Submitted to the journal "Human Computer Interaction" in April, 2009

An “Active Vision” Computational Model Of Visual Search For Human-Computer Interaction

Tim Halverson

Air Force Research Laboratory

Anthony J. Hornof

University of Oregon

RUNNING HEAD: ACTIVE VISION MODEL OF VISUAL SEARCH

Corresponding Author’s Contact Information:

Tim Halverson

Air Force Research Laboratory

Mesa Research Site

6030 South Kent St.

Mesa, AZ 85212

Email: thalvers@cs.uoregon.edu

Brief Authors’ Biographies:

Tim Halverson is a computer and information scientist with an interest in human-computer interaction, cognitive modeling, visual search, and eye tracking; he is a research associate in the Performance and Model Learning Group of the Air Force Research Laboratory. **Anthony J. Hornof** is a computer scientists with an interest in human-computer interaction, cognitive modeling, visual search, and eye tracking; he is an Associate Professor in the Department of Computer and Information Science at the University of Oregon.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **APR 2009**	2. REPORT TYPE	3. DATES COVERED **00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE
An 'Active Vision' Computational Model Of Visual Search For Human-Computer Interaction

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory,Mesa Research Site,6030 South Kent St.,Mesa,AZ,85212

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**75**	

Abstract

Visual search is an important part of human-computer interaction (HCI). The visual search processes that people use have a substantial effect on the time expended and likelihood of finding the information they seek. This work investigates visual search through computational cognitive modeling of empirical data. Computational cognitive modeling is a powerful methodology that uses computer simulation to capture, assert, record, and replay plausible sets of interactions among the many human processes at work during visual search. This work aims to provide a cognitive model of visual search that can be utilized by predictive interface analysis tools and to do so in a manner consistent with a comprehensive theory of human visual processing, namely active vision. The model accounts for the four questions of active vision: What can be perceived in a fixation? When do the eyes move? Where do the eyes move? What information is integrated between eye movements? The answers to these questions are important to both practitioners and researchers in HCI.

This work presents a principled progression of the development of a computational model of active vision. Three sets of data were modeled in the EPIC (Executive Process-Interactive Control) cognitive architecture. This work extends the practice of computational cognitive modeling by (a) providing the first detailed instantiation of the theory of active vision in a computational framework and (b) informing the process of developing computational models through the use of eye movement data. This instantiation allows us to better understand how these visual search processes can be used computationally to predict people's visual search behavior. The development of a

comprehensive model ultimately benefits HCI by giving researchers and practitioners a better understanding of how users visually interact with computers, and provides a foundation for tools to predict that interaction.

Contents

1. INTRODUCTION

Visual search is an important part of human-computer interaction (HCI). Users search news web sites to locate new stories of interest. Users search the interfaces of unfamiliar desktop applications to learn those applications. Users search the virtual environments of games to locate and identify objects that require more scrutiny or action. For sighted users, nearly every action requires some visual interaction and many of these actions require visual search to find familiar or novel information.

The visual search processes that people use in HCI tasks have a substantial effect on the time and likelihood of finding the information they seek. Visual search is a particularly fascinating human activity to study because it requires a complex and rapid interplay among perceptual, cognitive, and motor processes. Computational cognitive modeling is a very powerful methodology for capturing, asserting, recording, and replaying plausible sets of interaction among these processes.

The most important contribution of computational cognitive models to the field of HCI is a science base that is needed for predictive interface analysis tools. Projects such as CogTool (John & Salvucci, 2005) and CORE/X-PRT (Tollinger et al., 2005) are at the forefront of tools that utilize cognitive modeling to predict user interaction based on a description of the interface and task. These tools provide theoretically-grounded predictions of human performance in a range of tasks without requiring that the analyst be knowledgeable in cognitive, perceptual, and motoric theories embedded in the tool. Designers of device and application interfaces can use such tools to evaluate their visual layouts, identify potential usability problems early in the design cycle, and reduce the need for more expensive human user testing early in the development cycle.

Predicting people's visual interaction is one facet of user behavior that research with interface analysis tools is trying to improve. The most recent version of CogTool (Teo & John, 2008) now incorporates modeling work presented in this paper (based on an early summary of the work in Halverson & Hornof, 2007). However, interface analysis tools like CogTool do not yet account for the human eyes, where they move, and what they do and do not see. That is, automated interface analysis tools do not yet simulate *active vision*.

Active vision (Findlay & Gilchrist, 2003) is the notion that eye movements are a crucial aspect of our visual interaction with the world, and thus critical for visual search. When people interact with the environment (e.g. a user interface), they constantly move their eyes to sample information. Accounting for these eye movements will not only allow a better understanding of the processes underlying visual search, but also a better understanding of how people are using computer interfaces. Any simulation of active vision must address four questions, the answers of which are important to designers and those interested in HCI: When and why do we move our eyes? Where do we move our eyes? What information in the environment do we process during each fixation? What information from the environment do we maintain across fixations?

The goal of this work is to build a computational model of visual search for HCI that integrates a range of theory consistent with a contemporary understanding of the processes involved, as well as with the notion of active vision. This work presents a detailed step-by-step principled progression of the development of a computational cognitive model based on eye movement data and a synthesis of existing literature to provide answers to the questions put forth by active vision.

The remainder of this paper is arranged as follows: Section 2 reviews literature on cognitive modeling and visual search that is relevant to a computational cognitive model of active vision for HCI. Section 3 briefly discusses three eye tracking experiments whose data is used to inform the development and validation of the model discussed in subsequent sections. Sections 4 through 6 discusses the development of an active vision computational cognitive model of visual search. Section 7 summarizes the research, identifies key contributions, and suggests future directions.

2. MODELING VISUAL SEARCH

The focus of this work is to better understand and predict how people visually search computer displays in everyday, natural tasks. Typically, when people conduct a visual search task, the eyes are moved and independent shifts of attention (i.e. covert attention) are not used (Findlay & Gilchrist, 1998; Findlay & Gilchrist, 2003). Some accounts of visual search emphasize the use of covert visual attention. The study of covert visual attention is concerned with how information in a region of our visual field can be preferentially processed to the detriment of information in other regions, independent of eye movements (Posner & Cohen, 1984). While the idea that covert visual attention can move independently of eye movements is an intriguing idea and is theoretically important, the use of covert attention to explain visual search in natural settings can result in important factors being ignored such as: Different information is available depending on the orientation of the gaze (Bertera & Rayner, 2000; Findlay & Gilchrist, 2003), so where the eyes are pointing is important.

The movements of the eyes (and sometimes even head and body movements) are important for models of visual search in HCI. This is especially true due to the increasing size of computer displays and the increasing ubiquity of computing interfaces and hence the increased importance of where the eyes are physically pointing. Therefore, this work will emphasize the role of eye movements in visual search.

2.1. Previous Models of Visual Search in HCI

A variety of models have been developed to predict visual search behavior. Some models have been developed specifically to predict and explain performance in a narrow domain, such as graph perception (e.g. Lohse, 1993). Others have been developed to predict and explain the effects of specific visual features in a broad range of visual search tasks (e.g. Wolfe & Gancarz, 1996). The following is a brief overview of some previous modeling work to provide context for the remainder of the chapter.

Guided Search (GS; Wolfe, 1994; Wolfe & Gancarz, 1996) is a computational model of how visual features, such as color and orientation, direct visual attention. GS predicts that the order in which objects are visually searched is affected by the “strength” of objects’ visual features (e.g. their blueness, yellowness, steepness, shallowness, and so on), the differences between object features, the spatial distance between objects, the similarity to the target, and the distance of objects from the center of gaze (i.e. the eccentricity). GS predicts where the eyes move, when the eyes move, and what is perceived in each fixation.

The Area Activation Model (AAM; Pomplun, Reingold & Shen, 2003) is a computational model of how visual features direct visual attention. The AAM shares many characteristics with Guided Search, but differs in at least one important way. The AAM assumes that all objects near the center of gaze are searched in parallel and GS assumes that objects are searched serially. AAM predicts where the eyes move and what is perceived in each fixation.

Barbur, Forsyth, and Wooding (1990) propose a computational model to predict eye movements in visual search. The model uses a hierarchical set of rules to predict where people’s gaze will be deployed. Like the AAM, Barbur, et al.'s model assumes that all objects near the center of gaze are searched in parallel. It differs from the GS and AAM in that eccentricity is the only visual feature that determines where the gaze moves next. Barbur, et al.’s model predicts where the eyes move and what is perceived in each fixation

Understanding Cognitive Information Engineering (UCIE) is a computer model of human reasoning about graphs and tables (Lohse, 1993) based on *GOMS* (Goals, Operators, Methods, and Selection Rules; John & Kieras, 1996), an engineering model for predicting task execution time. UCIE extends GOMS with a model of visual search that includes the time to perceive objects, eye movements, and a limited memory for information. These processes provide constraints for the simulation of how people scan graphs and tables to answer questions about the graph or table. UCIE predicts where the eyes move, when the eyes move, what is perceived in each fixation, and how information is integrated across fixations.

All of the models above were built to predict a limited set of behavior. They are all “stand-alone” models that do a good job of predicting the behavior for reasonably well-defined tasks, but may not be appropriate for predicting a larger set of behavior of which visual search is only a part. An alternative approach to stand-alone models of visual search are models built in a more general framework that has been used to predict a wider range of behavior for a wider variety of tasks. One example of such a framework is EPIC.

EPIC (Executive Process-Interactive Control) is a framework for building computational models of HCI tasks that lends itself well to building models of visual search (Kieras & Meyer, 1997). EPIC provides a set of perceptual, motor, and cognitive constraints based on the psychological literature. Models of visual search built within EPIC tend to explain visual search as the product of cognitive strategies, perceptual constraints, and motor constraints.

2.2. Active Vision Theory

Active vision is the notion and collection of theories that asserts that eye movements are central to visual perception in general and for tasks such as visual search (Findlay & Gilchrist, 2003). Active vision poses four central questions: (a) *When* and why do the eyes move? (b) *Where* do the eyes move next? (c) *What* can be perceived when the eyes are relatively steady? (d) *What* information is integrated between eye movements?

When Do the Eyes Move?

When the eyes move from one visual element to another, or conversely how long the eyes linger on elements in a visual layout, is an important factor to consider in a model of active vision. Will the eyes remain on complex icons longer than simple icons? The time between eye movements is called saccade latency or fixation duration.

Four explanations of fixation duration control have been proposed in the literature (Hooge & Erkelens, 1996): (a) preprogramming-per-trial, (b) preprogramming-per-fixation, (c) strict process-monitoring, and (d) mixed-control. The first explanation, preprogramming-per-trial, is that the required fixation duration is estimated before the visual search task is initiated and this estimated fixation duration is used throughout the visual search task. The second explanation, preprogramming-per-fixation, assumes that fixation durations are dynamically estimated throughout a visual search task. If previous fixations were too short to perceive the stimuli before initiating a saccade, future fixation durations are lengthened. The third explanation, strict process-monitoring, is that fixation durations are not estimated, but rather directly determined by the time to perceive the fixated stimuli. The last explanation, mixed-control, assumes that saccades are sometimes initiated by the time to perceive the stimuli *and* at other times by previously estimated durations.

Existing computational models vary with respect to which explanation of fixation duration they instantiate. When eye movements are considered in Guided Search (Wolfe & Gancarz, 1996), fixation durations are fairly constant, with each fixation lasting 200 to 250 ms, and therefore can be considered to be an instantiation of the preprogramming-per-trial explanation. UCIE instead proposes a varying time for fixation durations based on the number, proximity, and similarity of all objects within a limited range of fixation center. Other models do not account for fixation durations, as so make no prediction about when the eyes move (e.g. Barbur et al., 1990; Pomplun et al., 2003).

Where Do the Eyes Move?

The order in which items are searched in a layout may have a large impact on usability and implications for the visual tasks that a layout will support well. Will a visitor to a web page look in the location the designer intends for a task?

The path the eyes follow is usually referred to as the scanpath. Research pertaining to the scanpath attempts to understand what factors guide visual attention or the eyes. A great deal of research has been conducted to determine the factors that influence the scanpath in visual search (see Wolfe & Horowitz, 2004 for a review).

Two influences on scanpaths are (a) guidance by features, or bottom-up guidance, and (b) guidance by strategy, or top-down guidance. The intrinsic features (e.g. color, size, shape, or text) of objects affect the order in which objects are visually searched. When features of the target are known and these features can be perceived in the periphery, this information can guide visual search. Most existing models of visual search use intrinsic properties to guide search in some way. For example, Guided Search 2 (Wolfe, 1994) builds an activation map based on the color and orientation of objects to be searched. Activation maps are spatial representations of where in the visual environment information exists. Visual search is then guided to the items in the order of greatest to least activation. Guided Search 3 (Wolfe & Gancarz, 1996) adds the additional constraint that objects closer to the center of fixation produce more activation. The Area Activation model (Pomplun et al., 2003) is similar to Guided Search 2, except that search is guided to regions of greatest activation instead of items.

Intrinsic features are not the only influence on the scan path, especially if (a) the peripherally available information cannot guide search or (b) the exact identity of the target is unknown. Strategic decisions, or top-down guidance, also influences the order in which objects are searched. For example, hierarchical menus have been found to motivate fundamentally different strategies than non-hierarchical menus (Hornof, 2004) and the ordering of menu items, either alphabetically or functionally, motivate fundamentally different strategies than randomly ordered menus that decreases search time (Card, 1982; Perlman, 1984; Somberg, 1987).

What Can Be Perceived?

What a user can visually perceive in an interface at any given moment is an important question that must be answered by a model of visual search. For example, will the user notice the notification that just appeared on their screen? Or, can the user perceive the differences between visited and unvisited links on a proposed web page? Most of the models previously reviewed make different assertions about the information perceived in each fixation and the region from which this information can be extracted.

One possible assumption about what can be perceived is that all objects within a fixed region can be perceived. Some models of visual search make this assumption. Barbur, et al. assume that all information within 1.2 degrees of visual angle of fixation center can be perceived (Barbur et al., 1990). UCIE (Lohse, 1993) assumes that all items within an unspecified radius are processed, but only the object of interest at the center of fixation is considered. Guided Search Wolfe & Gancarz, 1996) assumes that up to 5 objects near the center of fixation are processed during each fixation. The Area Activation model (Pomplun et al., 2003) assumes that all items within a “fixation field” are perceived. These fixation fields are two-dimensional normal distributions centered on the center of fixation and vary by the properties of the stimuli in the layout.

A straightforward model of visual search for HCI need only assume that a set number of objects or a set region is perceived during each fixation. This is what many existing models assume (Barbur et al., 1990; Hornof, 2004; Lohse, 1993; Wolfe & Gancarz, 1996). This simplifies the model, as only object location is required to determine which objects fall within the set region and are consequently perceived. A reasonable approximation for this region is one degree of visual angle radius, as this distance has been used to explain visual search for simple shapes (Barbur et al., 1990) and text (Hornof, 2004).

What Information Is Integrated Between Eye Movements?

Another important factor to consider in visual search is what information is integrated between eye movements. In other words, how does memory affect visual search? For example, when searching for a specific news article, will a user remember which headings have already been searched, or will the user repeatedly revisit them?

There are at least three types of memory that may affect the visual search process: visual working memory, verbal working memory, and spatial working memory (Logie, 1995). Interestingly, research has shown that when either verbal or visual working memory is used for a second (non-visual search) task, visual search is largely unaffected (Logan, 1978; Logan, 1979; Woodman, Vogel & Luck, 2001). However, the use of spatial memory (i.e. memory for locations in space) *does* appear to affect visual search (Oh & Kim, 2004). Other research suggests a possible use of spatial working memory in saccade selection: A study of the visual search in “Where’s Waldo?” scenes, in which a cartoon figure is hidden within complex scenes, found that saccades tend to be directed away from the locations of previous fixations (Klein & MacInnes, 1999).

In general, computational models of visual search do not incorporate limitations of memory for spatial locations. Many models of visual search assume a perfect memory for objects searched (Anderson, Matessa & Lebiere, 1997; Barbur et al., 1990; Byrne, 2001; Hornof, 2004; Kieras & Meyer, 1997; Pomplun et al., 2003; Wolfe, 1994). In general these models remember which items have been inspected and then do not re-inspect the objects unless all items have been searched without locating the target (i.e. search without replacement).

Active vision emphasizes the importance of eye movements. Active vision asserts that where and when the eyes move, and how information gathered during eye movements is utilized, is critical for understanding vision and, in particular, visual search. The literature reviewed suggests that no computational model of visual search has yet answered all of the questions of active vision. However, every question of active vision is addressed by at least one model. The proposed answers for the questions of active vision gleaned from the literature, along with empirical data, are used here to build an active vision model of visual search.

2.3. Computational Cognitive Modeling

Computational cognitive models are computer programs that behave like people, such as by simulating aspects of people's perceptual, motor, and cognitive processes. Cognitive modeling is used in two ways: (a) In a *post hoc* fashion to help explain how people performed a task. (b) In an *a priori* fashion to predict how people will perform a task. This article reports on post hoc modeling research that uses cognitive modeling to explain people's already observed behavior. Such post hoc cognitive modeling has been used to understand web link navigation behavior (Fu & Pirolli, 2007), driving behavior (Salvucci, 2001), and time interval estimation (Taatgen, Rijn & Anderson, 2007). Post hoc modeling is needed to inform predictive modeling. For example, the post hoc modeling of driving behavior (Salvucci, 2001) was used to inform the development of a predictive model of driver behavior while utilizing a cell phone (John, Prevas, Salvucci & Koedinger, 2004). The post hoc models developed in this article will inform predictive modeling.

Cognitive architectures provide a computational instantiation of psychological theory that are useful for modeling human performance both post hoc and a priori. The architecture constrains the construction of the models by enforcing human capabilities and constraints. Cognitive models consist of (a) a detailed set of if-then statements called *production rules* that describe the strategy used by the simulated human to carry out a task, (b) a set of hypothesized processors that interact with the production rules to produce behavior, and (c) parameters that constraint the behavior of the model (e.g. the velocity of a saccadic eye movement). While the parameters can be task-specific, the majority of the parameters are usually fixed across a wide variety of models. Cognitive models produce predictions of how a person may perform the task. The results of such simulations allow the testing of the theory instantiated in the models by comparing the performance against those observed from humans.

There is a special relationship between cognitive modeling and the study of eye movements. The eye movements provide a rich set of data for informing the construction and evaluation of the models. The data provide many constraints on the models, including the number, extent, sequence, and timing of eye movements. The models, in turn, provide a means for understanding and explaining the strategies and processes that produce the eye movement data. It is beneficial to use a modeling framework that makes explicit predictions of eye movements, such as the framework discussed next.

EPIC

EPIC (Executive Process-Interactive Control) is a cognitive architecture that instantiates and integrates theory of perceptual, motor, and cognitive constraints. Figure 1 shows the high-level components of EPIC (Kieras & Meyer, 1997). EPIC provides separate facilities for simulating the human and the task. In the task environment, a visual display, pointing device, keyboard, speakers, and microphone can be simulated. Information from the environment enters the simulated human through eyes, ears, and hands and moves into corresponding visual, auditory, and tactical perceptual processors.

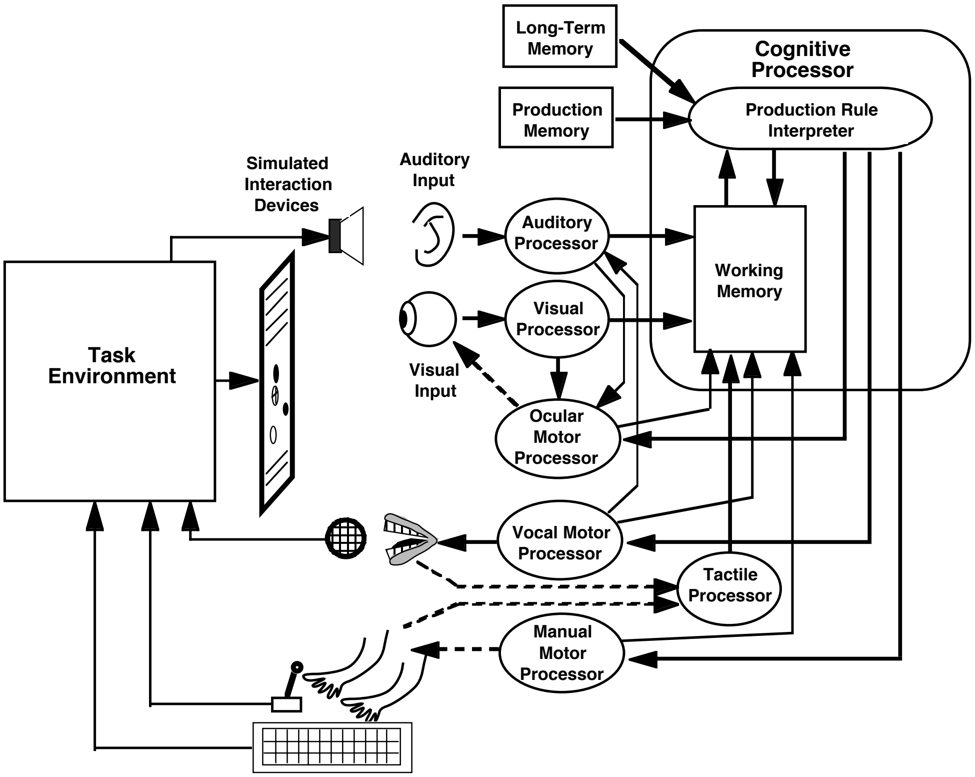


Figure 1. The high-level architecture of the EPIC cognitive architecture (Kieras & Meyer, 1997).

Information from the perceptual processors are deposited into working memory. Working memory is represented by a set of clauses that represent discrete facts about the world. In the cognitive processor, information in working memory interacts with a cognitive strategy, instantiated in production rules, to produce action through the ocular, manual, and voice motor processors. The motor processors control the simulated eyes, hands, and mouth to interact with the environment. All processors run in parallel.

The perceptual and motor processors impose many constraints on the simulated behavior that a set of production rules can generate. It is not possible for the architecture to generate just any arbitrary set of data. Perceptual processors determine what information in the environment is potentially available “downstream” to the cognitive processor, and when it will be available. Motor processors can affect the task environment and provide simulations of observable human data. Particularly relevant to the research presented here are the constraints imposed by the simulated eyes, as follows: The retina, the layer of cells at the back of the eye which detect light, provides more detailed processing of information in foveal (central) vision. The eyes take time to move to gather additional information. EPIC specifies retinal availability functions for different visual properties to simulate the limitations of the retina. For example, detailed information like text is only available within 1 degree of visual angle from the center of gaze. Other information, such as color, is available at a greater eccentricity. EPIC also simulates the ballistic eye movements, called saccades, to gather additional information. The cognitive processor sends commands to the ocular-motor processor to initiate eye movements. The ocular motor processor then prepares and executes the eye movements, imposing appropriate time delays for processing time and eyeball rotation.

EPIC's cognitive processor is a production rule interpreter. Production rules are if-then statements. The *if* part of each rule is the *condition*. The *then* part is the *action*. EPIC permits multiple production rules to fire in parallel. EPIC does not impose a central cognitive processing "bottleneck" but instead shifts serial processing limitations to the "peripherals", the motor and, to a lesser extent, the perceptual processors. Thus, an important aspect of modeling with EPIC is instantiating strategies that work with the constraints imposed by the peripheral processors, such as that the eyes can only be commanded to move to one location at a time.

The following is an example of a production rule that moves the eyes to a visual object whose information is in working memory. The IF part contains a number of clauses, each of which must exist in working memory for the rule to fire. When the rule fires, the THEN part of the rule commands the ocular motor processor to move the eyes to the ?Object variable, which will have been assigned to a random visual object with text in working memory.

```
// Look and point at a random word
(Move_eyes_to_random_word
IF
   (
       // This is the current goal and step
       (Goal Do Visual_Search Task)
    (Step Random Search)

       // The eyes are free to move
    // (i.e. they are not moving to another word)
    (Motor Ocular Processor Free)

    // Choose a random object with text
       (Visual ?Object Text ???)
    (Randomly_choose_one)
   )
THEN
   (
    // Move the eyes to the word
    (Send_to_motor Ocular Perform Move ?Object))
   )
```

EPIC and Visual Search

EPIC is particularly well-suited as a cognitive architecture for building models of visual search. Perhaps most importantly for active vision, EPIC simulates (a) perceptual availability as a function of an object’s eccentricity from the point of gaze and (b) eye movements as a means to get needed information into high resolution vision. Within the EPIC framework exists theory that constrains the simulation of (a) when visual features are perceived, (b) where the eyes move, (c) what can be perceived in each fixation, and

(d) how visual information can be integrated across fixations. These are all important for building a computational model of active vision.

When visual features are perceived: EPIC simulates the delay of sensory information in perceptual processors, namely sensory transduction time and perceptual recoding time. The encoding of visual objects and their properties into visual working memory takes time. If a visual property is available according to the availability function, that information travels through the visual sensory processor and the visual perceptual processor, each of which induces a delay, before being deposited into visual working memory. For example, if an object appears in the model's parafovea, the shape of that object would appear in the visual sensory store 50 ms after presentation and in the perceptual memory store (i.e. visual working memory) 100 ms after presentation. Different visual features have different delays.

Where the eyes move: The contents of working memory and the visual search strategies encoded by the analyst into the production rules determines where the simulated eyes move. The production rules must explicitly state under which circumstances the eyes are moved. When the contents of working memory satisfy the left hand side (the "if" side) of a production rule that moves the eyes, a motor movement command is sent to the ocular motor processor by the right hand side of the rules (the "then" side), which then initiates an eye movement to the object specified by the production rule.

What can be perceived: EPIC's *retinal availability functions* constrain the perception of visual information from the environment. The availability functions simulate the varying resolution of the retina, with greater resolution near the center of vision and lower resolution in the periphery. The retinal availability functions determine the eccentricity at which visual properties can be perceived. For example, text is available within one degree of visual angle from the center of fixation, roughly corresponding to foveal vision, whereas color is available within seven and a half degrees of visual angle.

What can be integrated across fixations: Memory decay time and the production rules determine what information is integrated across fixations. The perceptual parameters affect how information can be integrated across fixations. When the eyes move away from an object, one or more visual properties may no longer be available. The visual property decays from sensory memory and then from visual working memory. Production rules can extend the "life" of visual features and object identity by creating an amodal memory item called a "tag". Copying memory items into tag memory must be explicitly programmed into production rules that execute while the visual memory is still available to the production rules, before the item has decayed from visual working memory.

This research draws on visual search literature, specifically literature related to active vision and previous models of visual search. Issues central to the notion of active vision include when and where the eyes move, what is perceived, and how the information perceived is used over time. This paper will explore these issues through cognitive modeling using the EPIC cognitive architecture. All told, EPIC is very well-suited as a framework for simulating active vision in the context of HCI tasks. EPIC provides a theory of visual-perceptual and ocular-motor processing that is useful for guiding the development of active vision models of visual search. Now that the theoretical framework and background is in place, the next section will briefly present three experiments whose data is used to develop and validate the proposed model of active vision.

3. EYE TRACKING EXPERIMENTS TO DEVELOP THE MODEL

Throughout the remainder of this article, eye tracking data from three experiments are used to inform and validate a computational model of visual search. This section presents an overview of those three experiments. Each experiment investigated how a specific design decision affects users' visual search processes as revealed by reaction time and eye movements. All together, the experiments provide a useful set of data for building and refining an active vision model of visual search for HCI. The first experiment (Halverson & Hornof, 2004a; Halverson & Hornof, 2004b) investigated the effects of varying the visual density of elements in a structured layout. The second experiment (Hornof, 2004) investigated the effects of layout size and a visual hierarchy. The third experiment (Halverson, 2008) investigated how both semantic and visual grouping affect people's active vision.

All three experiments were conducted using a classic visual search experimental paradigm. The entire layout is displayed at the same moment, permitting any search order, and the trials are blocked by layout. Each trial proceeded as follows: The participant studied and clicked on the precue; the precue disappeared and the layout appeared; the participant found the target, moved the mouse to the target, and clicked on the target; the layout disappeared and the next precue appeared.

3.1. Mixed Density Task

The mixed density experiment was designed to explore how issues of text size and spacing should be incorporated into a general purpose, comprehensive, active vision model of visual search for HCI. The experiment is discussed in more detail in Halverson and Hornof (2004a; 2004b) and is presented here specifically with regards to developing a comprehensive model of active vision.

Layouts contained six groups of words. There were two types of groups: sparse groups containing five words, and dense groups containing 10 words. Both types of groups subtended the same vertical visual angle. There were three types of layouts: sparse, dense, and mixed-density. Sparse layouts contained six sparse groups. Dense layouts contained six dense groups. Mixed-density layouts contained three sparse groups and three dense groups. Figure 2 shows an example of a mixed-density layout. Twenty-four people participated in the experiment.

The results of the experiment suggest that people tend to search sparse groups first and faster. The search time data demonstrate that people spent less time per word searching sparse layouts. It appears that, with sparse groups, participants adopted a more efficient eye movement strategy that used slightly more and shorter fixations.

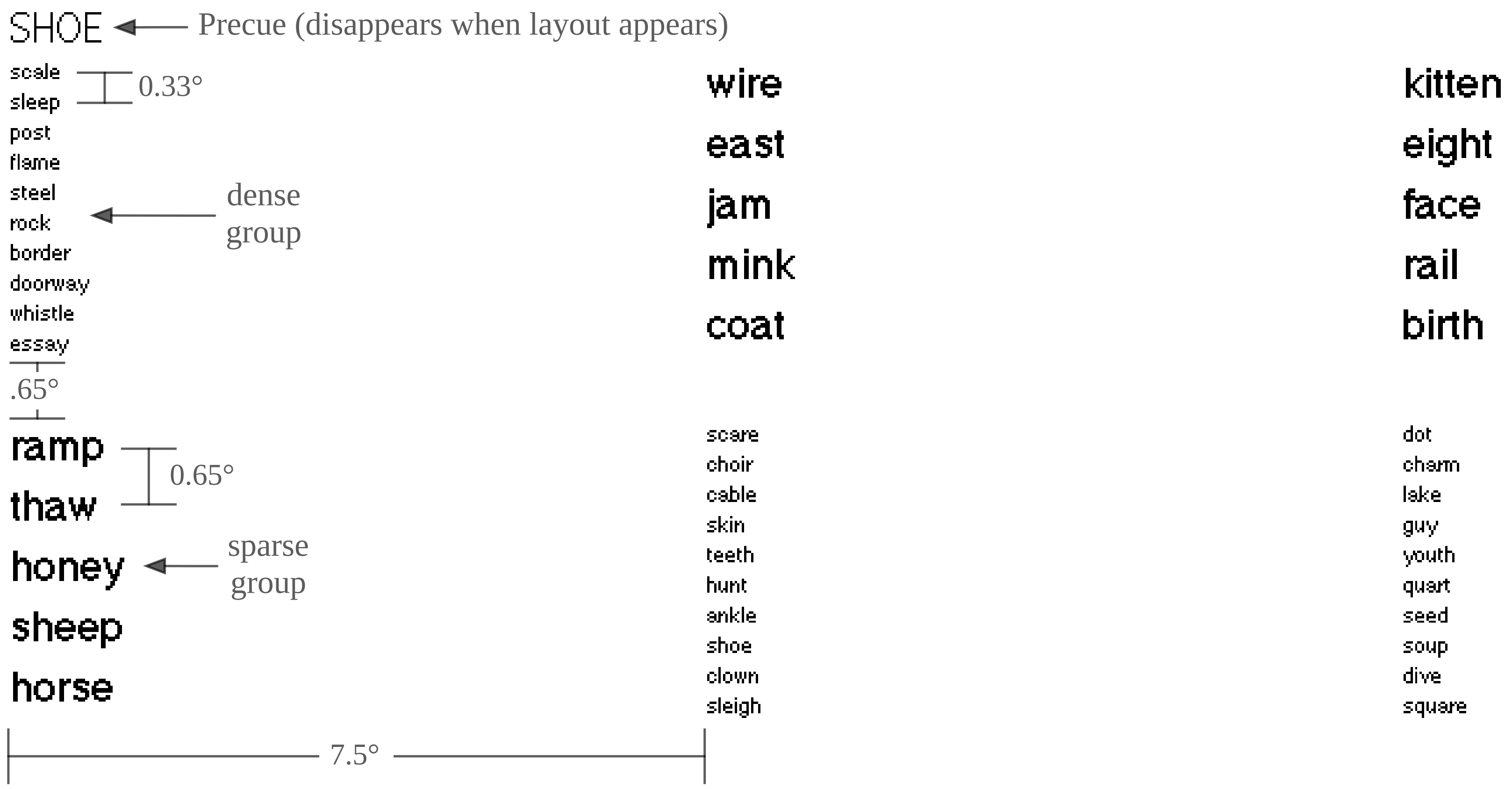


Figure 2. A mixed-density layout. All angle measurements are in degrees of visual angle.

3.2. CVC Task

The CVC (consonant-vowel-consonant) search task investigated the effects of layout size and a visual hierarchy (Hornof, 2004). The CVC task is called such because the task used three-letter consonant-vowel-consonant pseudowords (such as ZEJ) that controlled for word familiarity and other effects. The CVC task included layouts with and without a labeled visual hierarchy. Data from the tasks without a labeled visual hierarchy are used to inform the development of the model.

The CVC experiment was originally conducted by Hornof (2001) without eye tracking, and modeled by Hornof (2004). The experiment was run again by Hornof and Halverson (2003) to collect eye movement data to evaluate the models in more detail. Sixteen people participated in each study.

Each layout contained one, two, four, or six groups. Each group contained five objects. The groups always appeared at the same physical locations on the screen. Figure 3 shows a sample layout from the experiment. One-group layouts used group A. Two-

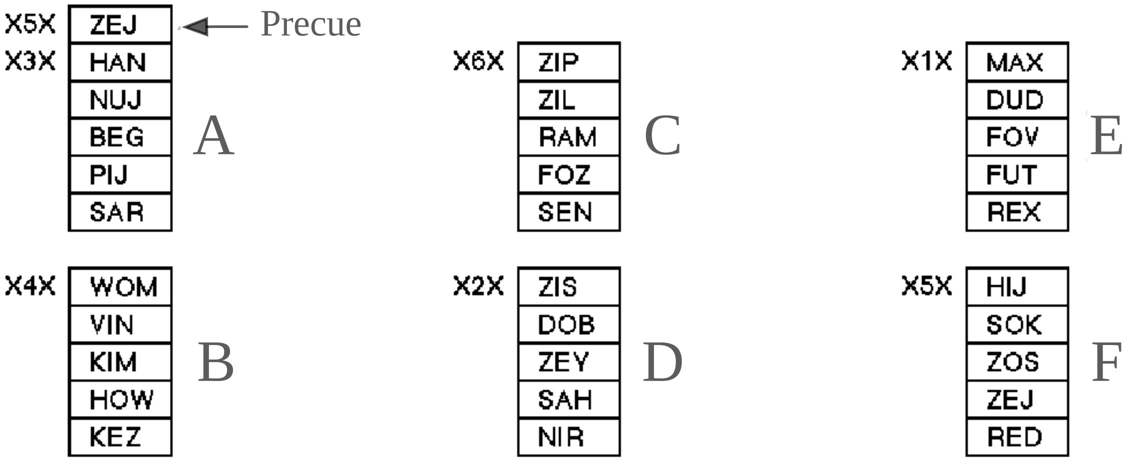


Figure 3. A layout without labels from Hornof's (2001) CVC search task.

group layouts used groups A and B. Four-group layouts used groups A through D. In the labeled condition, strings such as X5X were positioned next to the groups.

The results of the experiment show that people were able to search smaller layouts faster than larger, and labeled layouts faster than unlabeled. Further, people required disproportionately more time and fixations to find the target in large unlabeled layouts compared to small unlabeled layouts. It appears as if participants used a fundamentally different search strategy when a useful visual hierarchy was present.

3.3. Semantic Grouping Task

The semantic grouping experiment was conducted to determine how people search layouts that are organized based on the meanings of words. The experiment investigated effects of (a) positioning a target in a group of semantically similar words, (b) giving the groups identifying labels, and (c) further subdividing the layouts into meta-groups using graphic design techniques. The experiment is discussed in more detail in Halverson (2008).

Three variables were manipulated in the layouts: the semantic cohesion of groups of words, the presence of group labels, and the use of meta-groups. Groups of words were either semantically related (e.g. cashew, peanut, almond) or randomly grouped (e.g. elm, eraser, potato). Groups were either labeled or not. In some conditions, colored regions divided the groups into four meta-groups. When the meta-groups were used in a semantically-grouped layout, groups in the same colored region were further semantically related (e.g. nuts with candy, and clothing with cosmetics). Layouts always contained eight groups with five words each. Figure 4 shows a layout with semantically-cohesive groups, group labels, and meta-groups. Eighteen people participated in the study.

The results of the experiment show that people use the structure provided by the semantic content of the words in the layouts to guide their search. Further, people appear to judge the semantic relevance of all objects in semantically cohesive groups with a single fixation, regardless of the presence of labels. The semantic cohesion of words in a group can substitute, to some extent, for labels of those groups. The meta-groups did not appear to affect people's behavior.

Clearly, a good model of active vision need to accurately predict eye movements, as these are the most directly observable and measurable events of interest during a visual search task. In each experiment reported above, eye movements were recorded using a pupil-center and corneal-reflection eye tracker. In all analyses presented in the following sections, fixations are identified using a dispersion-based algorithm (Salvucci & Goldberg, 2000). Following established conventions, fixations are defined as a series of eye tracker gaze samples with locations within a 0.5° of visual angle radius of each other for a minimum of 100 ms. This is the data that our model will explain.

4. MODELING THE MIXED DENSITY TASK

The next few sections demonstrate a principled approach for building a model of visual search based on step-by-step improvement of the model using eye movement measurements to inform the refinement of model strategies and parameters. In this way, the model of active vision is developed, refined and enhanced by accounting for a

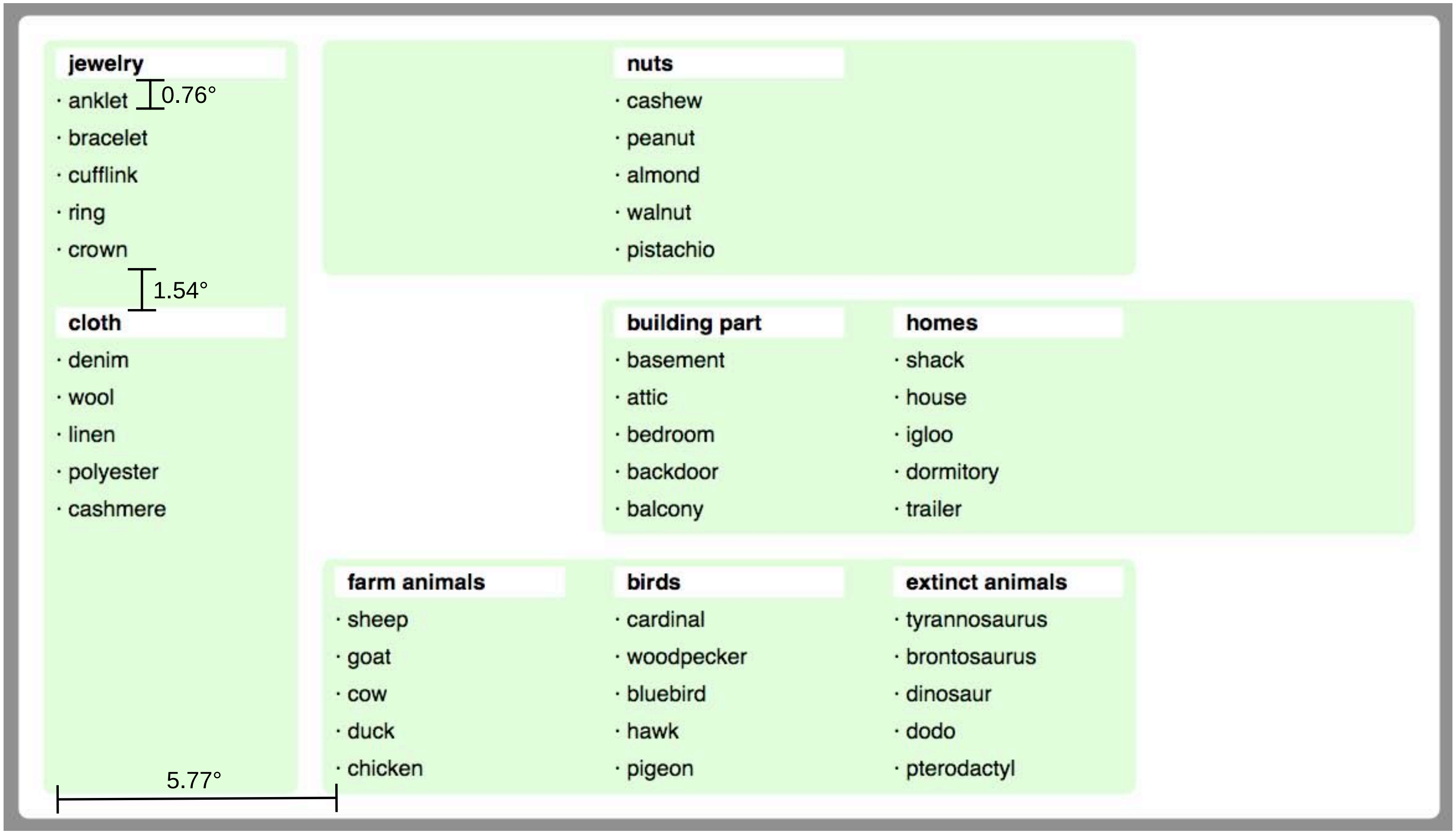


Figure 4. A layout from the semantic grouping experiment, with semantically-cohesive groups, group labels, and meta-groups.

growing set of eye movement data. Throughout the model development, the criterion that will be used to accept predictions by the model as “good enough” will be a 10% average absolute error (AAE) between the observed and predicted data. This is consistent with engineering practices (Kieras, Wood & Meyer, 1997).

All models were built using the EPIC cognitive architecture (Kieras & Meyer, 1997). Some modifications were made to EPIC’s visual processors during the iterative process of refining the models. These modifications are discussed in the following sections.

The development of the comprehensive model starts with a baseline model using reasonable initial assumptions, and progresses to a model that explains many features of the data with refinements related to what is perceived in a fixation and when saccades are initiated. The modeling focuses on issues raised by previous research on density, e.g., the number of items perceived per fixation (e.g. Bertera & Rayner, 2000; Ojanpää, Näsänen & Kojo, 2002), and other fundamental perceptual and ocular motor issues of visual search.

4.1. Baseline Random Search Model

The initial Baseline Model starts with the a small set of reasonable assumptions primarily supplied by the architecture. All of EPIC’s perceptual properties are left at established values, including: Text centered within 1° of the point of fixation will enter working memory after 149 ms; saccades take time to prepare, 50 ms if the previous saccade had the same direction and 150 ms if the previous fixation had a different direction and extent; saccades last 4 ms per degree of visual angle.

A couple of additional assumptions are extracted from the literature. First, the model searches “without replacement.” That is, any object for which the text has been perceived is excluded from being the destination point of future saccades. While there is some controversy over whether visual search proceeds without replacement (see for example Shore & Klein, 2000) or with replacement (i.e. amnesic-search; see for example Horowitz & Wolfe, 2001), the preponderance of evidence favors search without replacement. Second, saccade destinations are selected at random. It has been shown with previous modeling research that assuming a random search pattern provides a good initial prediction of search time (Hornof, 2004).

The Baseline Model includes a production-rule strategy that executes the task as follows: The model fixates and memorizes the target precue. As soon as the visual search layout appears, the model starts searching for the target. The model moves its eyes to a random word in the layout. As soon as the eyes arrive at the saccade destination, the model initiates the next eye movement to a random object whose text has not entered working memory. If at any time the target is identified, search is terminated, the eyes are moved to the target, and the target is clicked.

The Baseline Model includes nine production rules. Six of these production rules are used to memorize the precue, click on the precue, click on the target, and prepare for the next trial. Those actually involved in visual search, the core production rules of this random search model, are shown in Table 1. The rules have been slightly edited for readability, and some pre-conditions that do not directly contribute to the topics discussed

Table 1. The core production rules of the Baseline Random Search model during the step in which the search is conducted.

(Move_eyes_to_random_word IF ((Step Random Search) (Motor Ocular Processor Free)	If the ocular processor is free to prepare another eye movement
(Tag ?Word Object_Not_Fixated) (Randomly_choose_one))	for a randomly-chosen word that has not been fixated yet,
THEN ((Send_to_motor Ocular Perform Move ?Word)))	move the eyes to that randomly-chosen word.
(Set_items_as_fixated IF ((Step Random Search)	
(Visual ?Object Text ?T) (Tag ?Object Object_Not_Fixated))	If the text of an object has entered working memory
THEN ((Delete (Tag ?Object Object_Not_Fixated))))	then remember that the object has been fixated.
(Target_is_Located_Stop_Searching IF ((Step Random Search)	
(Visual ?Object Text ?T) (Tag Target_Text ?T)	If the text of an object that matches the target has entered visual working memory
THEN ((Delete (Step Random Search)) (Add (Step Move Gaze And Cursor To Target))	then stop the search and finish the trial.
(Add (Tag Target_Object ? Target_Object))))	

here have been removed. For example, the goal pre-conditions, (Goal Do Visual_Search Task), have been removed as they are all identical.

Figures 5, 6, and 7 show the predicted search times, fixation durations, and number of fixations per trial. As can be seen in Figure 7, this rudimentary model accurately predicts the observed number of fixations per trial for one condition. While this model incorrectly predicts the number of fixations for two of the conditions, the prediction for the sparse

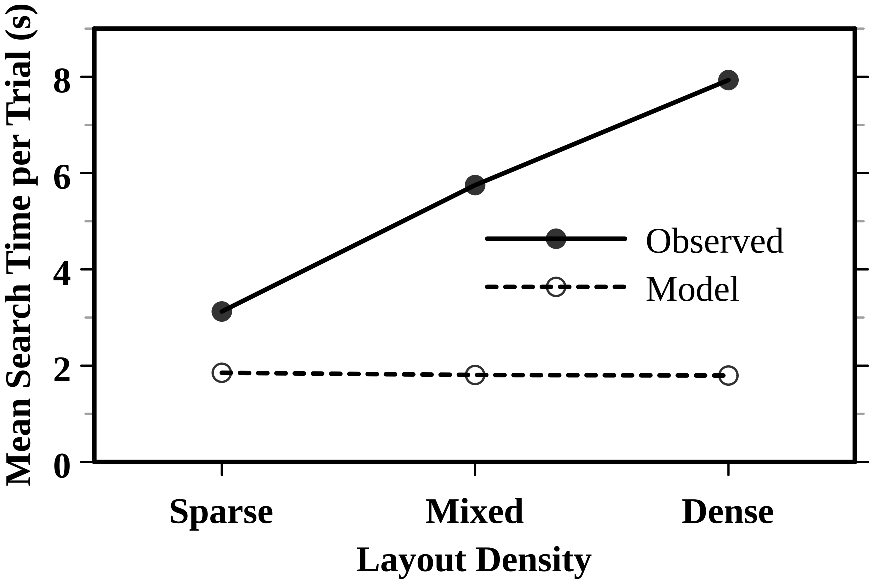


Figure 5. Mean search time per trial observed (solid line) and predicted (dashed line) by the baseline random search model for the mixed-density task. Average absolute error (AAE) = 62.1%

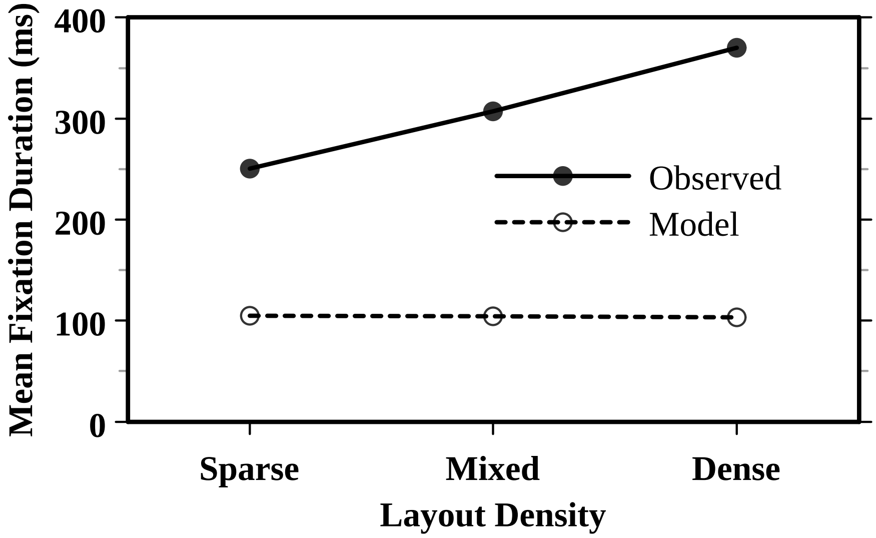


Figure 6. Mean fixation duration observed (solid line) and predicted (dashed line) by the baseline random search model for the mixed-density task. AAE = 65.5%

layouts is quite good. However, the Baseline Model is a poor predictor of human performance as most of the predictions are incorrect both in value and trend.

The model’s accurate prediction of the number of fixations per trial in the sparse layouts is promising and suggests that the purely random search model is a good starting

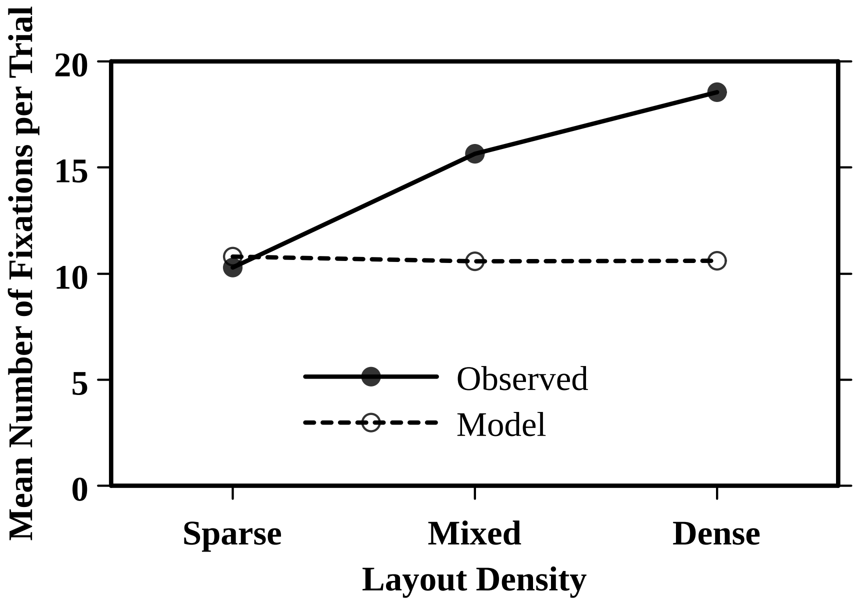


Figure 7. Mean number of fixations per trial observed (solid line) and predicted (dashed line) by the baseline random search model for the mixed-density task. AAE = 26.7%

point for modeling the characteristics of participant eye movements. While it is not likely that the participants are randomly selecting saccade destinations, such a strategy does provide an adequate starting point.

However, the fixation duration predictions show a strong need for an alternative means of initiating eye movements which is a key component of active vision. The greatest error is found in the fixation duration predictions, with a 65.5% AAE. The model initiates saccades as soon as possible given the constraints of the architecture. This proved to be too fast, as the model predicts a fixation duration of 100 ms, whereas the participants used fixations that were 250 ms or longer. Additionally, the participants used longer fixations for the denser text and the model did not. Therefore, the next round of modeling explores the initiation of saccades to improve the model's predictions of fixation durations, and the development of an active vision model of visual search for HCI proceeds.

4.2. Improving the Predictions of the Fixation Duration

The observed eye movement data from the Mixed Density experiment guides the model development. Because the predicted fixation durations have the greatest error of the eye movement measurements examined in the last section, this iteration of the model will focus on improving this prediction.

In the Baseline Model, saccades were initiated as soon as the previous saccade was complete. However, based on the results of the Mixed Density experiment, people appear to delay there saccades for a couple of hundred more milliseconds than predicted by the Baseline Model. This delay could be simulated a number of ways in the model. One approach would be for the production rules to directly set the fixation duration, though EPIC provides no such facility. Another would be to hold back each saccade until a certain amount of information is gathered from the currently fixated stimuli. We next look to the literature to improve the fixation duration predictions.

The model is developed considering the four explanations of fixation duration described by Hooge and Erkelens (1996), discussed earlier: preprogramming-per-trial, preprogramming-per-fixation, strict process-monitoring, and mixed-control. Preprogramming-per-trial and preprogramming-per-fixation do not appear to explain the mixed-density data very well. As shown in Figure 6, the observed fixation duration used in dense groups is always longer than the duration used in sparse groups, suggesting that the density (perhaps discriminability of the text) is driving the fixation durations. However, it could be that the participants were preprogramming fixation durations based on the current task. Then how should a model of visual search predict fixation durations? One way to answer this is to look for a parsimonious explanation with the architecture.

Newell (1990) encouraged researchers building cognitive models to “listen to the architecture.” By this, he meant researchers should develop parsimonious solutions to modeling a phenomenon that are in accordance with basic principles encoded in the architecture. The theory instantiated in EPIC lends itself to a process-monitoring explanation of saccade initiation, as the timing and retinal availability of visual features can be used in a straightforward manner to instantiate process-monitoring. Further, instantiating the preprogramming of saccade initiation would require additional mechanisms and parameters that are not required with the process-monitoring strategy and would therefore decrease the parsimony of the model. A preprogramming saccade initiation strategy might require a theory of time perception (such as Taatgen et al., 2007) that could be used to predict saccade time intervals. The introduction of such a temporal processor for eye movements may introduce unnecessary complexity to the model. Therefore, the current modeling effort explores the use of a strict process-monitoring to explain fixation durations, and doing so finds a fit between the theory (strict process-monitoring) and the architecture (EPIC).

Strict Process-Monitoring Model

The strategy rules are modified to include a strict process-monitoring strategy. Figure 8 shows a flow-chart based on the production rules. This strategy initiates saccades only

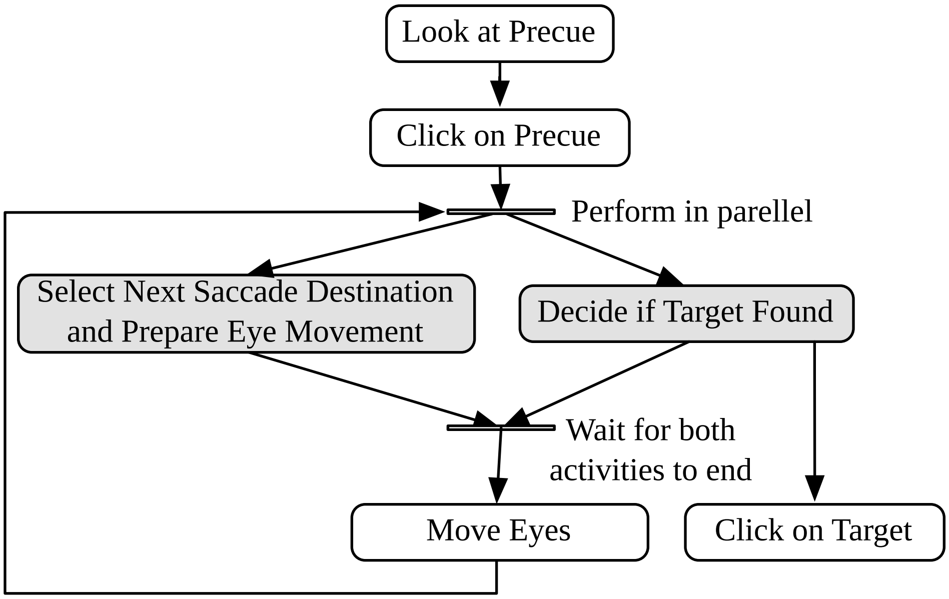


Figure 8. Flow chart of the production rules for an instantiation of the strict process-monitoring strategy.

after the text property for the current saccade destination becomes available and a decision is made whether the target has been found or not. This strategy is implemented by modifying one production rule, Move_eyes_to_random_word, and adding two other production rules, Prepare_eyes_to_random_word and Target_Not_Found_Yet. These rules are shown in Table 2.

EPIC’s visual perceptual processor is modified to accommodate a strict process-monitoring, as follows. The default recoding time for text is a constant 100 ms. This was modified to fit the human data. As shown in Figure 6, the observed fixation duration in the dense layouts was over 100 ms longer than in the sparse layouts. To model this, a stepped recoding function is introduced to the visual perceptual processor to calculate the perceptual time for a feature based on the proximity of adjacent items. If an object’s closest neighbor is closer than 0.15° of visual angle (a dense object), the text recoding

Table 2. New production rules in the Strict Process-Monitoring model.

(Prepare_eyes_to_random_word IF ((Step Prepare Eye) (Motor Ocular Processor Free)	
(Tag ?Word Object_Not_Fixated) (Not (Tag ?Word Current Destination)) (Randomly_choose_one))	Select a random word that, according to the tag memory, has not been fixated and that is not the destination of the current saccade.
THEN ((Send_to_motor Ocular Prepare Move ?Word) (Delete (Step Prepare Eye)) (Add (Step Move Eye)) (Add (Tag ?Word Next Destination))))	Prepare to move the eyes to that word, and remember that this word is the next saccade destination.
(Move_eyes_to_random_word IF ((Step Move Eye) (Motor Ocular Processor Free)	
(Not (Tag ??? Current Destination))	If the current saccade destination has been processed (see next rule)
(Tag ?Word Next Destination))	and a new word has been selected as the saccade destination (see the previous rule)
THEN ((Delete (Step Move Eye)) (Add (Step Prepare Eye)) (Send_to_motor Ocular Perform Move ?Word) (Delete (Tag ?Word Next Destination)) (Add (Tag ?Word Current Destination))))	then move the eyes to the next saccade destination.
(Target_not_found_yet IF ((Step Check Current Destination)	
(Tag ?Word Current Destination) (Visual ?Word Text ?X)	If the text of the current saccade destination has been perceived
(Tag Target_Text ?T) (Not (Visual ??? Text ?T)))	and the target has not been found
THEN ((Delete (Tag ?Word Current Destination))))	then allow another eye movement by removing the tag memory of the current destination.

time is 150 ms. Otherwise the text recoding time is 50 ms. The differentiation of the time to recode the text remains true to a principle in the EPIC architecture in which the processing of visual objects is differentiated based exclusively on the features of those visual objects. These changes instantiate an existing theory of saccade initiation into the active vision model of visual search.

As shown in Figure 9, the Strict Process-Monitoring model improves the predictions of fixation durations. Delaying the initiation of saccades until after the text information had entered working memory and increasing recoding time for dense objects results in a fixation durations similar to those in the observed data.

The predicted mean search time only improved slightly. As seen in Figure 10, there is now a very slight upward trend in the search time. However, the slope of the predicted search time line is not nearly as steep as the observed search time line. As shown in

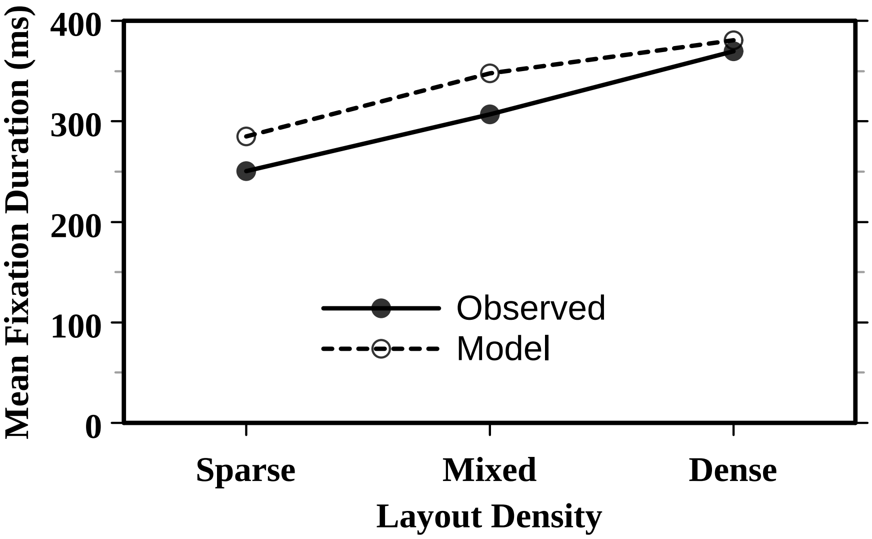


Figure 9. Mean fixation durations observed (solid line) and predicted (dashed line) by the strict process-monitoring model for the mixed-density task. AAE = 10.0%

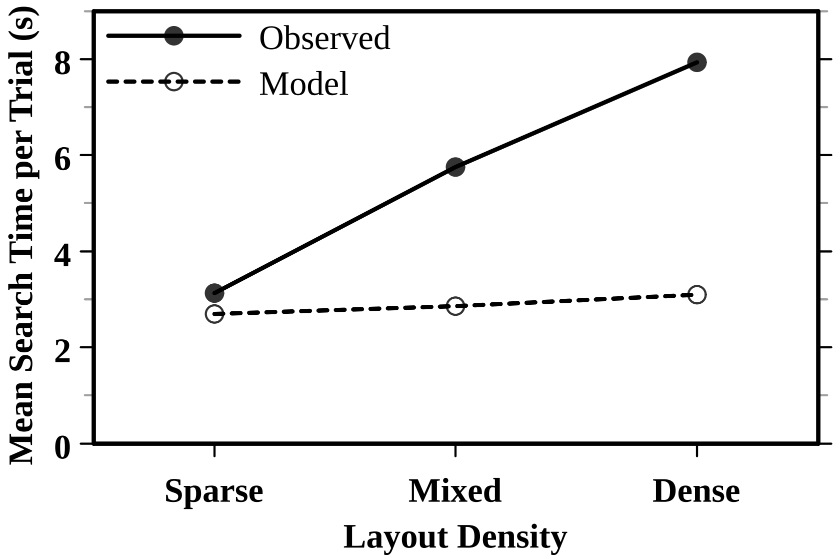


Figure 10. Mean search time per trial observed (solid line) and predicted (dashed line) by the strict process-monitoring model model for the mixed-density task. AAE = 41.7%

Figure 11, the model still does not make more fixations in layouts with dense objects, as is seen in the observed data.

The Strict Process-Monitoring model suggests a number of components that should perhaps be included in a comprehensive computational model of active vision for HCI. Using a strict process-monitoring strategy for saccade initiation provides straightforward, plausible predictions. The monitoring strategy is well supported by EPIC as the availability of features through the various visual processors produces a delay that is slightly less than the observed mean fixation duration in humans. Further, after including the time to decide, prepare, and execute the eye movement, the eye movement latency predicted by EPIC matches the mean fixation duration of humans very well.

While other explanations of fixation duration control (Hooge & Erkelens, 1996) could possibly be used to explain the observed fixation duration data, doing so would require

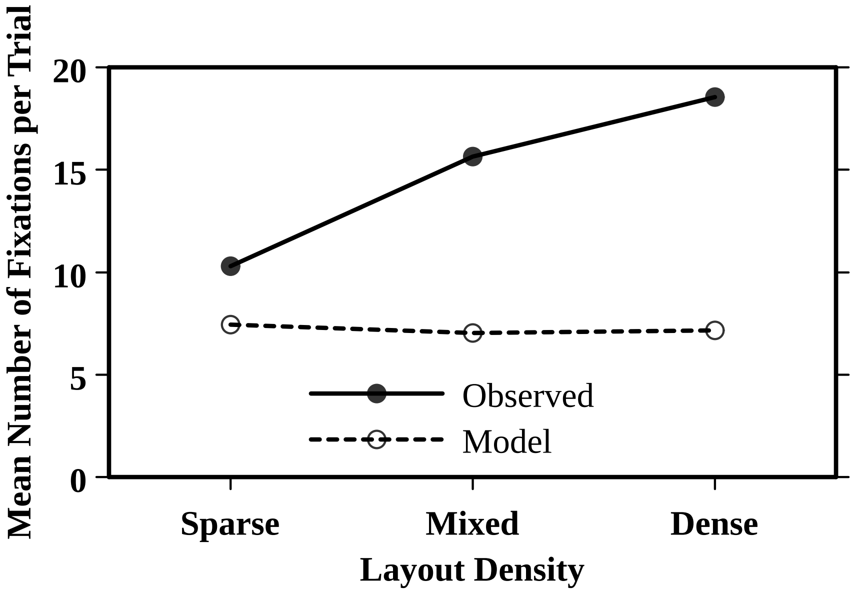


Figure 11. Mean number of fixations per trial observed (solid line) and predicted (dashed line) by the strict process-monitoring model model for the mixed-density task.
AAE = 48.1%

introducing addition processes and many more parameters in to the EPIC cognitive architecture. Therefore, the process monitoring strategy will remain as an important component of the model, as it is both parsimonious, predicts the observed data very well, and is supported by the literature.

Since the greatest error now lies in the predicted number of fixations per trial, the next model focused on improving the number of fixations predicted by the model.

4.3. Improving the Predictions of the Number of Fixations

The number of fixations predicted by the model so far was largely determined by a parsimonious assumption about the area in which text is perceived. The assumption is that all text within the fovea (1° of visual angle) was perceived during each fixation. This results in the model perceiving two to three sparse objects, or five to seven dense objects, in each fixation. Consequently, the model was able to perceive all items in a layout with an equal number of fixations, regardless of the layout density. The observed data suggest that humans do not do this. People require more fixations for dense text. The number of fixations predicted for dense objects can be increased in a number of ways. One way is to reduce the region within which dense text can be perceived. Another is to reduce the probability of correctly perceiving text based on the size or spacing of the text. Both methods were tested in the models.

Previous research suggests that predicting that two to three objects are processed per fixation can help account for the observed number of fixations in a search task (Hornof & Halverson, 2003). The EPIC visual perceptual processor availability function is modified so that 2 to 3 words were processed per fixation regardless of density. Sparse words within 1° of visual angle are processed (this matches EPIC's default availability function for text). Dense words within 0.5° of visual angle are processed. This modification results in a much better fit for the predicted number of fixations per trial. However, as shown in Figure 12, the model still under-predicts the number of fixations per trial in all layouts, and so this words-per-fixation approach is passed over in favor of the probability-of-encoding approach discussed next.

Text-Encoding Error Model

To adjust the probability of incorrectly encoding text, EPIC's perceptual processor is modified again so the probability of encoding the text of an object is based on the distance to the nearest neighboring object. Using the distance to the nearest neighboring object is one of several ways to measure density. One advantage of this measure for ease and practicality in predictive modeling is that it only requires the position of each item on

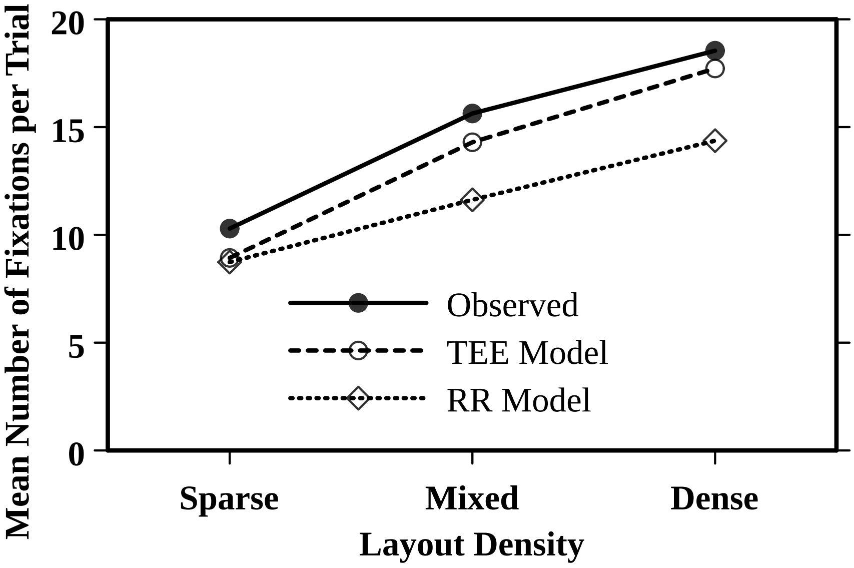


Figure 12. Mean number of fixations per trial observed (solid line), predicted by the Text-Encoding Error (TEE) model (dashed line with circles), and the Reduced-Region (RR) model (dashed line with squares) for the mixed-density task. The AAE of the TEE model is 8.8% and the RR model is 21.1%.

the screen. If an object's closest neighbor is 0.15° of visual angle away or more (sparse text), the probability of the model incorrectly perceiving the text is 10%. Otherwise, the probability of the model incorrectly perceiving the text is 50%. These probabilities were chosen because they result in two to three items, on average, perceived per fixation across densities, which appears to be the right number of items per fixation to explain the human data.

If the text of an object is incorrectly encoded, that object is marked as fixated just like objects that are encoded properly. This makes it possible for the target to be missed even if fixated. If the entire layout is searched (i.e. all objects have been marked as fixated) without finding the target, the model marks all objects as unfixated. This results in the visual search process starting over with the eyes starting at the last object fixated.

As seen in Figures 12 and 13, with text-encoding errors introduced to the model, the predicted number of fixations and the predicted search time improved considerably. The average absolute errors for the two measures are 8.8% and 6.5%. The number of fixations per trial now closely approximates the observed data. The accuracy of six predicted data points across the two groups is greatly increased by adding one perceptual parameter, a seemingly parsimonious improvement. Additionally, the modification made to the text-encoding property remains true to a principle in the EPIC architecture in which the processing of visual objects is differentiated based on the characteristics of visual objects as opposed to global parameter settings.

The modeling suggests that the use of encoding errors is a good method to simulate the perceptual constraints of density, at least for the perception of text in the current task. When all items in a fixed or varying region are perceived in every fixation, the model underpredicts the number of eye movements the humans need to find the target. When the

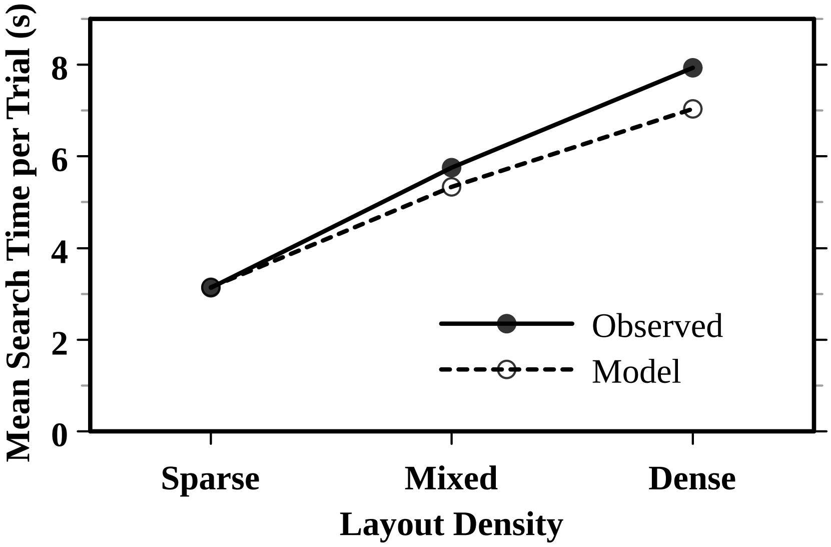


Figure 13. Mean search time per trial observed (solid line) and predicted (dashed line) by the TEE model for the mixed-density task. AAE = 6.5%

model is modified to include the possibility of misperceiving text, the model correctly predicts the number of fixations used in each layout.

4.4. Discussion

The process-monitoring strategy of saccade initiation instantiated in this model not only accounts for fixation durations in a straightforward and parsimonious manner but also suggests *when* saccade destinations are selected as is needed in a good model of active vision. In the model, saccades are initiated as soon as the relevant visual features (i.e. features that identify the target, like text) of the currently fixated objects enter working memory and a decision is made as to whether the target has been found or not. The observed fixation durations can be explained by such a model.

The modeling suggests that the use of encoding errors better simulates the perceptual constraints of density than changing the size of the region in which text can be perceived. One means of accounting for the number of fixations in a visual search of words is to limit the number of words perceived per fixation to two to three on average. Hornof (2004) found that limiting the number of objects perceived per fixation to two to three items helped predict observed search times. The same assumption here helps to predict the search time and number of fixations, but only when the number of items processed in each fixation is constrained by perceptual encoding errors and not by limiting the region in which objects are perceived.

The modeling of the mixed density task eye movement data provides two preliminary components that should be included in an active vision model of visual search for HCI, which are (a) initiating simulated saccades using a strict process-monitoring strategy, and (b) simulating what can or cannot be processed within a fixation using a region centered on the point of gaze with a probability of incorrectly encoding a small percentage of the objects in that region. The next section continues the development of the model by providing strategies and constraints to simulate where the eyes move next and what information is integrated across fixations. In addition, the instantiation of the strict process-monitoring strategy is refined.

5. MODELING THE CVC SEARCH TASK

A comprehensive model of active vision will need to account for how a person would deploy their active visual system to navigate a wide range of visual layouts and visual features. The progression towards a model of active vision continues here with the modeling of a second set of data, the CVC search task that was initially modeled by Hornof (2004). It is useful to return to the CVC search task because there are adequate differences between the mixed density task and the CVC search task, and differences between the corresponding models developed for each of the two tasks. For example, in the CVC task, the number of items in the layout varied with the size of the layouts and not as function of the density of the text in the layout as in the mixed density task. The original model for the CVC task pushed the eyes through the task using a fundamentally different, perhaps more rigid, strategy than developed for the mixed density task. This stage of the modeling primarily focused on two issues — evaluating assumptions in the active vision model and further developing the active vision model to also account for this previous set of data, but with a more general, flexible strategy than the original model.

5.1. Improving Saccade Distance

The model developed in the previous sections included a simplifying assumption that saccade destinations are selected at random from all items on the screen. This assumption was good enough to predict the mean search times, the mean number of fixations per trial, and the mean fixation durations. However, as it is unlikely that people select saccade destinations at random, and therefore the model is refined to more accurately simulate people's selection of saccade destinations.

One job of the human active vision system is to decide which objects to fixate. Though a completely random search strategy is very useful for predicting the mean layout search time, people do not search completely randomly. Instead, people move their eyes to objects that are relatively nearby more often than objects across the layout. Saccade destinations tend to be based on proximity to the center of fixation when the target is not visually salient (Motter & Belky, 1998).

Previous research supports the idea that people tend to fixate nearby objects. The original CVC task model suggests that moving to nearby objects is a reasonable strategy that explains the data. The best fitting model for the CVC task data in Hornof (2004) uses a strategy that moves the eyes a few items down each column of words on each saccade. While the strategy of the best fitting model did a good job, a more general strategy for saccade destinations is needed. Fleetwood and Byrne's model of icon search (2006) moved visual attention to the nearest icon that matched one randomly chosen feature of the target icon. In an effort to increase the fidelity of the model presented in the last section and to account for human active visual processes in a more general way than previous models, the development of an active vision model next explores the role of proximity in visual search.

Fixate-Nearby Model

The strategy used by the Text-Encoding Error model is next modified so that saccades are more likely to land on nearby items. Rather than searching randomly or following a prescribed search order, as with previous models, the strategy now chooses saccade destinations with the least eccentricity (distance from the eye position). To account for variability in saccade distances, as observed in Motter and Belky (1998), noise is added to the model's process of selecting the next saccade destination as follows: (a) After each saccade, the eccentricity property of all objects is updated based on the new eye position. (b) The eccentricity is scaled by a fluctuation factor, which has a mean of one and a standard deviation of 0.3 (determined iteratively to find the best fit of the mean saccade distance). This scaling factor is individually sampled for each object. (c) Objects whose text has not been identified and are in unvisited groups are marked as potential saccade destinations (i.e. search without replacement). (d) The candidate object with the lowest eccentricity that is not immediately surrounding the point of fixation is selected as the next saccade destination. The relevant new production rules is shown in Table 3. In order to support this rule, a new production rule predicate `Least` was implemented in EPIC to determine the object with the least eccentricity.

Table 3. New production rule in the Fixate-Nearby model.

(Prepare_eyes_to_nearest_object IF ((Step Prepare Eye) (Motor Ocular Modality Free)	If the ocular modality is free
(Tag ?Word Object_Not_Fixated) (NOT (Tag ?Word Current Destination))	and a word has not been fixated and is not the destination of the current saccade
(Visual ?Word In_Group ?Group) (Tag ?Group Unvisited)	in a group that had not been visited
(Visual ?Word Eccentricity ?ecc) (Greater_than ?Ecc 1.0) (Least ?ecc))	that has the least eccentricity and is not too close to the current fixation location
THEN ((Send_to_motor Ocular Prepare Move ?Word) (Delete (Step Prepare Eye)) (Add (Step Move Eye)) (Add (Tag ?Word Next Destination))))	then prepare to move the eyes to that word.

The strategy used by the model is also modified to reduce how often the model will revisit groups before visiting the rest of the layout. While the participants did revisit groups on occasion, approximately once every one to four trials, the majority of these revisits occurred either (a) after all groups had been visited once, or (b) because the target was overshot, resulting in a fixation in another group before refixating the target. One possible explanation for the low rate of revisits is that people tend to remember the regions they have explored. The model takes a straightforward approach to explain this behavior: A constraint is added to inhibit group revisits until the entire layout had been searched. This constraint is shown in the Prepare_eyes_to_nearest_object production rule in Table 3. Without this constraint, the model is much more likely to revisit a group than found in the observed data.

In an effort to explain the eye movement data and to depict the human information processing that is not directly observable, two mechanisms are introduced to the Fixate-Nearby model: (a) noisy saccades to nearby objects and (b) inhibition of group revisits. These two mechanism may interact to produce the same effect as the encoding errors introduced while modeling the mixed density search task. If the noise in the saccade selection strategy results in the gaze moving to another group before all words in the current group have been processed, the target can get passed over. Missing the target via encoding errors was used in the Text-Encoding Error model to explain the additional saccades sometimes required to re-examine the layout. Initial exploration of this reveled that removing the text-recoding errors resulted in a substantial decrease in the accuracy of the predictions for the number of fixations per trial. The Fixate-Nearby model *without* the encoding errors results in an AAE of 14.3% for the fixations per trial, which is not acceptable. Therefore, text-encoding errors are left in the model.

Figures 14, 15, and 16 show the predictions made by the Fixate-Nearby model as well as those made by the original CVC task model (Hornof, 2004). As shown in Figure 14, the model predicts the mean saccade distances very well, with an AAE of 4.2%, a considerable improvement over the AAE of 43.3% in the original model. As shown in Figure 15, the model also predicts the mean number of fixations per trial well, with an AAE of 4.2% a considerable improvement over the AAE of 37.8% in the original CVC model. As shown in Figure 16, the new model also does a good job of predicting the observed scanpaths. The figure shows the three most frequently observed scanpaths, and how the new model predicts the observed scanpath frequencies better than does the

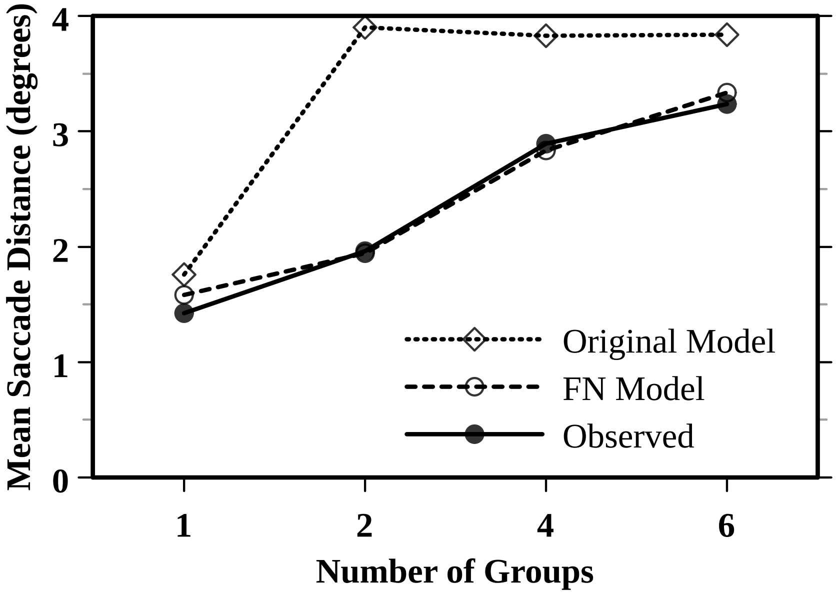


Figure 14. Saccade distance observed in the CVC search task (solid line), predicted by the original CVC search task model (dashed line with squares), and predicted by the Fixate-Nearby (FN) Model (dashed line with circles). The AAE of the original model is 43.3% and of the FN model is 4.2%.

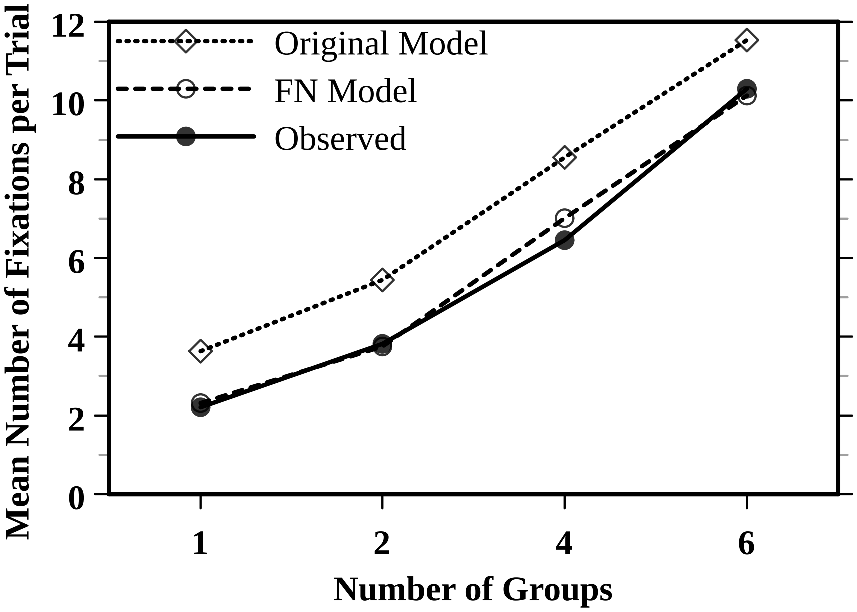
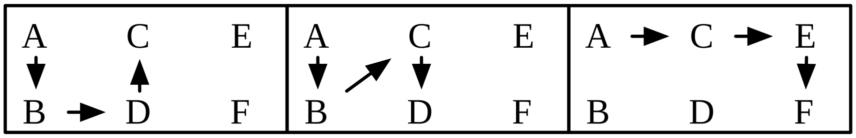


Figure 15. Fixations per trial observed in the CVC search task (sold line), predicted by the original model (dashed line with squares), and predicted by the Fixate-Nearby (FN) model (dashed line with circles). The AAE of the original model is 37.8% and of the FN model is 14.3%.



Observed:	30%	18%	11%
FN Model:	30%	21%	1%
Original Model:	0%	70%	1%

Figure 16. The most commonly observed scanpaths in the CVC search task in six-group layouts and how often each path was taken by the participants (observed), and predicted by the original model and predicted by the Fixate-nearby (FN) model. The dashed boxes emphasize the good predictions of the FN model.

original model.

One way that the new model improves over the original CVC model (Hornof, 2004) in predicting the number of fixations is that the new model adopts the strict process-monitoring strategy discussed in the previous section. The original model did not have such a strategy and would thus consistently move the eyes past the target before realizing the target had been fixated, thus requiring an additional two fixations on most trials. This overshooting of the target was not observed in the human data. The process-monitoring strategy is carried forward in subsequent models discussed here on the way to a comprehensive model of active vision for HCI.

These results reinforce the findings presented earlier with the mixed density task that people occasionally miss the target, even when looking directly at it. A failure rate of approximately 10% predicts human performance in this respect across multiple tasks. The increased accuracy in the model’s predictions and the similarity between the best-fitting text-encoding failure rate found here and the rate found in past research provides support for the use of the text-encoding failure rate parameter.

Results from this modeling suggest that people select saccade destinations partly based on eccentricity from the center of fixation. The selection of saccade destinations based on proximity results in a good fit of both the mean saccade distance and the scan paths that people used in this task. The original CVC model (Hornof, 2004), which moves the eyes a few words down each column on each saccade, predicts saccade distances much larger than is seen in the observed data. Additionally, the original model predicts little difference based on the size of the layout. When the saccade destination selection uses proximity, the effect of the size of the layout on observed saccade distances seen in Figure 14 is accounted for. Further, the two most frequent scanpaths, which account for nearly half of all observed scanpaths, are matched very well by the model that uses proximity.

This Fixate-Nearby strategy used in the model has a couple of benefits for predicting visual search compared to models whose predictions are based on particular visual structures or saliency of visual features. First, a predictive tool using this strategy need only encode the location information from a device representation. This is beneficial if other properties in the layout are unknown or difficult to automatically extract from the device representation. Second, this search strategy can be used when visual saliency alone cannot predict visual search, as is the case with goal-directed search (Koostra, Nederveen & de Boer, 2006). Unlike the original CVC model (Hornof, 2004), the Fixate-Nearby Model does not require a predefined notion of how the eyes will move through the layout to predict the observed scanpaths, which might be difficult for a model to predict for any arbitrary visual layout.

While the fidelity of the model improved overall, a problem was found with the implementation of the strict process-monitoring strategy when used with fixate-nearby strategy. This problem is discussed next as we continue to refine our computational model of active vision.

5.2. Revisiting the Predictions of the Fixation Duration

In the modeling of the mixed density task, it was found that a strict process-monitoring strategy predicted people’s fixation durations well. However, the particular implementation of the strict process-monitoring used was designed to work in a model in which saccade destinations were selected at random. The next saccade destination was prepared in advance of the decision to move the eyes and so the eyes could move to the next saccade destination shortly after the objects from the current fixation were processed. However, the Fixate-Nearby model requires that the model decides where to move the eyes next after the eyes have arrived at the current saccade destination, because the decision is based on the eccentricity of objects relative to the current fixation location. Therefore, the preparation of the eye movement must occur later than in the previous implementation of the process-monitoring strategy. The development of an active vision model of visual search for HCI proceeds to integrate the Fixate-Nearby strategy and the Process-Monitoring strategy.

Process-Monitoring Revisited Model

To address issues identified with the previous implementation of the process-monitoring strategy, a new saccade initiation strategy is proposed and implemented in this iteration of the model. This new process-monitoring strategy differs from the previous strategy in two ways: First, ocular motor movement preparation is removed from the EPIC architecture and is replaced in the model by a multi-stage process for selecting the saccade destination. As identified in research by Kieras (2003), only a constant motor movement initiation time (50 ms) is required to correctly simulate eye movement preparation. Second, multiple stages are used in selecting the saccade destination. In the first stage, different sub-strategies can each nominate saccade destinations. In the second stage, one of the nominated saccade destinations is selected. Table 4 shows rules from both states of this strategy.

Figures 17 and 18 show the predictions made by the revised model as well as by the original CVC model (Hornof, 2004) As shown in Figure 17, the Process-Monitoring Revisited Model predicts the fixation durations for unlabeled layouts very well, with an AAE of 4.6%. The new implementation of the strict process-monitoring strategy may accurately represent human active vision processes. As shown in Figure 18, the model also predicts the observed search time well, AAE = 9.7%. While the original CVC model predicted the search time slightly better than the active vision model, the active vision model still predicts the search time within our intended AAE of 10%. Additionally, as shown in this and previous sections, the active vision model predicts the eye movement data better than the original CVC model.

5.4. Discussion

The active vision model does a good job of predicting the search time, number of fixations, fixation duration, saccade distance, and scanpaths for two tasks. The model does so primarily by employing four constraints and associated visual features: (a) a

Table 4. New production rules in the Process-Monitoring Revisited model. The first two rules fire in parallel.

(Nominate_Unlabeled_All_Unidentified IF ((Step Nominate) (Motor Ocular Modality Free)	If the eyes have stopped moving and it is time to nominate fixation locations
(Tag ?Word Object_Not_Fixated) (Not (Tag ?Word Current Destination)) (Visual ?Word In_Group ?Group) (Tag ?Group Unvisited))	and an item not previously fixated is in a group that has not been visited
THEN ((Add (Tag ?Word Nominee Object))))	then nominate that word.
(Nominate_control IF ((Step Nominate) (Motor Ocular Modality Free))	If the eyes have stopped moving and a nomination is in progress
THEN ((Delete (Step Nominate)) (Add (Step Move))))	then change the next step to move.
(Move_eyes_to_nearest_nominee IF ((Step Move)	If it is time to move the eyes
(Tag ?Dest_Object Current Destination) (Visual ?Dest_Object Text ???)	and the text of the saccade destination has been perceived,
(Tag ?Word Nominee Object) (Not (Visual ?Word Text ???)) (Visual ?Word Eccentricity ?Ecc) (Least ?Ecc))	select the saccade destination nominee with the least eccentricity
THEN ((Send_to_motor Ocular Perform Move ?Word) (Add (Tag ?Word Current Destination)) (Delete (Tag ?Old Current Destination))	and move the eyes to that word, tagging it as the next saccade destination.
(Delete (Step Move)) (Add (Step Nominate))))	

strict-processing model to account for saccade durations; (b) text-encoding errors to help account for total fixations; (c) fixating nearby objects to help account for saccade distances and scanpaths; and (d) inhibiting group revisits to help account for saccades distances and scanpaths. These help to answer the active vision questions of when and

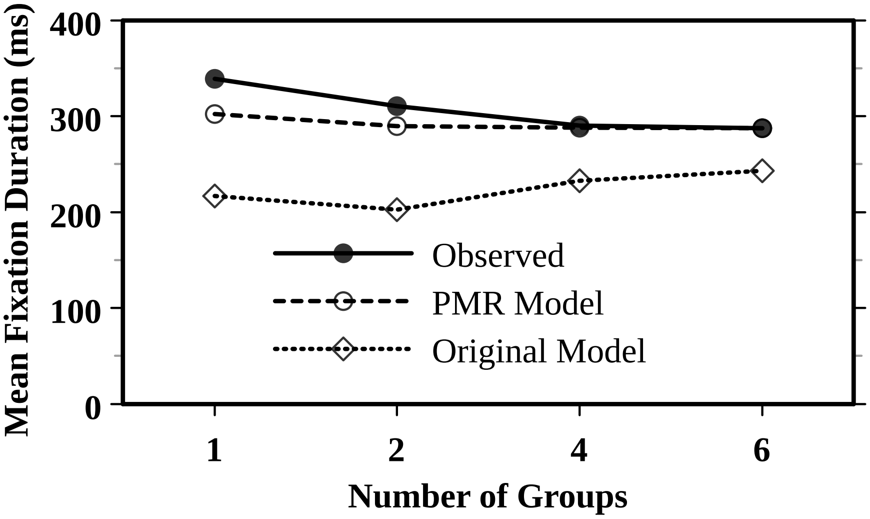


Figure 17. Fixation duration observed in the CVC search task (solid line), predicted by the original model (dashed line with circles), and predicted by the Process-Monitoring Revisited (PMR) model (dashed line with diamonds). The AAE of the original model is 26.5% and the PMR model is 4.6%.

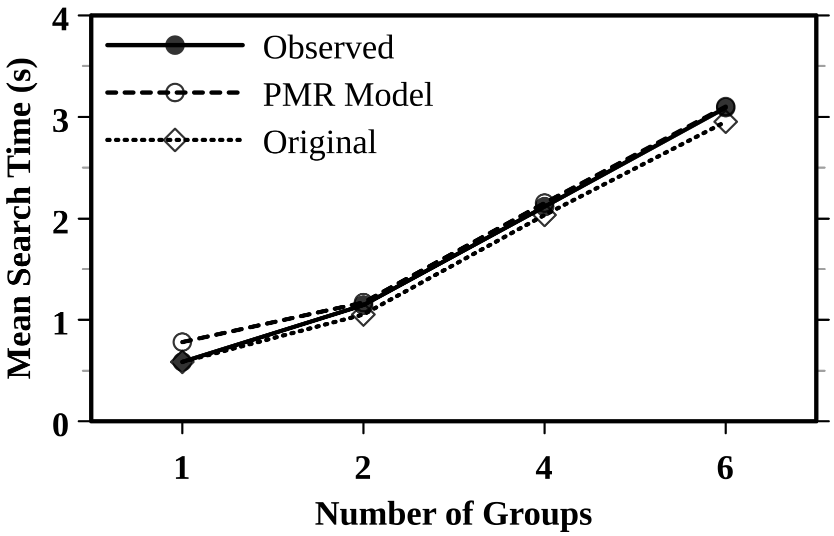


Figure 18. Search time observed in the CVC search task (solid line), predicted by the original model (dashed line with diamonds), and predicted by the Process-Monitoring Revisited (PMR) model (dashed line with circles). The AAE of the original model is 4.2% and the PMR model is 9.7%.

where the eyes move, what the eyes perceive in a fixation, and what information is integrated across fixations. The model details are motivated by eye movement data and previous research, and can be applied to other modeling research. The next section further validates the active vision model.

6. MODEL VALIDATION WITH THE SEMANTIC GROUPING TASK

An aim of this research is to inform the development of predictive, automated interface analysis tools and, as such, a validation is required to test the *a priori* predictive power of the *post hoc* model. The active vision model described in the previous sections is next applied to the non-semantic conditions of the semantic grouping task (Halverson, 2008). This task provides a rich set of reaction time and eye movement data for a task that is arguably more ecologically valid than the other tasks on which the model was built, so this should be a good test of the model. One of many ecologically valid details in the experimental design included that the precue always appeared at the location of the target from the *previous* trial, much like how a web page search typically starts at the location of the link you clicked on to arrive at the new page. Human performance for the semantically cohesive and random layouts was compared to the model's predictions across measures of search time, number of fixations, and fixation duration.

As shown in Figures 19, 20, and 21, the model did a very good job of predicting the search times, number of fixations, and saccade distances for the random-group conditions. In all three measures, when only considering the random conditions, the model predicted the observed data with accuracies well below the intended AAE of 10%.

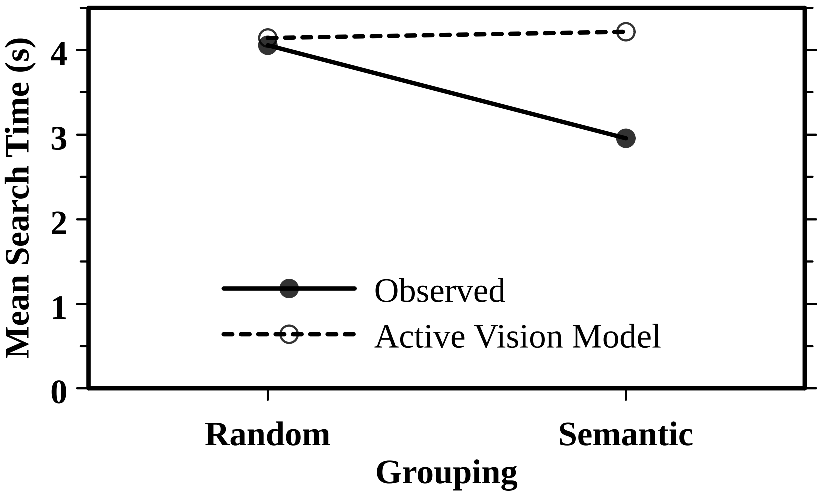


Figure 19. Search time observed in the semantic grouping task (circles), predicted by the Active Vision model (squares). The AAE is 20.7% and for the random layout alone, 6.5%.

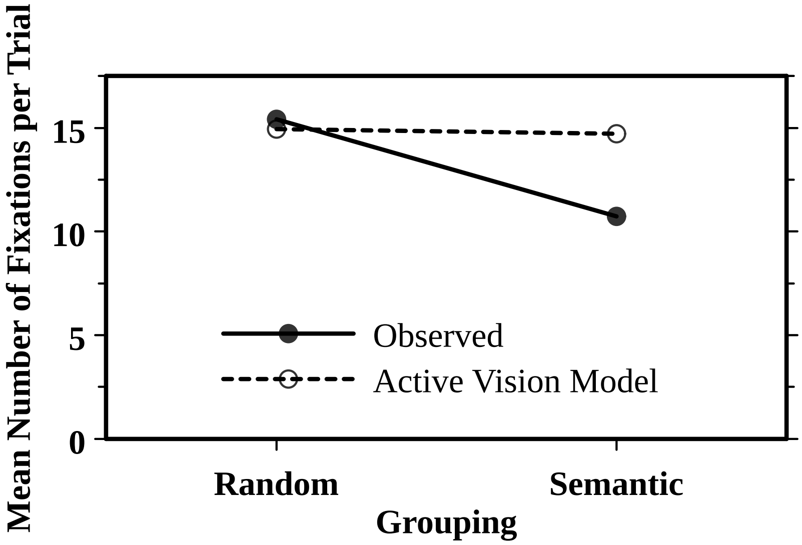


Figure 20. The number of fixations per trial observed in the semantic grouping task (circles), and predicted by the Active Vision model (squares). The overall AAE is 20.2% and for the random layout alone, 3.0%.

With the exception of saccade distance, the model did not accurately predict human performance in the semantic conditions. Since the cognitive model had no representation for semantic information, it is not surprising that the model makes more fixations than people, who could sometimes pass over an entire group of words with a single fixation that captured the semantic content of the group.

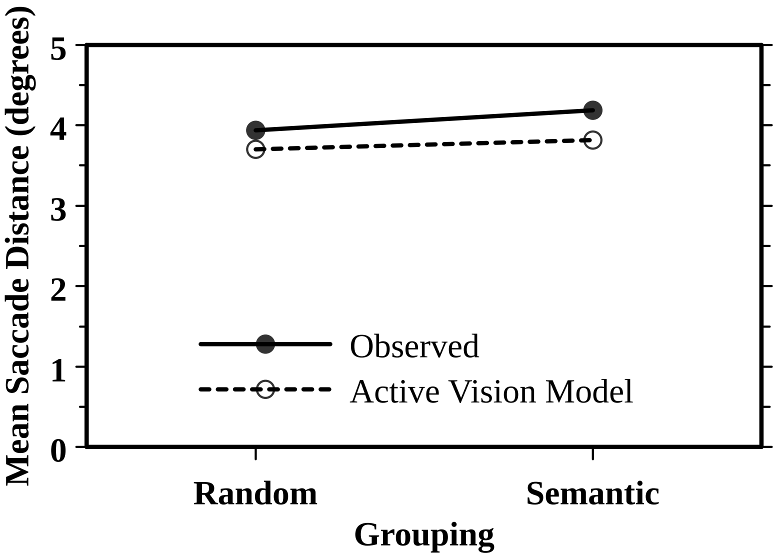


Figure 21. Saccade distance observed in the semantic grouping task (circles), and predicted by the Active Vision model (squares). The AAE is 7.4%.

But the model does a good job predicting human data from the semantic grouping task for layouts without organized semantic relationships. The model predicts the search time and eye movement data within our target AAE of 10%. The ability to predict visual search behavior *a priori* for a task that includes a larger layout, more words, and a different word set provides validation for the active vision model. These results suggest that the model would be an appropriate starting place for modeling more complex tasks and more complex stimuli, such as with the addition of a semantic processor.

The correct and incorrect predictions made by the model in the semantically-grouped conditions provide guidance for future work. That the saccade distances are correctly predicted for both semantic and non-semantic layouts suggests that one important aspect of the model, the basis of saccade destination selection, may be correct. The results suggest that certain constraints of human information processing are invariant across tasks and that the active vision model has captured many of those constraints. The semantic grouping task offers a third set of data that has now been accurately modeled in the EPIC cognitive architecture with a refined, comprehensive model of visual search, the result of which extends our understanding of how people visually search computer displays and provides a basis for predictive modeling.

7. DISCUSSION

This paper presents a comprehensive computational model of active vision for visual search. This model is a substantial push towards a model for predicting visual search in human-computer interaction tasks. Such a model is needed for automated interface analysis tools, like CogTool (John & Salvucci, 2005), which do not yet have a fully developed active vision model that can simulate people's visual search behavior. The model instantiates proposed answers to the important questions of active vision (Findlay & Gilchrist, 2003): What can be perceived during a fixation? When and why are saccades initiated? What do the eyes fixate next? What information is integrated across fixations? That the model is built into a computer program that actually runs, and generates predictions, demonstrates a certain level of completeness. All of the necessary components are account for at some degree of detail. The model is sufficient to account for the major processes involved in active vision.

7.1. Contributions to Cognitive Modeling

This research moves the field of HCI closer to a powerful, detailed, computational understanding of how people apply their active vision processes to visual HCI tasks. This work extends the practice of computational cognitive modeling by (a) informing the process of developing such computational models by using eye movement data in a principled manner and (b) addressing the four questions of active vision for the first time in a computational framework, setting a standard of completeness for future modeling of visual search in HCI. Critical theoretical contributions were identified along the way that will be useful to incorporate into future models of visual search.

The model of visual search proposed here accounts for a variety of eye movement data, from fixation duration to scanpaths, and works well to predict people's ability to locate a target of visual search. The model does so by employing visual search strategies and constraints, informed by eye movement data and previous research, that can be applied to other modeling research. The strategies and constraints in the model suggest answers to the four questions of active vision (Findlay & Gilchrist, 2003), which are: (a) When and why are saccades initiated? Answer: A strict process-monitoring saccade initiation strategy predicts peoples' fixation durations well. The model shows how the simulated flow of information through perceptual processors, like the transduction times instantiated in the EPIC cognitive architecture, can be used to explain the observed fixation durations. While other hypotheses of saccade initiation (Hooge & Erkelens, 1996) are not ruled out by this research, the instantiation of the process-monitoring strategy used in this modeling is able to predict visual search behavior without additional mechanisms or parameters that would be necessary to implement the other saccade initiation strategies. (b) What do the eyes fixate next? Answer: The eyes tend to go to nearby objects. When the target does not "pop out", a strategy of selecting saccade destinations based on proximity to the center of fixation does a good job of predicting people's eye movement behavior. The model predicts people's saccade distributions and scanpaths by utilizing only the location of the objects in the layout, a further contribution to predictive modeling in HCI in that object location is one of the few visual characteristics that can be automatically translated from a physical device to a predictive modeling tool. (c) What can be perceived during a fixation? Answer: Items nearer the point of gaze are more likely to be perceived, with varying eccentricities for different features. However, the visual features (e.g. text) of nearby objects are sometimes misidentified. The modeling supports the use of text-encoding errors, even for objects very near the center of fixation, as text-encoding errors may do a better job of explaining the limitations of what information is processed in a fixation than can be done by varying the effective field of view (Bertera & Rayner, 2000). A text-encoding error rate of 10% predicts human performance well across multiple tasks. (d) What information is integrated across fixations? Answer: The memory for the locations previously visited is required across fixations. While identifying the constraints of working memory on visual search was not an explicit goal of this research, the modeling does suggest something about the use of memory during the visual search of structured layouts. The proposed model uses the memory for previous regions visited to help explain the observed saccade distances and scanpaths. So memory for previously fixated locations may be integrated across fixations to guide search toward unexplored areas (Klein & MacInnes, 1999).

This research informs the process of building computational models of visual search in a principled way. The model is based on a variety of eye movement measures, informed by previous research literature on visual search, and guided by the principles underlying the EPIC cognitive architecture. Using eye movements to inform the building of computational models of visual search is useful. The original CVC model (Hornof, 2004) predicted the search time slightly better than the active vision model of visual search. However, the original model did not do as well at predicting the eye movements. This is not surprising since the original model was not informed by eye movement analysis. However, this discrepancy between predicting visual search time and predicting detailed visual search behavior (i.e. eye movements) shows a strong need for utilizing eye movement data when building models. The comprehensive active vision model is informed by a variety of precise eye movement measurements at every step of the process, which provides more support for the resulting model.

The visual search literature also provides support for the model. Previous claims in the research literature were computationally instantiated and integrated within the proposed model (Bertera & Rayner, 2000; Hooge & Erkelens, 1996; Motter & Belky, 1998). These instantiations provide potential refinements of previous claims, such as with Bertera and Rayner’s (2000) finding that effective fields of view do not change as a function of density. The modeling reinforces and refines Bertera and Rayner’s claim by showing that using text-encoding errors, with the error rates differentiated by text density, explains the data better than varying the region in which text can be perceived as a function of density. The tasks used to inform the active vision model differ from those used by Bertera and Rayner, which used randomly arranged single letters. In our mixed-density task, density is manipulated by varying the size of text and spacing which is arguably more ecologically valid than the stimuli used by Bertera and Rayner. Even with these differences, both this and the previous research conclude that the region that is perceived during a fixation does not vary with density.

The model currently predicts the visual search of text-based displays with an acceptable level of accuracy for engineering based models. Such an active vision model of visual search will be useful for automated interface analysis tools, discussed next.

7.2. Informing the Development of Automated Interface Analysis Tools

An aim of this research is to provide the theoretical underpinnings needed for automated interface analysis tools, and to provide a useful method of predicting users' gaze interaction with novel visual displays. Interface designers can use such tools to evaluate visual layouts early in the design cycle before user testing is feasible. Work is required to integrate the results of this modeling, and future related modeling, into one or more interface analysis tools like CogTool (John & Salvucci, 2005) and CORE/X-PRT (Tollinger et al., 2005).

At least two directions can be taken to improve the predictive power of interface analysis tools with respect to predicting users' visual interaction: (a) Improve the predictive power of models of visual search, as is done with the active model of visual search in this article, and (b) enhance interface analysis tools with a more robust model of visual search based on such a model. Regarding the second goal, some progress has already been made, but more is needed. Aspects of this research have already been demonstrated to be useful for such tools. As evidence of the need for and impact of an active vision model of visual search, CogTool-Explorer (an research extension to

CogTool) was recently updated (Teo & John, 2008) to include aspects of the visual search strategies reported in earlier, partial reports on this modeling (Halverson & Hornof, 2007). The accuracy with which CogTool-Explorer predicts visual search behavior improved when augmented with principles identified in the active model of visual search. For example, CogTool-Explorer searches visual objects in an order based on the eccentricity of the objects relative to one another. However, CogTool-Explorer and the computational model on which it is partially based, SNIF-ACT (Fu & Pirolli, 2007), do not embrace many aspects of active vision. These tools do not simulate eye movements, and incorporate limited simulations of visual perception. For example, all visual objects on a web page have equal visual saliency regardless of location on the page.

7.3. Future Directions

While the progression of models presented in this research is a substantial step towards a unified theory of visual search for HCI, more work is required to achieve a truly unified theory of visual cognition. The proposed model answers questions that are important to the study of active vision. However, it does so for a limited domain, that of structured layouts of text. More work is needed.

Integration of Models of Visual Search

Currently, models of visual search cannot accurately predict the behavior of users' visual interaction with the complex visual layouts of today's computer applications. Individual models exist that separately instantiate different strategies that people use when visually searching. However, a unified visual search theory is needed. Newell proposed a unified theory of cognition, which he described as "…a single system [that] would have to take the instructions for each [task], as well as carry out the task. For it must truly be a single system in order to provide the integration we seek" (Newell, 1973, p. 305). His vision of a unified theory of cognition has to some extent been realized in cognitive architectures such as EPIC (Kieras & Meyer, 1997) and ACT-R (Anderson et al., 1997). However, the independence of the models instantiated in the architectures has a decentralizing effect if there is no unification of the theory embedded in the separate architectures or individual models. Therefore, future work is required to integrate across multiple architectures and models, including models from different cognitive architectures.

Integration with Other EPIC Models

Other computational models of visual search have been proposed in EPIC that propose slightly different answers to some of the questions of active vision. EPIC is conducive to the modeling of active vision as it emphasizes perceptual and motor processes that are central to active vision, such as the visual processor and ocular motor processor. The variation in different models is a good thing for a number of reasons. For one, until the theory is nailed down, the architecture should not unnecessarily restrict the modeling but should instead leave room for competing theoretical explorations. For another, a wide variety of tasks need to be simulated before a truly comprehensive model can be developed.

A current area of research using the EPIC cognitive architecture is the investigation of the perceptual constraints of the visual system (Kieras & Marshall, 2006; Kieras, 2003). Recent modeling efforts have refined EPIC's visual availability functions, which are the equations that determine what visual properties are available to cognitive processes as a function of where the object is in the visual field. For example, the default availability function for text is that text can be perceived out to 1° of visual angle from the center of gaze. Availability functions for a range of visual features are necessary to accurately describe visual search behavior.

Both Kieras’s availability functions and the Fixate-Nearby strategy can be used to explain people’s saccade behavior in different tasks. Further research is required to determine whether both are necessary to predict observed scanpaths, how the two methods may be integrated, and whether one strategy subsumes the other. Such integration will be useful for extending our active vision model to a wider variety of visual tasks.

Integration with Models of Semantic Search

While the model was able to explain some of the eye movement behavior in the Semantic Grouping task, the effects of semantics on saccade destination selection was not explained. Research and modeling (Brumby & Howes, 2004; Brumby & Howes, 2008; Fu & Pirolli, 2007) has provided much insight in to how the semantics can guide visual search. Computational models such as Brumby and Howes’s interdependence of link assessment model (2004) and Fu and Pirolli’s SNIF-ACT 2.0 (2007) have accounted quite well for the influence of text semantics on people’s visual search processes. Conversely, these models use over-simplified scanpaths and do not account for many aspects of active vision. Our active vision model does a good job of predicting how people select saccade destinations. The integration of these models would benefit predictive modeling in HCI tasks, as important factors to consider in screen design are how users move their eyes, and how the semantic content influences their navigation of information.

8. CONCLUSION

To better support users and predict their behavior with future human-computer interfaces, it is essential that we better understand how people search visual layouts. Computational cognitive modeling is an effective means of expanding visual search theory in HCI, and ultimately will provide a means of predicting visual search behavior to aid in the evaluation of user interfaces. The active vision computational cognitive model of visual search presented here illustrates the efficacy of using eye movements in a methodical manner to better understand and predict visual search behavior. Additionally, the results from the modeling solidify and extend an understanding of active vision in a manner that is useful for future HCI research by instantiating the theory in a computational model. This instantiation allows us to better understand the effects and interactions of visual search processes and how these visual search processes can be used computationally to predict people’s visual search behavior. This research ultimately benefits HCI by giving researchers and practitioners a better understanding of how users visually interact with computers, and provides a foundation for tools to predict that interaction.

References

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(4)*, 439-462.

Barbur, J. L., Forsyth, P. M., & Wooding, D. S. (1990). Eye movements and search performance. In D. Brogan, A. Gale, & K. Carr (Eds.), *Visual Search 2*. (pp.253-64). London: Taylor & Francis.

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of effective stimulus in visual search. *Perception & Psychophysics, 62(3)*, 576-585.

Brumby, D. P., & Howes, A. (2004). Good enough but I'll just check: Web-Page search as attentional refocusing. *Proceedings of the International Conference on Cognitive Modeling*, Pittsburgh, PA, 46-51.

Brumby, D. P., & Howes, A. (2008). Strategies for guiding interactive search: An empirical investigation into the consequences of label relevance for assessment and selection. *Human-Computer Interaction, 23(1)*, 1-46.

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

Card, S. K. (1982). User perceptual mechanisms in the search of computer command menus. *Proceedings of the Conference on Human Factors in Computing System*s, Gaithersburg, MD, 190-196.

Findlay, J. M., & Gilchrist, I. D. (1998). Eye guidance and visual search. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception*. (pp.295-312). Amsterdam: Elsevier.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Fleetwood, M. D., & Byrne, M. D. (2006). Modeling the visual search of displays: A revised ACT-R/PM model of icon search based on eye tracking data. *Human-Computer Interaction, 21(2)*, 153-197.

Fu, W., & Pirolli, P. (2007). SNIF-ACT: A cognitive model of user navigation on the world wide web. *Human-Computer Interaction, 22(4)*, 355 - 412.

Halverson, T. (2008). *An "active vision" computational model of visual search for human-computer interaction.* Unpublished doctoral dissertation, University of Oregon, Eugene, OR.

Halverson, T., & Hornof, A. J. (2004a). Explaining eye movements in the visual search of varying density layouts. *Proceedings of the International Conference on Cognitive Modeling*, Pittsburgh, Pennsylvania, 124-129.

Halverson, T., & Hornof, A. J. (2004b). Local density guides visual search: Sparse groups are first and faster. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, New Orleans, LA, 1860-1864.

Halverson, T., & Hornof, A. J. (2004c). Strategy shifts in mixed-density search. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Chicago, IL, 529-534.

Halverson, T., & Hornof, A. J. (2007). A minimal model of for predicting visual search in human-computer interaction. *Proceedings of the Conference on Human Factors in Computing System*s, San Jose, CA, 431-434.

Hooge, I. T. C., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Perception and Psychophysics, 58*, 969-976.

Hornof, A. J. (2001). Visual search and mouse pointing in labeled versus unlabeled two-dimensional visual hierarchies. *ACM Transactions on Computer-Human Interaction, 8(3)*, 171-197.

Hornof, A. J. (2004). Cognitive strategies for the visual search of hierarchical computer displays. *Human-Computer Interaction, 19(3)*, 183-223.

Hornof, A. J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. *Proceedings of the Conference on Human Factors in Computing Systems*, Ft. Lauderdale, FL, 249-256.

Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception & Psychophysics, 63(2)*, 272-285.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Which technique. *ACM Transactions on Computer-Human Interaction, 3(4)*, 287-319.

John, B. E., & Salvucci, D. D. (2005). Multi-Purpose prototypes for assessing user interfaces in pervasive computing systems. *IEEE Pervasive Computing, 4(4)*, 27-34.

John, B. E., Prevas, K., Salvucci, D. D., & Koedinger, K. (2004). Predictive human performance modeling made easy. *Proceedings of the Conference on Human Factors in Computing Systems*, Vienna, Austria, 455-462.

Kieras, D. E. (2003, November). Modeling visual search in the EPIC architecture. Meeting of Office of Naval Research Grantees in the Area of Cognitive Architectures, University of Pittsburgh, PA.

Kieras, D. E., & Marshall, S. P. (2006). Visual availability and fixation memory in modeling visual search using the EPIC architecture. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vancouver, BC, Canada,

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(4),* 391-438.

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

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science, 10(4),* 346-352.

Koostra, G., Nederveen, A., & de Boer, B. (2006). On the bottom-up and top-down influences of eye movements. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2538.

Logan, G. D. (1978). Attention in character classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General, 107,* 32-63.

Logan, G. D. (1979). On the use of concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception & Performance, 5,* 189-207.

Logie, R. H. (1995). *Visuo-Spatial working memory*. Hove, UK: Lawrence Erlbaum.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction, 8,* 353-388.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38(12)*, 1905-1815.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual Information Processing*. (pp.283-308). New York: Academic Press.

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

Oh, S., & Kim, M. (2004). The role of spatial working memory is visual search efficiency. *Psychonomic Bulletin & Review, 11(2),* 275-281.

Ojanpää, H., Näsänen, R., & Kojo, I. (2002). Eye movements in the visual search of word lists. *Vision Research, 42(12),* 1499-1512.

Perlman, G. (1984). Making the right choices with menus. *Proceedings of Interact '84*, London, England, 317-321.

Pomplun, M., Reingold, E. M., & Shen, J. (2003). Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science, 27(2),* 299-312.

Posner, M. I., & Cohen, Y. (1984). Components of attention. In H. Bouma, & D. G. Bouwhuis (Eds.), *Attention and Performance X*. (pp.55-66). Hillsdale, NJ: Erlbaum.

Salvucci, D. D. (2001). Predicting the effects of in-car interface use on driver performance: An integrated model approach. *International Journal of Human-Computer Studies, 55*, 85-107.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocol. *Proceedings of the Eye Tracking Research and Applications Symposium*, Palms Beach Gardens, FL, 71-78.

Shore, D. I., & Klein, R. M. (2000). On the manifestations of memory in visual search. *Spatial Vision, 14(1),* 59-75.

Somberg, B. L. (1987). A comparison of rule-based and positionally constant arrangements of computer menu items. *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface*, Toronto, ON, Canada, 79-84.

Taatgen, N. A., Rijn, H. v., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention and learning. *Psychological Review, 114(3),* 577-598.

Teo, L., & John, B. E. (2008). Towards a tool for predicting goal-directed exploratory behavior. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, New York, 950-954.

Tollinger, I., Lewis, R. L., McCurdy, M., Tollinger, P., Vera, A., Howes, A., et al. (2005). Supporting efficient development of cognitive models at multiple skill levels: Exploring recent advances in constraint-based modeling. *Proceedings of the Conference on Human Factors in Computing Systems,* Portland, OR, 411-420.

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

Wolfe, J. M., & Gancarz, G. (1996). Guided search 3.0: A mode of visual search catches up with jay enoch 40 years later. In V. Lakshminarayanan (Ed.), *Basic and Clinical Applications of Vision Science*. (pp.189-92). Dordrecht, Netherlands: Kluwer Academic.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*, 1-7.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science, 12(3)*, 219-224.