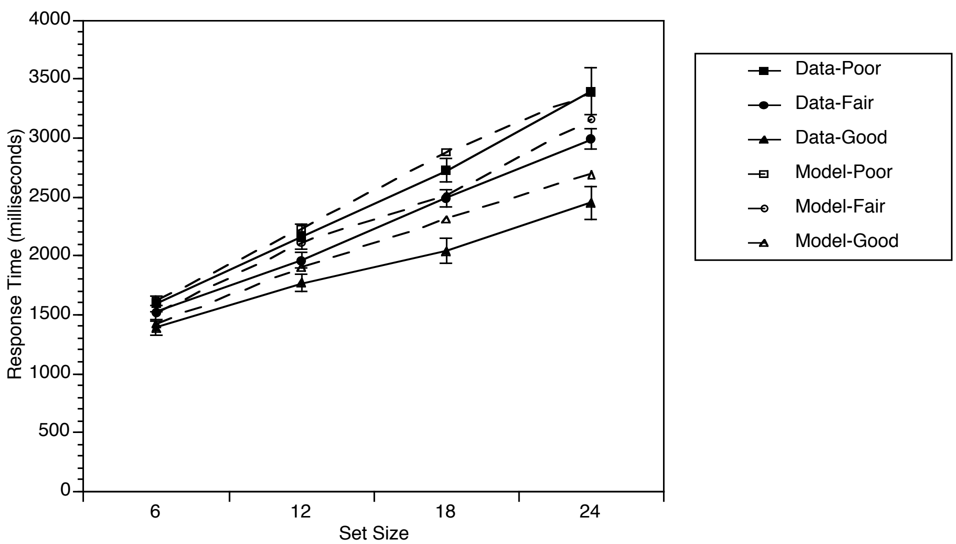


Figure 3. Comparison of model and experiment data. Reproduced from Fleetwood & Byrne (2002).

The proportion of variance explained (R^2) by the model relative to the data from the experiment was 0.98. The root mean square error (RMSE) and percent average absolute error (PAAE) between the model and the data were 126 ms and 4.27%, respectively. The quality of the fit suggests that the model does an excellent job of accounting for the major trends in the data.

Overall, we felt that the performance of the model compared to the experimental data was encouraging. However, the response time data only provides a single metric for comparison and does little to tell us if the model accomplished the task in a human-like manner. Numerous models have fit response time data well but have not necessarily fit other metrics of human performance. For example, Byrne et al. (1999) describes two models of human performance when using click-down menus. Both models fit the response time data well, but both models use different underlying strategies, neither of which were particularly good fits to participant eye movement data for the same task. Similarly, the original icon search model adequately fit the participant response time data, and the strategy employed by the model produces several predictions with respect to the visual search strategies of participants. In order to evaluate the eye movement predictions made by our model, we conducted an eye-tracking study of the task.

4. Eye Tracking the Icon Search Task

Researchers have used eye tracking to make fine distinctions regarding the processes used in a visual search task. For example, researchers were able to identify oculomotor distinctions between parallel and serial search processes (Zelinsky & Sheinberg, 1997) and develop models that account for visual performance effects, such as the “global” effect of the visual environment in making saccades (Becker & Jurgens, 1979; Findlay, 1982). Also, researchers have used eye tracking to gather information on the features of objects that drive visual search (Gould & Dill, 1969; Viviani & Swensson, 1982; L. G. Williams, 1966, 1967).

The use of eye-tracking has also made its way into studies of human-computer interaction and as a potentially applied procedure in the computer industry in the form of “gaze-based” interfaces (Salvucci & Anderson, 2000; Sibert & Jacob, 2000). On a different level, it has been used as a means of understanding the processes underlying the behavior of computer users (e.g. Byrne et al., 1999; Ehret, 2002; Hornof & Halverson, 2003; Jacko et al., 2001).

4.1. Model Predictions

The ACT-R model of icon search just described makes several predictions regarding the eye movement patterns of participants.

<u>Number of shifts of attention per trial:</u> The model predicts that the total number of shifts of visual attention per trial increases as set size increases and as icon quality decreases. Most specifically, in the good quality condition, the model only examines target-matching icons. The model evaluates a potential icon (one sharing an attribute pair with the target icon) in one shift of visual attention. It assumes that a target-matching icon can be located preattentively, i.e. exhibits the “pop-out” phenomena found in visual search. Hence, it predicts that the number icons examined by participants should approximate the average number of good quality icons that must be examined to locate the target icon. For example, in a set size of six icons, there are two target-matching icons; on average the model must examine 1.5 target-matching icons in order to locate the target icon.

<u>Number of shifts of visual attention to target-matching icons:</u> The model only shifts attention to icons sharing an attribute pair with the target icon. Because this attribute pair is unique in the good quality set, the model only examines target-matching icons in the good quality set. Target-matching (TM) icons were icons exactly matching the target icon, which was one-third of the icons in each distractor set. The model examines a decreasing proportion of TM icons as the quality of the icons decreases.

Hence, the model predicts that participants will examine a high proportion of TM icons in the good quality condition and that this proportion will decrease as quality decreases. However, even in the poor quality condition, participants should examine a higher proportion of TM icons than if fixations were randomly directed.

Search Strategy: The model examines any icon sharing the attribute pair that was selected in the precue stage of the trial and that has not yet been examined, but the order in which it examines them is random. Hence the model prediction is that participants will show a preference for examining TM icons but show no preference for the order in which the TM icons are examined.

Reexamination of Icons: The model predicts that participants will occasionally and non-systematically reexamine icons. This is consistent with visual search studies that show that people have little memory for where they have looked in a static visual scene (Horowitz & Wolfe, 1998). The model has no way of marking or remembering which icons it has previously attended. This is because the icons themselves are never actually attended, just the filenames below the icons, and ACT-R only “remembers” locations to which it has shifted attention. Hence, the model occasionally and randomly redirects attention to icons and filenames that it has previously examined. Because this revisitation is stochastic, analytic predictions are difficult to derive and Monte Carlo simulations are required to calculate the likelihood of revisitation.

4.2 Methods

4.2.1. Participants

The participants in the experiment were 10 undergraduate students at Rice University who were participating in order to meet a requirement for a psychology course. Although some variation with regard to computer experience was expected, users in this population are generally familiar with computer use.

4.2.2. Design and Procedure

The design and procedure of the experiment was identical to that described in the General Methods section with the addition of the eye tracker to record the participants' eye movements while engaged in the task.

4.2.3. Apparatus/Materials

The eye tracker used was an ISCAN RK726/RK520 HighRes Pupil/CR tracker with a Polhemus FASTRACK head tracker. Head-mounted optics and a sampling rate of 60 Hz were used in the experiment. This system, like many other laboratory eye trackers, works by shining an infrared light on the eye and taking a video image of the eye. From that image, it is possible to determine the pupil center and the point on the cornea closest to the camera (the corneal reflection) and take the vector between them. This vector changes as the eye orients to different positions on the screen, and with calibration to known points, it is possible to compute visual point of regard (POR, also referred to as "point of gaze"). The magnetic polhemus is used to compensate for head movements.

POR reports by this system of eye-tracking equipment are typically accurate to within one-half degree of visual angle.

POR and mouse position were recorded at 60 Hz by the experimental software. Stimulus and POR/mouse data for each trial were recorded so that all individual trials could be “replayed” at various speeds. An experimenter monitored each experimental trial and recalibrated the eye tracker if there appeared to be a sizable disparity between reasonable expectations about where users were looking (in particular, users needed to look at the target on each trial) and the position reported by the tracker.

4.2.4. Analysis Technique

From the raw data it is possible to compute where and when fixations occur. This can be done either by assuming that any eye position within a given region for more than some threshold number of milliseconds is a fixation (dwell-based) or assuming that any period of time showing relatively low velocity is a fixation (velocity-based). For the data set, both methods were initially used and examined to verify that they both yielded approximately the same result. For ease of calculation, the dwell-based method was used for further analyses.

For analyses in which a direct comparison of the eye tracking data and the ACT-R model was made, gazes were used as the metric of analysis in lieu of fixations. An uninterrupted series of subsequent fixations on a region of interest (an icon, in this case) was considered a gaze. Aggregating fixations into gazes on a region of interest is a commonly used technique in the analysis of eye tracking data (e.g. Rayner, K &

Pollatsek, A, 1998; Rayner, 1995). For analyses in which fixations were attributed to a particular icon, the fixation was attributed to the nearest icon.

Gazes, rather than fixations, were analyzed here in order to make a more direct comparison with the data output by the model, shifts of visual attention. As noted, ACT-R 5.0 describes patterns of visual attention, but does not explicitly predict eye movements or fixations. It is well established that visual attention guides eye movements, i.e., visual attention is shifted to locations in the visual field, and the eyes may or may not follow. Specifically, for any shift of visual attention, three possibilities may occur with respect to eye movements and fixations. For a given shift of visual attention, a saccade and a single fixation may be made to the new locus of visual attention. In this case, there would be a one-to-one correspondence between fixations, gazes, and shifts of visual attention. A second possibility occurs when a shift of visual attention is followed by several fixations towards or on the region of interest before a new shift of visual attention is made. In this case, in order to get a direct correspondence between the number of shifts of visual attention and the number of fixations on a region, one would collapse the number of fixations into a single gaze. A third possibility occurs when multiple shifts of visual attention occur before any eye movements are made. In this case, there is no eye movement data that may be compared to visual attention predictions making any sort of analysis regarding the two metrics quite difficult.

Our model made several predictions regarding the visual attention shift patterns of participants. By collapsing the fixation data from the eye tracking study into gazes, we were able to directly compare the predications from our model to the data from the eye tracking study. We were able to account for two of the three aforementioned conditions, when there is an equal or greater number of fixations relative to shifts of visual attention. Accounting for the condition when there are a greater number of visual attention shifts than fixations would have been impossible given the methodology employed.

The disadvantage of collapsing fixations into gazes for the purposes of analysis is that some level of precision in the data is lost. Hence, whenever we were interested in analyzing the eye movement patterns of participants, but were not making comparisons between the model and the eye movement data, fixations are used as the metric of analysis.

4.3. Results

When the user did not correctly identify the target icon, the trial was considered an error and removed. Outliers were also removed when the response time was more than three standard deviations from the 5% trimmed mean of the user for the corresponding set size. In total, less than 2% of the trials were removed due to errors and outliers. For statistical tests, where response times had been removed as errors or outliers, they were replaced with the individual participant's overall mean. Trials on which there were no fixations on the region of the screen containing icons were removed from the analysis (6.6%). The removal of a relatively high percentage of trials from the analysis was due to an equipment problem during data collection in which a poor calibration of the equipment to the individual participant was obtained. The equipment problem was not systematic in nature and was corrected before the following trial was presented.

Over the course of all of the trials, the average duration of fixations was 291 ms. Across all trials, participants made approximately 11.1 fixations and 3.3 gazes per trial.

The response times in the eye-tracking study (presented in Figure 4) corresponded with those from the previous study. As icon quality decreases (good to fair to poor), response times increase, $F(2, 18) = 58.71$, $p < 0.001$. Also, as set size increases, response times increase, $F(3, 27) = 71.89$, $p < 0.001$. Finally, an interaction between set size and icon quality indicated that response time increased proportionately more as set size increased for poor icons than for good quality icons (F 6, 54) = 2.38, $p < 0.05$.

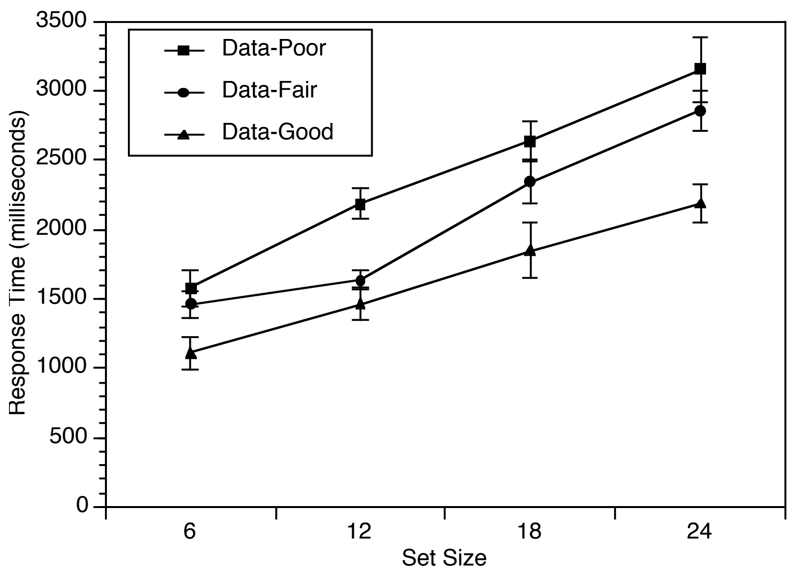


Figure 4. Average response time by set size and icon quality in the eye-tracking study.

The average number of gazes per trial are plotted as a function of icon quality and set size in Figure 5 (the solid lines represent gaze data from participants). Patterns in the gaze data were similar to those found in the response time data—i.e. as set size increases and icon quality decreases, the average number of gazes increases (as does response time). This is consistent with other studies that have found qualitatively similar patterns in RT data and the number of fixations per trial (e.g. Shen et al., 2000; D. E. Williams, Reingold, Moscovitch, & Behrmann, 1997; Zelinsky & Sheinberg, 1997). Revealed in the average number of gazes per trial data were reliable main effects of set size, $F(3, 27) = 77.08$, $p < 0.001$, and icon quality, $F(2, 18) = 56.60$, $p < 0.001$, and a reliable interaction between set size and quality, $F(6, 54) = 5.64$, $p < 0.001$. The number of gazes made by each participant on each trial was highly correlated with their response time for that trial. (Correlations for each subject ranged from $r(143) = 0.55$, $p < 0.01$ to $r(143) = 0.86$, $p < 0.01$.) Again, this is consistent with previous studies.

However, the model over-predicted the number gazes at all levels of icon quality and set size. The RMSE was 2.53 fixations; the PAAE was 77.08%, and the R^2 was 0.96. The relatively high RMSE and PAAE indicate a poor absolute model-to-data fit; however, the high R^2 indicates that the model did a good job of fitting the general trends in the data.

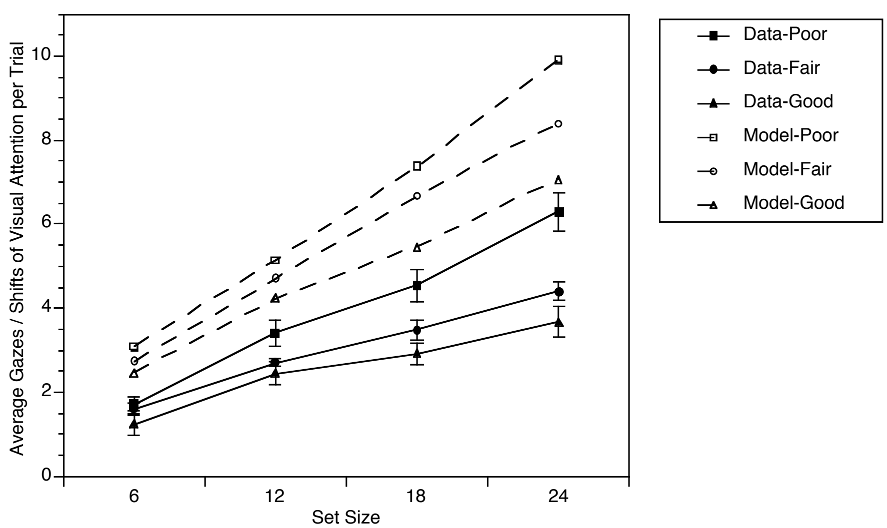


Figure 5. Mean number of shifts of visual attention made by the model relative to the number of gazes made by participants in the eye-tracking study (Data) as a function of set size and icon quality.

In Figure 6, the ratio of target-matching (TM) fixations (fixations to TM icons) to total fixations is presented as a function of icon quality and set size. Non-target fixations are fixations to any icon other than a TM icon in the distractor set. Fixations to areas outside of the distractor set of icons (i.e., when a participant fixated the region of the screen that was not part of the icon set) were excluded from this analysis (approximately 8% of the total number of fixations). Participants had a higher proportion of TM fixations relative to non-TM fixations as icon quality increased, $F(2, 18) = 7.87$, $p < 0.01$ with Huynh-Feldt correction. Additionally, participants made a higher proportion of fixations to target-matching icons than would be expected if fixations were randomly directed, $t(9) = 6.90$, $p < 0.01$.

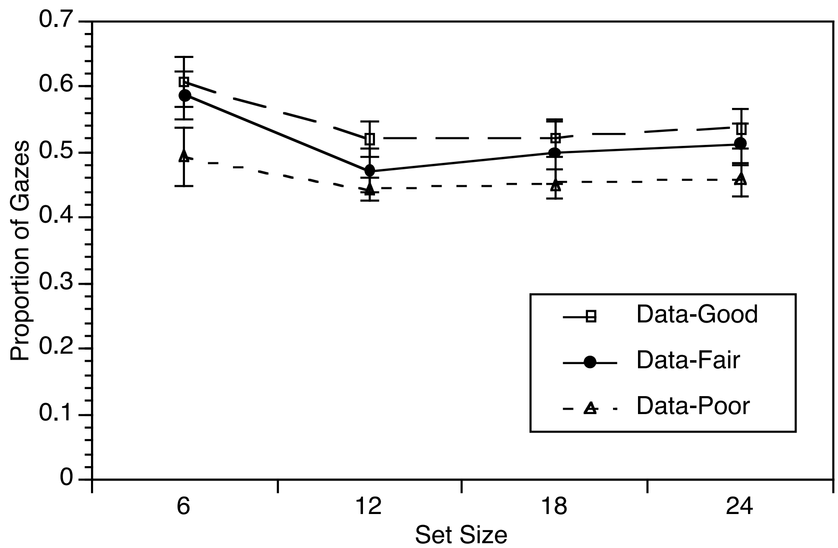


Figure 6. Ratio of target-matching (TM) gazes to total gazes by icon quality and set size, indicating that participants made a higher proportion of target-matching gazes with better quality icons.

Several qualitative patterns emerged in the data, which were reflective of the aforementioned patterns in the fixation and gaze data. First, it seemed that participants in the experiment used different search strategies depending on the level of quality of the icons. For instance, in the good quality condition the search strategy used by participants was often directed specifically TM icons. For an example, see Figure 7. In this case, the saccades were nearly all directed to a TM icon or fell in the area between two groups of TM icons, leaving whole areas of the distractor set unexamined. Second, this “directed” strategy often began with the largest group of TM icons (icons adjacent to one another) and proceeded to smaller groups of TM icons until the target was identified. In contrast, search strategies in the poor quality condition were not directed at TM icons and might cover the whole set of icons, possibly in a circular or zigzag pattern (Figure 8).

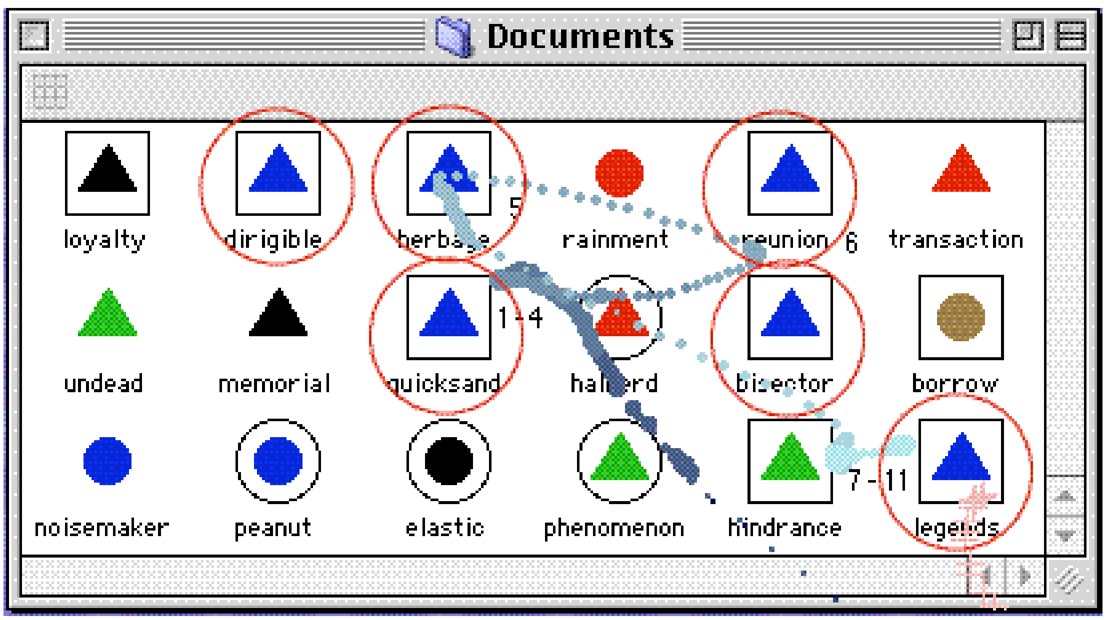


Figure 7. Example of a directed search with good quality icons. The round dots indicate point of regard, going from darker to lighter with time. The numbers to the right of an icon represent the number and order of fixations that were attributed to that icon in the analysis of the data—i.e. the first four fixations were attributed to the icon labeled “quicksand.” The cross-hairs (in the lower right) indicate the position of the mouse. The target-matching icons are circled. (The circles were not part of the experiment stimuli.) Note that the participant only examines a small subset of icons—those matching the target icon. The participant begins with the largest group of target-matching icons and eventually proceeds to the single target-matching icon in the lower right.

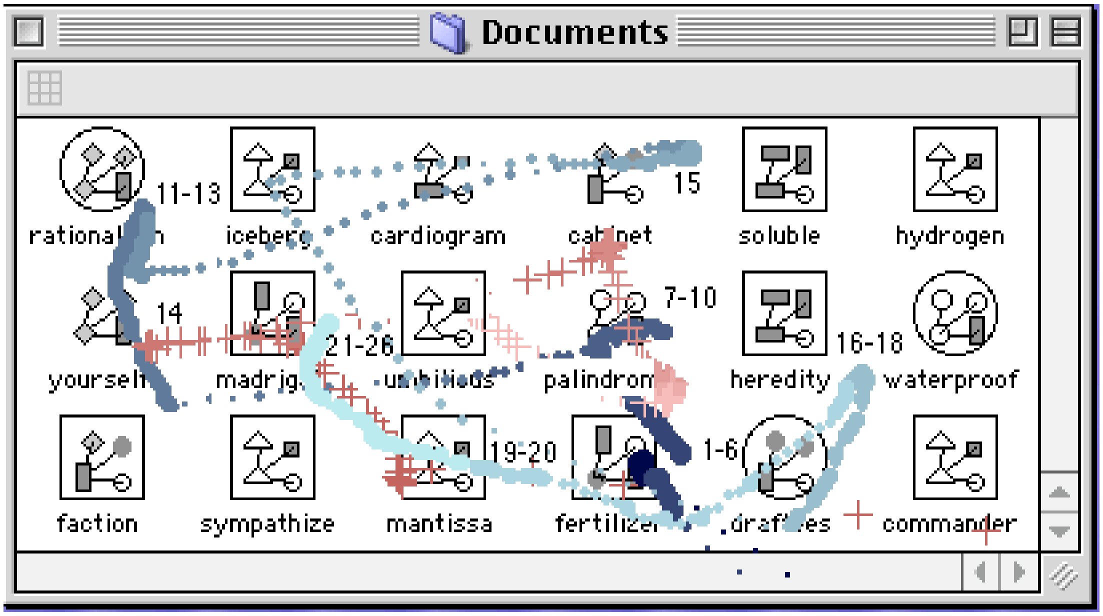


Figure 8. Example of an undirected search with poor quality icons. Following the dots, which indicate point of regard, from dark to light (with time) and the numbers to the right of an icon (which indicate which fixations were attributed to that icon in the analysis), indicates that the subject examined nearly the entire set of icons in a zigzag manner. The mouse position (cross-hairs) follows a similar pattern.

4.3.1. Analysis of Fixation Contingency

The model predicted that participants would show a preference for fixating TM icons, but it predicted that they would have uniform preference for all TM icons—i.e. that the probability that a participant would fixate any TM icon was equal for all TM icons. It was clear from watching replay videos of the trials that this was not the case—that participants were not random in their search through the TM icons. The authors were familiar with a computational model of vision, EMMA (which will be discussed in some detail subsequently), which predicted that an efficient icon search strategy in terms of average saccade distance would be to examine the target-matching icon nearest to the current point of regard. In this case, a participant's next fixation would be contingent on the location of his or her current fixation. To examine whether the fixation patterns of participants exhibited any evidence of this "nearest" strategy, we investigated the probabilities of subsequent fixations landing on target-matching icons; specifically, given a current fixation, what was the probability that the participants' next fixation would be directed to (1) a target-matching icon and (2) the nearest target-matching icon to the current point of regard.

For all trials, the final fixation was not considered in the analysis as a current fixation, since there was no subsequent fixation to examine. For the same reason, all trials in which all fixations were directed to only a single icon on the region of screen containing icons (2.4%) were eliminated from the analysis.

The proportion of fixations where the contingent fixation was to a target-matching icon is presented in Figure 9. Participants were able to direct their subsequent fixations to TM icons at above chance accuracy. (Accuracy in this context is defined as the likelihood of fixating a target-matching icon.) Even in the poor icon quality condition, where the icons were designed so that the TM icons would be difficult to distinguish from non-TM icons, the proportion of TM fixations differs reliably from the proportion one would expect if fixations were randomly directed (one-third), $t(9) = 4.93$, $p < 0.001$. Inverse patterns to those observed in the response time and fixation data are apparent. Specifically, as icon quality increases the accuracy of participants fixations increases, $F(2, 18) = 20.92$, $p < 0.001$. And as the set size increases the fixation accuracy decreases, $F(3, 27) = 3.98$, $p < 0.05$.

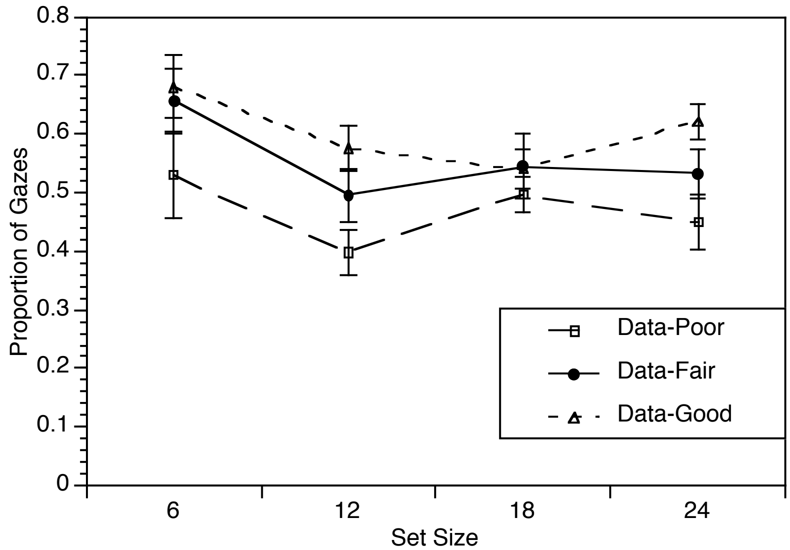


Figure 9. Proportion of fixations in which the next fixation was to a target-matching icon.

The proportion of contingent target-matching fixations in which the subsequent fixation was to a nearest target-matching icon was also calculated. A nearest TM icon was defined in terms of the number of icons lying in-between the currently fixated icon and a TM icon. Note that multiple TM icons could qualify to be a nearest TM icon. For instance, for any icon there could be multiple TM icons immediately adjacent to it, and each of these adjacent target-matching icons would qualify as a nearest TM icon. If there were no TM icons adjacent to the current icon, then TM icons adjacent to the adjacent icons would be considered the nearest TM icons, and so on and so forth. Also, note that there is a high probability that the next target-matching fixation would be to a nearest TM icon. For instance, in a set size of six icons, there are two TM icons. Hence, if the next fixation were to a target-matching icon, it would have to be to the nearest target-matching icon since there is only one other TM icon in the set.

Across all conditions of icon quality and set size, nearly all of the participants' contingent fixations were directed to a nearest TM icon. Where the subsequent fixation was to a TM icon, the percentage of fixations directed to a *nearest* TM ranged from approximately 99% to 95%. Even at the largest set size in the poor quality condition, where the model predicted that participants would be the least efficient in their visual search, nearly all (approximately 95%) of the TM fixations were to a nearest TM icon.

Again, patterns corresponding to those found in the RT and fixation data were observed in the proportion of subsequent fixations to a nearest TM icon. A reliable effect of set size, $F(3, 27) = 4.53$, $p < 0.05$, and a reliable effect of icon quality, $F(2, 18) = 22.32$, $p < 0.001$, indicate that participants fixated a higher proportion of TM icons nearest the current POR at lower sizes and better quality icons. Also, even in the poor icon quality condition, the proportion of TM fixations differs reliably from the proportion one would expect if fixations were randomly directed (one-third), $t(9) = 4.72$, $p < 0.01$.

4.3.2. Reexamination of Icons

In order to examine the model prediction that participants would reexamine icons, the proportion of fixations to an icon that had previously been examined was calculated. A fixation was considered in this category if there was at least one intervening fixation to another icon in-between a fixation or fixations to a single icon—i.e. the participant looked at an icon, then looked at other icons, then returned to the icon. In the poor quality condition, where there was the most reexamination of icons, icons were reexamined very infrequently, a maximum of approximately 4% of the time at the largest set size. This suggests that people reexamine icons very infrequently and at a rate that is within the margin of error of the system employed.

4.4. Discussion of Eye-tracking Results

The eye-tracking data revealed a strong correspondence with the reaction time data from the previous experiments. As we saw increases in reaction time with increases in set size and decreases in icon quality, we saw corresponding increases in the number of gazes per trial under the same conditions. The model also predicted the response time data well (Figure 3).

Despite predicting the response-time data well, the model over-estimated the number of gazes per trial across all set sizes and levels of quality (Figure 5). This indicates that the estimated time to make shift of visual attention in the model is faster than the average gaze duration of participants. This may be due to two possible reasons. One reason is simply that the model-estimated time to make a shift of visual attention and encode the item at that location, 135 ms (50 ms for a production to fire and 85 to shift attention and encode an item), is simply too fast for the current task. The 85 ms setting is the unadjusted estimation of this time in ACT-R 5.0, and it is quite possible that we may achieve more accurate predictions by adjusting this parameter. A second possibility is that participants are making many covert shifts of visual attention, i.e., shifting visual attention and encoding information without making a measurable fixation on the information. Both of these possibilities are considered further in the next section.

The model predicted that participants would be more accurate in locating target-matching icons as icon quality increased. This effect was manifested in the fixation data through the proportion of TM fixations to total fixations, which increased with each level of improvement in icon quality (Figure 6). There was also some evidence for this effect at a qualitative level, manifested in the "directed" search strategies in the good quality icons and the "undirected" search strategies seen with poor quality icons.

One notable pattern in the data arises in the average number of gazes per trial across the four different set size conditions (Figure 5). The average number of gazes per trial more than doubles from the smallest set size (6 icons) to the largest set size (24 icons). However, the proportion of target-matching fixations across the range of set sizes only decreases approximately 10% across the four different set sizes (Figure 5). From these two patterns in the data, it can be inferred that although the number of fixations increases greatly with set size (nearly a "100% effect"), the ability of participants to shift attention to TM icons does not change nearly so dramatically, as measured by the proportion of participants' fixations directed to target-matching icons.

The accuracy of the participants' fixations, as measured by the frequency of contingent fixations on target-matching icons, suggests that participants were able to perform a relatively efficient conjunctive visual search for target-matching icons. By definition, the proportion of subsequent fixations to a nearest target-matching icon is lower than the proportion of subsequent fixations to target-matching icons (the set of fixations to the nearest TM icons are a subset of fixations to TM icons). What is remarkable, however, is that the proportions are remarkably similar. This indicates that in nearly all cases where the participants' next fixation went to a TM icon, it was to a nearest TM icon. Participants showed a clear preference for fixating the TM icon nearest to their current point of regard. The model did not capture this aspect of the participants' behavior. The model predicted that participants' fixations would be randomly directed to icons sharing some level of similarity (represented in the model as an attribute pair) with the target icon.

Close examination of the data also speaks to two other possible search strategies potentially employed by participants. One potential strategy of users would be to simply shift attention to the icons near to the current point of regard. However, with such a strategy, the proportion of TM icons attended to on subsequent fixations would be approximately 0.33, as one-third of the distractor set is composed of TM icons, a level far below that found in the data. The data also refute the possibility that users are simply searching the display in a systematic left to right, top to bottom, or some other directional manner. With such a strategy, users would not show such a high proportion of subsequent fixations to TM icons, as shifting attention according to such a rote strategy would cause participants to frequently shift attention to a TM icon that is farther away than a nearest TM icon. Hence, the proportion of contingent fixations to TM icons would be lower. However, it is possible that some combination of the aforementioned strategies is responsible for producing the pattern of visual search activity manifested in the eye movement data. Users clearly show a preference for directing their attention to TM icons, to icons near to the current POR, and they may even do so using a directional strategy (although the qualitative data do not show evidence for a directional strategy).

Regarding the average number of gazes per trial, participants made fewer gazes than the model predicted. The greater number of shifts of visual attention made by the model may be due to the model’s behavior of reexaming icons. Evidence in the form of low icon revisitation rates indicates that participants have an accurate memory for where they have looked in this task and only reexamine icons very infrequently. The greater number of model shifts of visual attention may also be due to a disassociation between visual attention and shifts of POR. It is possible that the participants are able to examine multiple icons within a single gaze.

5. Revising the Model

The eye-tracking study highlighted several areas where the model's strategies did not match those of users in terms of search efficiency. This section outlines some of the potential revisions to the models suggested by the results of the eye-tracking study.

5.1. Number of Gazes per Trial

The poor fit of our model to the eye-tracking data in terms of the average number of gazes per trial led us to consider an issue in the underlying cognitive architecture of ACT-R that other authors have discussed previously (Salvucci, 2001). ACT-R by default only makes predictions regarding unobservable attention shifts. Yet the data used in our analysis of eye movements was, by necessity, based on observable movements in participants' POR. It is well-established in the research community that eye movements do not necessarily mimic movements of visual attention, i.e., people do not always move their eyes to their focus of attention (Henderson, 1992; Rayner, 1995). The experiments modeled here may provide an example where this is the case. Fortunately, there is an extension to ACT-R's vision module that addresses the disassociation between eye movements and movements of attention.

5.2. Eye Movements and Movements of Attention (EMMA)

EMMA is a computational model that serves as a bridge between observable eye movements and the unobservable cognitive processes and shifts of attention that produce them. The model describes whether or not eye movements occur, when they occur, and where they land with respect to their targets (Salvucci, 2001).

Concerning visual encoding, the model describes how peripheral viewing and object frequency affect the time needed to encode a visual object into an internal representation. When cognition requests a shift of attention to a new visual object, EMMA encodes the visual object into an internal representation. The time T_{enc} needed to encode object i is computed as follows:

$$T_{enc} = K\,[-\log f_i]\, e^{k\varepsilon_i}$$

The parameter f_i represents the frequency of the object encoded, specified as a normalized value in the range (0,1). The parameter ε_i represents the eccentricity of the object, measured as the distance from the current eye position to the object in units of visual angle. Thus, encoding time increases as object eccentricity increases and as object frequency decreases. (The constants K and k are scaling constants.) The encoding equation is based on an algorithm from the E-Z Reader model of eye-movement control in reading (Rayner, Reichle, & Pollatsek, 1998; Reichle, Pollatsek, Fisher, & Rayner, 1998).

The time needed to make an eye movement is also calculated in EMMA. The majority of the eye movement time is based on fixed parameters, but it is also based partly on the eccentricity of the object, i.e. the longer the saccade, the greater the calculated eye movement time (2 ms for each degree of visual angle subtended by the saccade).

With respect to spatial characteristics, EMMA provides a simple formalization of where eye movements land with respect to a desired destination. Given a saccade to a particular object, the model assumes that the landing point follows a Gaussian distribution around the center of the object. The variance of the distribution is a function of the distance of the saccade such that longer saccades are generally less accurate.

The control flow of the EMMA system describes how cognition, visual encoding, and eye movements interact as interdependent processes. When cognition requests an attention shift to a new visual object (such as a new icon in our paradigm), EMMA begins encoding the object while an eye movement is prepared and (possibly) executed. Eye movements occur in two stages, preparation, which is the retractable or “labile” stage of the eye movement program, and execution. If the encoding of the object completes and cognition requests a subsequent shift of attention before the preparation of the eye movement is complete, then the eye movement is canceled and a new eye movement may begin. If the attention shift occurs during eye-movement execution, execution continues to run to completion while preparation for a new eye movement is begun. If the eye movement completes before encoding completes, encoding continues and a new eye-movement is prepared. However, because the eye movement has (presumably) brought the POR nearer to the visual object, encoding speed increases accordingly.

5.2.1. Incorporating EMMA

We incorporated EMMA to improve the performance of the models in those areas that the eye-tracking study indicated were ripe for improvement, specifically the number and location of fixations and the timing of movements of visual attention.

As noted previously, our models made too many shifts of visual attention relative to the number of fixations made by participants in the eye-tracking study. With EMMA, the number of shifts of visual attention will not decrease; however, because eye movements do not directly correspond with shifts of attention in EMMA, we could expect the number of eye movements, or shifts of POR, to decrease. When the encoding time for a visual object is less than the time to make the labile portion of the eye movement, then the eye movement is not made, even though the object has been examined.

In addition to seeing a decrease in the number of shifts of POR made by the models, we also expected to see increasingly similar patterns of the location of shifts of the simulated POR relative to the eye movements of participants. Although visual attention will be focused on the filename selected by the model, the actual point of regard calculated by EMMA is based on a Gaussian distribution around the center of the object. Thus, the model will not always shift its POR directly to the center of an object. However, when visual attention remains focused on an object for an extended period of time, this will allow the model to make successive shifts in its POR, each one presumably more accurate than the last (i.e. focused on the center of the target object). This will occur due to longer encoding times for the object or because cognition has not requested an additional shift of visual attention, as when the model is making a mouse movement to the object. One of the instances where participants consistently and overtly attended the target icon was when selecting it with the mouse; EMMA predicts this behavior.

We also hoped that the incorporation of EMMA into our modeling efforts would provide greater predictive power of our models regarding the timing of fixations. Our previous models used a fixed estimate of time to make a shift of visual attention and encode and object of 85ms, the default value in ACT-R. However, there is a large body of evidence that suggests that the time to make a saccade and encode an object is a function of a number of different factors, one of the most influential of which is the eccentricity of the target object (e.g. Fuchs, 1971; Russo, 1978). Because the eccentricity of the object is taken into account when calculating the time to make a shift of the POR in EMMA, the incorporation of EMMA will allow us to make predictions as to the relative efficiency of various icon search strategies based on the average length of shifts of visual attention. Specifically, strategies that make shorter shifts of visual attention can be expected to be more efficient.

5.3. Improving the Model Search Strategy

The most significant finding from the eye-tracking study that we wanted to incorporate into the model stemmed from the efficiency of participants' fixations. Participants followed a strategy of looking at a TM icon near to their current POR. In order to accommodate this strategy in the model, we adopted a "nearest" strategy. The model would simply select the target-matching icon nearest to the current focus of visual attention. Thus, if examining an icon in a group of target-matching icons, the model will examine all of the icons in the group before moving on to the next group. Such a strategy also ties in with the predictions made by EMMA regarding the advantages of making shorter shifts of visual attention. Specifically, a strategy that makes the shortest possible shift will be the most efficient strategy.

An additional aspect that was changed was the model's behavior of revisiting icons that it had already examined. Because the model did not actually shift visual attention to an icon, it had no memory for which icons it had examined. We changed the code in ACT-R's vision module to allow us to mark specific objects at a location as having been attended even when visual attention had not explicitly been directed there. Specifically, we had the visual system mark an icon as having been attended when the filename below the icon was examined. The new model would not shift attention to locations that it had previously attended.

5.4. Modeling Results

The model was run for 80 blocks of trials; predictions are the averages over those 80 blocks. When the three model improvements were incorporated into the model, EMMA, nearest-TM-icon strategy, and marking icons as attended, the RMSE was 129 ms; the PAAE was 5.89%, and the R^2 was 0.99 (see Figure 12). On the basis of response time alone, relative to our previous models, the new model maintained the accuracy of the original model despite the introduction of the new features.

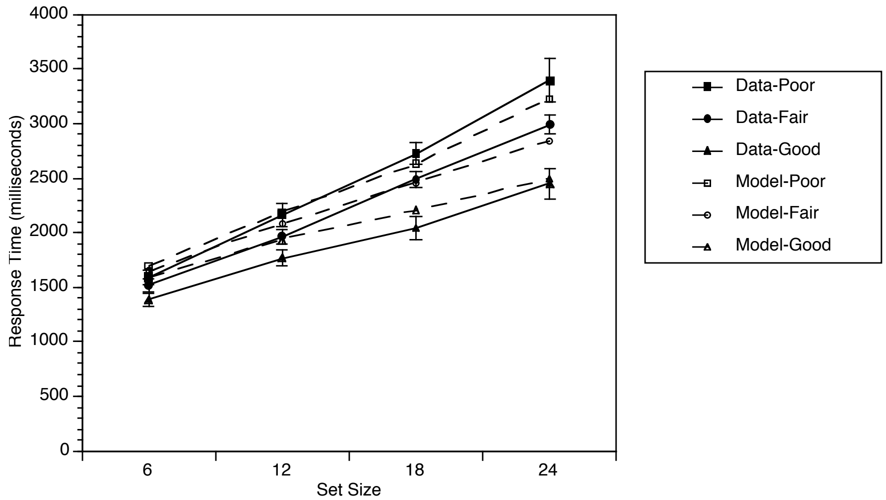


Figure 12. Response time by set size and icon quality for the revised model and the experiment data. The revised model data includes the incorporation of EMMA, marking icons as attended, and using the "nearest" search strategy.

We also compared the mean number of gazes made by participants to the mean number of shifts of visual attention made by the model. This is presented for the revised model in Figure 13 (See Figure 5 for a comparison of the original model to the eye data). Relative to the previous model, the revised model fares much better although the model makes slightly more overt shifts than subjects. Using the same metrics for comparing the new model visual shift data to the experiment fixation data as for the response time data compared previously, the RMSE was 0.58 fixations; the PAAE was 15.79%, and the R^2 was 0.99.

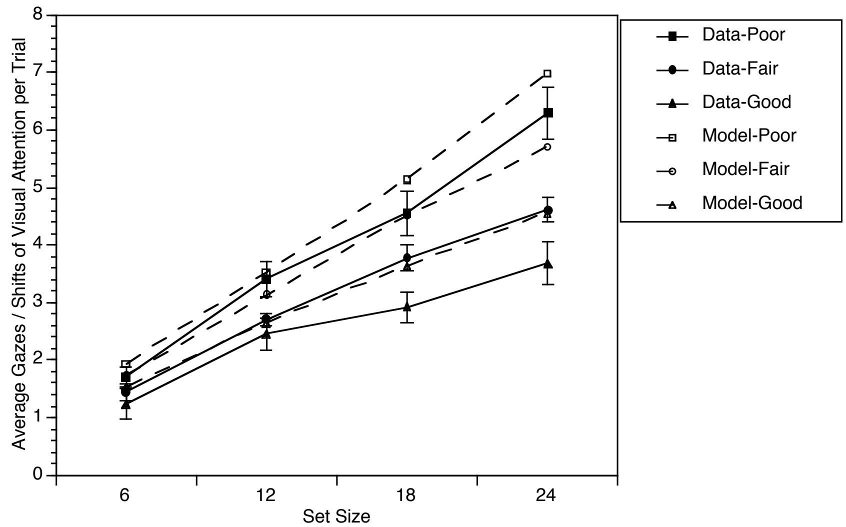


Figure 13. Mean number of shifts of visual attention per trial made by the model relative to the mean number of gazes per trial made by participants (Data).

We also found that the qualitative performance of the model was quite improved in many respects. One aspect of the eye-tracking study that we discussed was the general search patterns of participants. We noted that participants employed a “directed” strategy that was quite efficient in terms of only examining target-matching icons (at least with the good quality icons). There was also some evidence for a grouping strategy, whereby the icons in a group of target-matching icons were examined before moving on to another group of target matching icons. An example of a trial where these strategies were employed was shown (Figure 7). The new versions of the model were able to reproduce these qualitative aspects of the data quite well. Because the revised model examines TM icons nearest to the currently attended icon, it generally searches within a group of adjacent TM icons before jumping to a separate group of TM icons (because the adjacent TM icons are nearer to each other). As an example of the capability of the models, the exact same trial as was presented to the user in Figure 7 was run with the model (see

Figure 14). The line running through the figure shows the resulting trace of the POR data of the revised model. The model begins its search from the "Ready" button and enters the depicted portion of the trial from the lower-right corner. The model proceeds in a fashion quite similar to that of the human participant, first examining the largest group of icons before moving on to the nearest group and finally to the target icon in the lower-right corner of the window (labeled "legends"). The model would not follow this exact search pattern every time (the first icon that the model selects for search is randomly chosen from the target-matching icons presented), but the capability of the model to mimic human performance in this respect is encouraging. Two caveats are that the model never revisits an icon and always shifts attention to a nearest TM icon. Hence, at least in the good quality condition, the model is extremely efficient in its search.

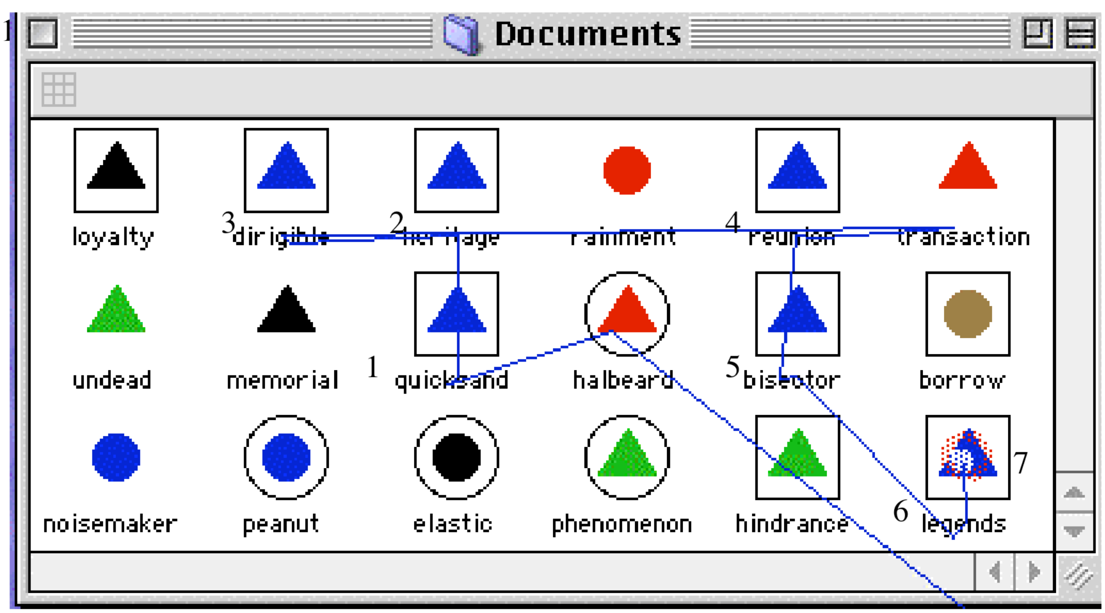


Figure 14. Example of the Text-Look model running an identical trial to that presented in Figure 7. The line indicates the POR path of the model, and the numbers represent the order of the visual attention shift. (All shifts were to file names except the final shift, which was to the icon above "legends.") The model POR data begins at the "Ready" button (not show), enters the view in the lower-right corner and finishes by selecting the icon above the filename "legends."

5.5. Discussion of Modeling Revisions

One improvement in the revised model was the inclusion of the EMMA model to disassociate eye movements and movements of attention. When the EMMA system was incorporated into the model without any other changes, the effect was an overall increase in response time. The previous models used a constant parameter of 85 ms for each shift of visual attention. EMMA uses a set of algorithms to compute the encoding and saccade time based primarily on the eccentricity of the target object from the current POR and the frequency of encountering the target object. After incorporating EMMA into the model, we found that the values computed by EMMA for shifting visual attention to and encoding each new icon average to a value greater than 85 ms. A closer examination of this attribute of EMMA revealed that longer shifts of visual attention, such as those from one side of the distractor set to the other side, took an estimated time much greater than 85 ms, and were thus responsible for much of the increase in average saccade time.

The increase in average time shift visual attention was compensated for in the revised model through the other two major improvements in the model suggested by the eye tracking study, the "nearest" strategy and marking icons as attended. The nearest strategy, always examining TM icons nearest to the currently attended icon, resulted in shorter average shifts of visual attention. Shorter shifts of attention correspond to shorter average times to make each shift and encode the item at the new location, as calculated by EMMA, and shorter average model response times. By marking icons as attended, even though only filename below each icon was actually attended by the model, the model no longer revisits icons. No revisitation of icons by the model meant less shifts of visual attention per trial and lower average model response times.

The aggregate effect on response time of incorporating EMMA, the nearest strategy, and marking icons as attended into the revised model was minimal. According to the metrics of comparison employed, RMSE, PAAE, and R^2, the revised model did not fit the participant response time data any better or worse than the original model. However, the revised model showed substantial improvement in fitting human performance in terms of correspondence to the eye-tracking data. This is an important point in the creation of simulated human users—SHUs must show human-like performance on more aspects of a task than just response time. Our revised model showed marked improvement to fitting the average number of gazes per trial made by participants. Also, the search patterns of the revised model were a much better approximation of the visual search patterns exhibited by experiment participants. Specifically, the model now exhibited a preference for examining icons nearest to the currently attended icon. Also, as a result of the nearest strategy, the model now searches within groups of TM icons before searching between groups, a pattern also exhibited by participants. However, it should be noted the model is now slightly too good with respect to searching the nearest TM icons. In the good quality condition, the next icon examined by the model is *always* the nearest TM icon.

6. General Discussion

One of the more pronounced effects seen in the studies presented here was the effect of icon quality. This effect was modeled by assuming that participants were able to locate icons which matched a feature of the target icon preattentively and that they could direct their visual attention to these locations with greater than chance accuracy. The evidence in the eye-tracking studies presented here suggests that participants were indeed able to do that. The ability of users to preattentively discriminate subsets of visual objects (such as "all blue objects") in conjunctive search tasks is not a new discovery (e.g. Alkhateeb et al., 1990; McLeod et al., 1991; Nakayama & Silverman, 1986; Treisman & Sato, 1990; Wolfe, 1992), and it is predicted by the guided search model (Wolfe, 1994; Wolfe et al., 1989; Wolfe & Gancarz, 1996). Additionally, previous research has shown that participants are able to adapt their search strategies to the visual environment "on the fly"—i.e., from trial to trial or with each new visual display (Shen et al., 2000). The contribution of this research is to show that these findings and predictions hold in a more complex visual environment and task approximating that commonly encountered by modern GUI users (McDougall et al., 2000).

Further analysis of the eye-tracking data revealed that participants made virtually no fixations on icons that they had previously fixated; that is, participants had almost perfect memory for where they had looked. Whether people have memory for where they have looked in a visual search context is currently the subject of some debate in the research community. One set of results suggest that participants have no memory for where they have searched (e.g. Horowitz & Wolfe, 1998). However, other researchers (e.g. Peterson, Kramer, Wang, Irwin, & McCarley, 2001) have found that people do indeed show search patterns that would indicate they have memory for where they have looked. The evidence from our eye-tracking study clearly agrees with the latter, as our participants clearly had memory for where they had looked. This may be because the task required that each icon was processed to a level of depth that included location information or simply because it was a task that required eye movements. While our data do not weigh in on the source of this memory, our data suggest that memory of visual search generalizes to HCI tasks. Many of the visual search tasks encountered by computer users, such as searching through menus or through lists of filenames, require reading or at least that attention be shifted directly to individual items. To the extent that eye movements and direct examination of individual objects are the precursors of memory in visual search tasks as evidence suggests (Boot, McCarley, Kramer, & Peterson, in press), it is likely that computer users exhibit visual search memory in such tasks.

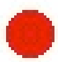
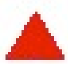
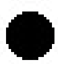
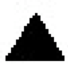
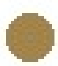
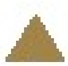
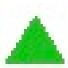
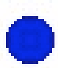
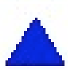
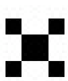
The search strategy of searching the icon nearest to the currently fixated icon also has implications well beyond the realm of icon search. Tullis (1997 p. 510) discusses the grouping of information in the realm of screen design issues and techniques. “The ways in which the elements are grouped plays an important role in both the ease with which the users can extract the information and the interpretations that they assign to it.” Other researchers have made similar distinctions. For example, “grouping similar items together in a display format improves their readability and can highlight relationships between different groups of data.” (Cakir, Hart, & Stewart, 1980. p. 114). From this perspective, the organization of information on the screen has value to the user by giving them some additional categorical information regarding what is presented on the screen as well as improving the general “readability” of the information. Although there is surely value in categorical information, from the perspective of our modeling effort, there is additional value in grouping the information on the screen that is reflected at a much lower level in the cognitive system—in the visual search strategies employed by users. Grouping information will tend to reduce the number and average distance of shifts of visual attention made by the user while searching for a desired piece of information. Shorter shifts and less of them will result in finding the desired information more quickly.

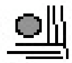
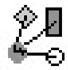
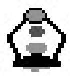
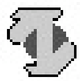
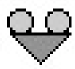
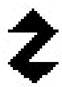
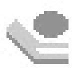
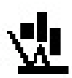
It is worth noting that we began to investigate the possibility that participants were using the "nearest" strategy in the eye-tracking data as a result of the modeling endeavor. While looking for ways to improve the efficiency of the model, we explored the possible addition of a computational model for vision, EMMA, to the ACT-R model. EMMA made the clear prediction that an efficient search strategy should minimize the average saccade distance, which is what we discovered in the eye-tracking data. Without the aid of the modeling endeavor, it is unclear whether we would have considered looking for this strategy.

This research has implications beyond the specific domain of icon search as well. In particular, it speaks to issues of model complexity and constraint on cognitive models. A general problem with applying computational cognitive models to real-world domains is the general lack of constraints on such models. What our research has shown is that using response time alone is not a strong enough constraint on the model-building process; we were able to fit the response time data well with an inaccurate strategy. By providing the model with a more human-like strategy, we were able to capture the effects found in the eye-tracking data without sacrificing the ability to correctly predict response time. While this did slightly increase the complexity of the ACT-R model, we believe this complexity was justified by the richer and more complex eye-tracking data, and our success in accounting for the key results found there. As we continue towards the development of simulated human users capable of making *a priori* predictions of human performance, it is essential that the criteria by which we judge the models become increasingly stringent. In the visual world of graphical user interfaces, eye-tracking data will not only inform the development of the models, but also provide additional criteria on which they may be judged.

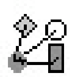
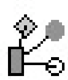
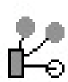
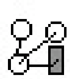
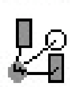
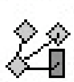
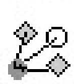
Appendix A.

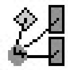
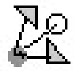
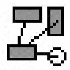
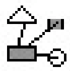
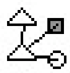
For each icon: ratings of complexity and distinctiveness, and list of ACT-R/PM attribute pairs.

Icon	Complexity Rating	Distinctivenes s Rating	Automated Complexity Rating (Forsythe, et al. 2002)	ACT-R Feature List
Good	**Quality Icons**			
	1.05	3.86	48	circle yellow
	1.09	3.91	55	triangle yellow
	1.09	3.95	48	circle red
	1.14	4.23	55	triangle red
	1.00	4.05	48	circle black
	1.09	4.18	55	triangle black
	1.09	3.50	48	circle brown
	1.05	3.45	55	triangle brown
	1.05	3.73	48	circle green
	1.14	3.77	55	triangle green
	1.05	3.82	48	circle blue
	1.09	4.00	55	triangle blue
Good Avg.	1.08	3.87	51.50	
Fair	**Quality Icons**			
	1.95	4.59	116	square black; square white; checkers b-and-w

	4.00	4.50	90	diagonal-left dark-gray; stripes black; diagonal-right gray
	3.68	3.86	126	circle-large gray; stripes black
	4.05	3.36	101	rectangle gray; triangle gray; circle-small gray
	3.82	3.41	133	circle-small; gray circle; gray stripes; black diagonal black
	3.09	3.73	85	rectangle gray; diagonal-left gray; diagonal-right gray
	2.64	3.50	73	circle-large gray; triangle gray
	2.41	4.45	79	square black; diagonal black; square white; triangle black
	3.05	3.68	105	oval-targ gray; diagonal-right gray; rectangle gray; stripes gray
	2.32	4.14	76	oval-targ gray; triangle gray; circle gray
	4.50	3.59	150	rectangle gray; rectangle dark-gray; stripes black
	2.95	4.09	103	rectangle gray; square black; diagonal-right black
Fair Avg.	3.20	3.91	103.08	

Poor Quality Icons

	4.14	1.32	106	square gray; circle-top white; rectangle-btm dark-gray; circle-btm dark-gray
	4.09	1.32	101	square gray; circle-top dark-gray; circle-btm white; rectangle-btm dark-gray
	3.86	1.95	103	circle-top dark-gray; circle-btm white; rectangle-btm dark-gray; double-circle gray
	4.00	2.14	114	circle-top white; circle-btm white; rectangle-btm dark-gray; triple-circle white
	4.00	2.09	118	circle-top white; rectangle-top dark-gray; circle-btm dark-gray; double-bar dark-gray
	3.95	1.82	100	square gray; rectangle-btm dark-gray; triple-diamond gray
	4.05	1.55	93	square gray; circle-top white; circle-btm dark-gray; double-diamond gray

	4.14	2.68	113	square gray; rectangle-top dark-gray; circle-btm dark-gray; double-bar dark-gray
	4.14	2.32	113	square gray; circle-top white; circle-btm dark-gray; double-triangle gray
	4.14	2.00	123	rectangle-top dark-gray; circle-btm white; horizontal dark-gray; triple-bar dark-gray
	4.27	1.73	104	circle-btm white; horizontal dark-gray; triangle white
	3.91	1.86	100	square gray; circle-btm white; double-triangle white
Poor Avg.	4.06	1.90	107.33	

References

Alkhateeb, W. F., Morland, A. B., Ruddock, K. H., & Savage, C. J. (1990). Spatial, colour, and contrast response characteristics of mechanisms which mediate discrimination of pattern orientation and magnification. *Spatial Vision, 5*(2), 143-157.

Altmann, E. M. (2001). Near-term memory in programming: A simulation-based analysis. *International Journal of Human-Computer Studies, 54*(2), 189-210.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Quin, Y. (in press). An integrated theory of mind. *Psychological Review*.

Anderson, J. R., Bothell, D., Byrne, M. D., & Lebiere, C. (in press). An integrated theory of mind.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*, 439–462.

Bacon, W. F., & Egeth, H. E. (1997). Goal directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 948-961.

Becker, W., & Jurgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research, 19*, 967-983.

Boot, W. R., McCarley, J. M., Kramer, A. F., & Peterson, M. S. (in press). Automatic and intentional memory processes in visual search. *Psychonomic Bulletin & Review*.

Byrne, M. D. (1993). *Using icons to find documents: Simplicity is critical*. Human Factors in Computing Systems: Proceedings of INTERCHI'93, 446-453.

Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies, 55(1)*, 41–84.

Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review, 108*, 847–869.

Byrne, M. D., Anderson, J. R., Douglass, S., & Matessa, M. (1999). Eye tracking the visual search of click-down menus. In *ACM CHI'99 Conference on Human Factors in Computing Systems* (pp. 402-409). New York: ACM.

Cakir, A., Hart, D. J., & Stewart, T. F. M. (1980). *Visual display terminals: A manual covering ergonomics, workplace design, health and safety, task organization*. England: Wiley.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Carter, R. C. (1982). Visual search with color. *Journal of Experimental Psychology: Human Perception & Performance, 8*(1), 127-136.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception & Performance, 10*, 32-39.

Ehret, B. D. (2002). *Learning where to look: Location learning in graphical user interfaces*. Paper presented at the ACM INTERCHI 2002 Conference on Human Factors in Computing Systems.

Everett, S. P., & Byrne, M. D. (2004). *Unintended Effects: Varying Icon Spacing Changes Users' Visual Search Strategy*. Paper presented at the Proceedings of ACM CHI '04: Conference on Human Factors in Computing Systems, Vienna, Austria.

Findlay, J. M. (1982). Global visual processing for saccadic eye movements. *Vision Research, 22*(8), 1033-1045.

Fleetwood, M. D. (2001). *Computational modeling of icon search*. Unpublished Master's Thesis, Rice University, Houston.

Fleetwood, M. D., & Byrne, M. D. (2002). Modeling icon search in ACT-R/PM. *Journal of Cognitive Systems Research, 3*, 25–33.

Forsythe, A., Sheehy, N., & Sawey, M. (2003). Measuring icon complexity: An automated analysis. *Behavior Research Methods, Instruments, & Computers, 35*(2), 334-342.

Fuchs, A. F. (1971). The saccadic system. In P. Bach-y-Rita, C. C. Collins & J. E. Hyde (Eds.), *The Control of Eye Movements* (pp. 343-362). New York: Academic Press.

Gould, J., & Dill, A. (1969). Eye-movement parameters and pattern discrimination. *Perception and Psychophysics, 6*, 311-320.

Gray, W. D., & Altman, E. M. (2001). Cognitive modeling and human-computer interaction. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors*. New York: Taylor & Francis, Ltd.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8*(3), 237-309.

Green, B. F., & Anderson, L. K. (1956). Color coding in a visual search task. *Journal of Experimental Psychology, 51*, 19-24.

Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye Movements and Visual cognition: Scene Perception and Reading*. New York: Springer-Verlag.

Hornof, A. J. (2001). Visual search and mouse-pointing in labeled versus unlabeled two-dimensional visual hierarchies. *ACM Transactions on Computer-Human Interaction, 8*, 171–197.

Hornof, A. J., & Halverson, T. (2003). *Cognitive strategies and eye movements for searching hierarchical computer displays*. Paper presented at the ACM CHI 2003: Conference on Human Factors in Computing Systems, New York.

Hornof, A. J., & Kieras, D. E. (1997). Cognitive modeling reveals menu search is both random and systematic. In *Human Factors in Computing Systems: Proceedings of CHI 97* (pp. 107–114). New York: ACM Press.

Horowitz, H., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 357*, 575-577.

Jacko, J. A., Scott, I. U., Barreto, A. B., Bautsch, H. S., Chu, J. Y. M., & Fain, W. B. (2001, August 5-10). *Iconic visual search strategies: A comparison of computer users with AMD versus computer users with normal vision*. Paper presented at the 9th International Conference on Human-Computer Interaction, New Orleans , LA.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction, 3*, 320-351.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12*, 391–438.

Kieras, D. E., Wood, S. D., & Meyer, D. E. (1997). Predictive engineering models based on the EPIC architecture for multimodal high-performance human-computer interaction task. *Transactions on Computer-Human Interaction, 4*(3), 230-275.

Kitajima, M., & Polson, P. G. (1997). A comprehension-based model of exploration. *Human-Computer Interaction, 12*(4), 345-389.

McDougall, S. J. P., Curry, M. B., & De Bruijn, O. (1999). Measuring symbol and icon characeristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers, 31*, 487-519.

McDougall, S. J. P., De Bruijn, O., & Curry, M. B. (2000). Exploring the effects of icon characteristics on user performance: The role of icon concreteness, complexity, and distinctiveness. *Journal of Experimental Psychology: Applied, 6*, 291-306.

McLeod, P., Driver, J., Dienes, Z., & Crisp, J. (1991). Filtering by movement in visual search. *Journal of Experimental Psychology: Human Perception & Performance, 17*(1), 55-64.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320*(264-265).

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

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science, 12*, 287-292.

Poisson, M. E., & Wilkinson, F. (1992). Distractor ratio and grouping processes in visual conjunction search. *Perception, 21*, 21-38.

Rayner, K. (1995). *Eye movements and cognitive processes in reading, visual search, and scene perception. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.).* Eye Movement Research: Mechanisms, Processes, and Applications.

Rayner, K. & Pollatsek, A. (1989). The psychology of reading. Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., Reichle, E. D., & Pollatsek, A. (1998). Eye movement control in reading: An overview and model. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 243-268). Oxford, England: Elsevier.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(125-157).

Ritter, F. E., Baxter, G. D., Jones, G., & Young, R. M. (2000). Supporting cognitive models as users. *ACM Transactions on Computer-Human Interaction, 7*, 141-173.

Russo, J. E. (1978). Adaptation of cognitive processes to the eye movement system. In J. W. Senders, D. F. Fisher & R. A. Monty (Eds.), *Eye Movements and the Higher Psychological Processes* (pp. 89-112). Hillsdale, NJ: Lawrence Earlbaum Associates.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1*(4), 201-220.

Salvucci, D. D., & Anderson, J. R. (2000). Intelligent Gaze-Added Interfaces. In *Proceedings of ACM CHI 2000 Conference on Human Factors in Computing Systems* (pp. 273-280). New York: ACM.

Salvucci, D. D., & Macuga, K. L. (2002). Predicting the effects of cellular-phone dialing on driver performance. *Cognitive Systems Research, 3*, 95-102.

Shen, J., Reingold, E. M., & Pomplun, M. (2000). Distractor ratio influences patterns of eye movements during visual search. *Perception, 29*, 241-250.

Sibert, L. E., & Jacob, R. J. K. (2000). *Evaluation of eye gaze interaction*. Paper presented at the ACM CHI '00 Conference on Human Factors in Computing Systems, The Hague, Netherlands.

Smith, S. L. (1962). Color coding and visual search. *Journal of Experimental Psychology, 64*, 434-440.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance, 16*(3), 459-478.

Tullis, T. S. (1997). Screen design. In M. Helander, T. K. Landauer & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 503–532). New York: North-Holland.

Viviani, P., & Swensson, R. G. (1982). Saccadic eye movements to peripherally discriminated targets. *Journal of Experimental Psychology: Human Perception & Performance, 8*, 126-133.

Williams, D. E., Reingold, E. M., Moscovitch, M., & Behrmann, M. (1997). Patterns of eye movements during parallel and serial visual search tasks. *Canadian Journal of Experimental Psychology, 51*, 151-164.

Williams, L. G. (1966). The effects of target specification on objects fixated during visual search. *Perception and Psychophysics, 1*, 315-318.

Williams, L. G. (1967). The effect of target specification on objects fixated during visual search. *Attention and Performance, 1*, 355-360.

Wolfe, J. M. (1992). The parallel guidance of visual attention. *Current Directions in Psychological Science, 1*(4), 125-128.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202-238.

Wolfe, J. M. (2000). Visual attention. In K. K. De Valios (Ed.), *Seeing* (2 ed., pp. 335-386). San Diego, CA: Academic Press.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An alternative to the Feature Integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance, 15*, 419-433.

Wolfe, J. M., & Gancarz, G. (1996). Guided Search 3.0. In *Basic and Clinical Applications of Vision Science* (pp. 189-192). Dordrecht, Netherlands: Kluwer Academic.

Young, R. M., Green, T. R. G., & Simon, T. (1989). *Programmable user models for predictive evaluation of interface designs*. Paper presented at the CHI '89 Conference Proceedings: Human Factors in Computing Systems, New York, NY.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception & Performance, 23*(1), 244-262.

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? *Perception, 18*, 191-200.