# A Computational Model of “Active Vision” for Visual Search in Human-Computer Interaction

**Tim Halverson**

*Oak Ridge Institute of Science and Education*

*Air Force Research Laboratory*

**Anthony J. Hornof**

*University of Oregon*

**RUNNING HEAD: COMPUTATIONAL MODEL OF ACTIVE VISION**

## *Corresponding Author’s Contact Information:*

Tim Halverson
Air Force Research Laboratory
Mesa Research Site
6030 South Kent St.
Mesa, AZ  85212
Email: thalverson@gmail.com

***Brief Authors’ Biographies:***

**Tim Halverson** is a cognitive scientist with an interest in human-computer interaction, cognitive modeling, eye movements, and fatigue; he is a post-doctoral research associate in the Performance and Learning Model Team of the Air Force Research Laboratory. **Anthony J. Hornof** is a computer scientist 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.

Accepted to the journal *Human-Computer Interaction* on August 20, 2010.

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*

| | | |
|---|---|---|
| 1. REPORT DATE<br>**AUG 2010** | 2. REPORT TYPE | 3. DATES COVERED<br>**00-00-2010 to 00-00-2010** |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>**A Computational Model of ’Active Vision’ for Visual Search in 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)<br>**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

**Human visual search plays an important role in many human-computer interaction (HCI) tasks. Better models of visual search are needed not just to predict overall performance outcomes, such as whether people will be able to find the information needed to complete an HCI task, but to understand the many human processes that interact in visual search, which will in turn inform the detailed design of better user interfaces. This article describes a detailed instantiation, in the form of a computational cognitive model, of a comprehensive theory of human visual processing known as ?active vision? (Findlay & Gilchrist, 2003). The computational model is built using the EPIC (Executive Process-Interactive Control) cognitive architecture. Eye tracking data from three experiments inform the development and validation of the model. The modeling asks?and at least partially answers?the four questions of active vision: (1) What can be perceived in a fixation? (2) When do the eyes move? (3) Where do the eyes move? (4) What information is integrated between eye movements? Answers include: (1) Items nearer the point of gaze are more likely to be perceived, and the visual features of objects are sometimes misidentified. (2) The eyes move after the fixated visual stimulus has been processed (i.e., has entered working memory). (3) The eyes tend to go to nearby objects. (4) Only the coarse spatial information of what has been fixated is likely maintained between fixations. The model developed to answer these questions has both scientific and practical value in that the model gives HCI researchers and practitioners a better understanding of how people visually interact with computers, and provides a theoretical foundation for predictive analysis tools that can predict aspects of that interaction.**

15. SUBJECT TERMS

| | | | | | |
|---|---|---|---|---|---|
| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT<br>**Same as Report (SAR)** | 18. NUMBER OF PAGES<br>**40** | 19a. NAME OF RESPONSIBLE PERSON |
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | | | |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std Z39-18

## ABSTRACT

Human visual search plays an important role in many human-computer interaction (HCI) tasks. Better models of visual search are needed not just to predict overall performance outcomes, such as whether people will be able to find the information needed to complete an HCI task, but to understand the many human processes that interact in visual search, which will in turn inform the detailed design of better user interfaces. This article describes a detailed instantiation, in the form of a computational cognitive model, of a comprehensive theory of human visual processing known as "active vision" (Findlay & Gilchrist, 2003). The computational model is built using the EPIC (Executive Process-Interactive Control) cognitive architecture. Eye tracking data from three experiments inform the development and validation of the model. The modeling asks—and at least partially answers—the four questions of active vision: (1) What can be perceived in a fixation? (2) When do the eyes move?  (3) Where do the eyes move? (4) What information is integrated between eye movements? Answers include: (1) Items nearer the point of gaze are more likely to be perceived, and the visual features of objects are sometimes misidentified. (2) The eyes move after the fixated visual stimulus has been processed (i.e., has entered working memory). (3) The eyes tend to go to nearby objects. (4) Only the coarse spatial information of what has been fixated is likely maintained between fixations.  The model developed to answer these questions has both scientific and practical value in that the model gives HCI researchers and practitioners a better understanding of how people visually interact with computers, and provides a theoretical foundation for predictive analysis tools that can predict aspects of that interaction.

## CONTENTS

# 1. INTRODUCTION

Visual search is an important part of human-computer interaction (HCI). Users search news web sites to locate stories of interest, search user interfaces to learn how to use desktop applications, and search virtual environments 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 that a user seeks. 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 proposing and evaluating plausible sets of interaction among these processes.

The most important contribution of computational cognitive models to the field of HCI is that the modeling provides a science base that is badly needed for predictive interface-analysis tools. Projects such as CogTool (John, Prevas, Salvucci & Koedinger, 2004) and CORE/X-PRT (Tollinger et al., 2005) are at the forefront of the development 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 for a range of tasks without requiring that the analyst be knowledgeable in the cognitive, perceptual, and motoric theories embedded in the tool. Designers of device and application interfaces may be able to utilize such tools to evaluate their visual layouts, identify potential usability problems early in the design cycle, and reduce the need for more human user testing early in the development cycle.

Predicting people's visual interaction is one aspect of user behavior that research with interface analysis tools is trying to improve. To this end, a recent version of CogTool (Teo & John, 2008, 2010) now incorporates modeling work presented in this paper based on an early summary of the work (Halverson & Hornof, 2007). However, interface analysis tools such as CogTool and CORE/X-PRT do not yet fully account for human vision, as in where the eyes move and what they do and do not see. A partial account of visual information processing is provided by EMMA (Salvucci, 2001), which is an extension to the ACT-R (Anderson, Matessa & Lebiere, 1997) modeling framework underlying CogTool. EMMA provides a simulation of the eyes including where the eyes move and how quickly visual information is processed. But this falls short of a complete account of active vision; automated interface analysis tools do not yet simulate active vision.

*Active vision* (Findlay & Gilchrist, 2003) embraces 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 continually move their eyes to sample information. Accounting for these eye movements not only allows a better understanding of the processes underlying visual search, but also a better understanding of how people use computer interfaces.

This paper describes a computational model of visual search for HCI that integrates a contemporary understanding of visual processing in the context of active vision. The remainder of this paper is arranged as follows: Section 2 introduces the EPIC (Kieras & Meyer, 1997) cognitive architecture that was used to build the model, and describes two eye tracking experiments that helped to guide the development of the model. Section 3 asks and answers the four questions of active vision in the context of the model. Section 4 assesses the validity of the model with a new set of data. Section 5 summarizes the research, identifies key contributions, and suggests future directions.

## 2. MODELING ACTIVE-VISION VISUAL SEARCH

The goal of this work is to better understand and predict how people visually search computer displays in everyday tasks. It is increasingly important for models of visual search in HCI to account for eye movements (and sometimes even head and body movements). 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. One way to improve models of visual search is to address the questions raised by active vision (Findlay & Gilchrist, 2003).

A model of active vision should address the four questions posed by active vision, the answers to which are important to designers and those interested in HCI: (1) When and why do we move our eyes? (2) Where do we move our eyes? (3) What information in the environment can be perceived when the eyes are held steady? (4) What information from the environment is integrated between eye movements? The research presented here proposes answers to these four questions within a larger set of psychological theory using a cognitive architecture, specifically EPIC (Kieras & Meyer, 1997).

A cognitive architecture provides a computational instantiation of psychological theory that is useful for modeling human performance. The architecture constrains the construction of the models by enforcing human capabilities and constraints. The cognitive models discussed in this paper consist of (a) a detailed set of if-then statements called *production rules* that encode 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 constrain 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.

The theory, which is computationally instantiated in the models, generates predictions of how a person would perform the task. The results of such simulations allow the testing of the theory by directly comparing the model’s performance with human performance.

There is a special synergetic relationship between cognitive modeling and the study of eye movements. Eye movements provide data for informing the construction and evaluation of the models at a more detailed level than reaction time data. Eye movement 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 motivate the observed eye movements. The modeling framework used here—EPIC—is particularly well-suited to model active vision because EPIC makes explicit predictions of eye movements.

## 2.1. The EPIC Cognitive Architecture

EPIC (Executive Process-Interactive Control) is a cognitive architecture that computationally instantiates and integrates theories of perceptual, motor, and cognitive processing 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, speaker, 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. Information from the perceptual processors is 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, represented 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 with each other.

FIGURE 1 ABOUT HERE

The perceptual and motor processors constrain the behavior that a set of production rules can generate. Particularly relevant to active vision are the constraints imposed by the simulated eyes, including EPIC’s *retinal availability functions*, which 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. EPIC also simulates the ballistic eye movements, called saccades, which are made to gather visual 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. To illustrate, Figure 2 annotates the contents of a production rule that selects the next saccade destination and prepares an eye movement to that location.

FIGURE 2 ABOUT HERE

The encoding of visual objects and their properties into visual working memory takes time. EPIC simulates these encoding times by delaying information as it flows through the visual sensory processor and the visual-perceptual processor, each of which induces a delay. For example, if an object appears in the model’s parafovea, the shape of that object would appear 50 ms later in the visual sensory store and another 50 ms later in the perceptual memory store (i.e., visual working memory). Different visual features have different delays, which are detailed in Kieras (2004) and Kieras and Meyer (1997).

**Model Development**

All models presented in this article started with the core EPIC cognitive architecture. Some modifications were made to EPIC's visual processors during the iterative process of refining the models. All of EPIC's perceptual properties were kept 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 extent, and 150 ms if the previous saccade had a different direction and extent; saccades require 4 ms per degree of visual angle to rotate the eyeball.

The models search "without replacement." That is, any object for which the text has been perceived is excluded from being the destination of future saccades. While there is some controversy over whether visual search proceeds with replacement (i.e., amnesic-search; see for example Horowitz & Wolfe, 2001) or without replacement (see for example Shore & Klein, 2000), the preponderance of evidence favors search without replacement.

The development of the model proceeded in a principled manner, guided by the questions raised by active vision (Findlay & Gilchrist, 2003). The modeling started with a baseline model that was based on the reasonable initial assumptions identified above, and progressed to a model that explains many features of observed eye movement data. With a model of active vision as the goal, the modeling focused on details in the data that related to questions such as what is perceived in a fixation and when saccades are initiated.

Throughout the development of the models presented in this paper, a model's prediction is considered to be accurate, or at least adequate, if its prediction falls within 10% of the observed data. This is consistent with engineering practices (Kieras, Wood & Meyer, 1997).

## 2.2. Eye Tracking Experiments to Develop the Model

The computational model of active vision was developed using eye tracking data from two experiments: a mixed density search task and a CVC (consonant-vowel-consonant) search task. The mixed density experiment (Halverson & Hornof, 2004b) investigated the effects of varying the visual density of elements in a structured layout. The CVC search experiment (Hornof, 2004) investigated the effects of layout size and visual hierarchy. Together, the two experiments provide a useful set of data for building and refining an active-vision model of visual search for HCI because there are substantial differences between the two tasks, and because the model needs to predict performance for a range of visual layouts and features.

Both experiments were conducted using a classic visual search experimental paradigm in which the entire layout is displayed at the same moment, permitting any search order, and the trials are blocked by experimental condition, which in this case is layout type. 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.

**Mixed Density Task**

The mixed density experiment explored how the size and spacing of text affects the visual search of structured layouts. The experiment is discussed in more detail in Halverson and Hornof (2004b) and is presented here specifically with regard to developing a comprehensive model of active vision.

Layouts in the mixed density task contained 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 3 shows an example of a mixed density layout. Twenty-four people participated in the experiment.

FIGURE 3 ABOUT HERE

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 when searching sparse layouts. It appears that, with sparse groups, participants adopted a more efficient eye movement strategy that used slightly fewer and slightly shorter fixations.

**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), which controlled for word familiarity and other effects. The CVC task included layouts with and without a labeled visual hierarchy. When labels were used, groups were randomly labeled with single numerical digits flanked by Xs (e.g., “X1X”). Data from the tasks without a labeled visual hierarchy are used to inform the development of the model presented here.

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 that were used 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 4 shows a sample layout from the experiment. One-group layouts used group A. Two-group layouts used groups A and B. Four-group layouts used groups A through D.

FIGURE 4 ABOUT HERE

The results of the experiment show that people were able to search smaller layouts faster than larger, and to search 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 that participants used a more consistent search strategy when a useful visual hierarchy was present.

A good model of active vision needs to accurately predict eye movements, as these are the most directly observable and measurable events of interest during a visual search task. In all experiments reported in this article, eye movements were recorded using a pupil-center and corneal-reflection eye tracker. In all analyses presented in the following sections, the accuracy of the eye tracking data is assured using the required fixation locations post-hoc method (Hornof & Halverson, 2002), and 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. These are the data that our model will explain in order to answer the four questions of active vision, presented next.

## 3. ANSWERING THE FOUR QUESTIONS OF ACTIVE VISION

Building on the special synergetic relationship between cognitive modeling and eye tracking, this section describes the development of models of the mixed density and CVC search tasks using the EPIC cognitive architecture. The result is a single comprehensive model that answers the four questions of active vision: (1) When do the Eyes Move? (2) What Can Be Perceived? (3) Where Do the Eyes Move? (4) What Information is Integrated Between Eye Movements?

### 3.1. When do the Eyes Move?

The question of when to move the eyes from one visual element to another, or conversely how long the eyes should linger on elements in a visual layout, is an important factor to consider in a model of active vision. For example, the eyes might remain on complex icons longer than simple icons in order to gather more visual details.

Four explanations of the control of fixation duration have been proposed in the literature: (a) preprogramming-per-trial, (b) preprogramming-per-fixation, (c) process-monitoring, and (d) mixed-control. The first explanation, preprogramming-per-trial, asserts that the fixation duration required for the task is estimated before the visual search task is initiated, and that this estimated fixation duration is used throughout the entire visual search task. The second explanation, preprogramming-per-fixation, assumes that fixation durations are similarly preset, but dynamically estimated throughout a task and, if previous fixations were too short to perceive the stimulus before initiating a saccade, then future fixation durations are lengthened. The third explanation, process-monitoring, asserts that fixation durations are not estimated, but instead directly determined based on the time that is required to perceive a stimuli during a fixation. The fourth explanation, mixed-control, assumes that saccades are sometimes initiated by the time to perceive the stimuli and sometimes by previously estimated durations. Hooge and Erkelens (1996) review these four explanations of fixation duration control.

A variety of research supports the mixed-control explanation of fixation duration both for reading (e.g., Yang, 2009) and natural scene viewing (e.g., Henderson & Pierce, 2008). However, other data (such as Henderson and Pierce, 2008) suggest that, at least in natural scene viewing, the majority of fixations are process-monitoring.

Existing computational models of visual search implement different mechanisms for the control of fixation duration, each of which is aligned with one of the four explanations of fixation duration. Fixations in Guided Search (Wolfe & Gancarz, 1996) are best characterized as preprogramming-per-trial, as fixation durations are fairly constant, each lasting 200 to 250 ms. Understanding Cognitive Information Engineering (UCIE; Lohse, 1993), on the other hand, proposes a varying time for fixation durations—akin to process-monitoring—with durations based on the number, proximity, and similarity of objects near the point of gaze.

## Modeling Fixation Duration

The model presented here was developed considering the four explanations of fixation duration described by Hooge and Erkelens (1996): preprogramming-per-trial, preprogramming-per-fixation, process-monitoring, and mixed-control. The observed fixation durations for dense groups were longer than the durations for sparse groups, suggesting that a factor such as the density, size, or discriminability of the text influenced the fixation durations. But such a factor could have been used by either a preprogramming-per-fixation or process-monitoring scheme. One way to answer the question of which scheme the active-vision model should use is to look for a parsimonious explanation based on the constraints of the architecture.

Newell (1990) encouraged researchers building cognitive models to “listen to the architecture.” By this, he meant that researchers should develop models that, at least in part, derive their parsimony from the basic principles encoded in the architecture. It turns out that EPIC lends itself to a process-monitoring explanation of saccade initiation because the timing and retinal availability of visual features that are built into the cognitive architecture can be used in a very straightforward manner to simulate process-monitoring. Modeling the preprogramming of saccade initiation would require additional mechanisms and parameters to be added to EPIC, and would therefore decrease the parsimony of the model. (For example, a preprogramming strategy might require a theory of time perception to predict saccade time intervals.) The current modeling effort uses process-monitoring to explain fixation durations and in doing so finds a good fit between the theory and the architecture.

EPIC’s visual-perceptual processor was used to simulate process-monitoring as follows: EPIC’s default recoding time for text (a constant 100 ms) was modified to fit the human data from the mixed density experiment in which fixation durations differed as a function of text density. As shown in Figure 5, 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 was 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 time is 150 ms. Otherwise the text recoding time is 50 ms. The differentiation of the time to recode the text is consistent with a principle in the EPIC architecture in which the processing of visual objects is differentiated based on the features of those visual objects.

FIGURE 5 ABOUT HERE

Figure 6 shows a flowchart that represents the model's production rules used for the process-monitoring. After the text property for the current saccade destination becomes available, and after it is decided whether the target has been found, then the model initiates a saccade. These rules, along with the delays that represent the time to process the visual features, encode the process-monitoring theory of saccade initiation into the active-vision model of visual search.

FIGURE 6 ABOUT HERE

All models presented in this article use process-monitoring, but the mechanism that is used to select the next saccade destination matured throughout the model-building process. The mixed-density task model used EPIC's default ocular motor preparation parameters and only a single production rule. In the final model, the ocular motor preparation time was removed and the model used a sequence of several production rules. The removal of the ocular motor preparation time is consistent with research that suggests that the processes for preparing an eye movement are better represented as decisions implemented in production rules rather than motor preparation time (Kieras, 2003).

The model's predictions are compared to the human data from the mixed density task, in which fixation durations varied systematically as a function of visual layout features. As shown in Figure 5, the process-monitoring (PM) model correctly predicts the fixation durations in the mixed density task. The PM model delays the initiation of saccades until after the text information has entered working memory, and increases recoding time for dense objects. Figure 5 also shows EPIC's prediction with the minimum fixation duration (MFD) model in which saccades are initiated as quickly as possible, somewhat akin to a preprogrammed duration model. As can be seen, the predictions of the process-monitoring model are much better than a similar model in which saccades were initiated without regard to whether the text property has entered working memory. Additional details on this model and its development can be found in Halverson and Hornof (2004a).

The process-monitoring model suggests a number of components that should be included in a comprehensive computational model of active vision for HCI. A process-monitoring strategy for saccade initiation provides straightforward, plausible predictions. The fixation durations predicted by EPIC match the observed mean fixation duration very well, with an average absolute error (AAE) of 10%, by including the time to decide where to move the eyes, to wait on the relevant features to enter working memory, and to execute the eye movement.

While other explanations of the control of fixation duration might also work to explain the observed fixation duration data, this would require introducing additional processes and parameters into the EPIC cognitive architecture. The process-monitoring strategy is an important component of the model as it is parsimonious, predicts the observed data very well, is supported by the literature, and provides a satisfactory answer to the question of when do the eyes move.

## 3.2. What Can Be Perceived?

Another important question that must be answered by a comprehensive model of visual search for human-computer interaction is what can be visually perceived in an interface at any given moment. For example, a user may or may not notice a notification that just appeared on their screen.

One reasonable assumption about what can be perceived is that all objects within a fixed region can be perceived around the point of gaze. Some previous models of visual search make this assumption. Barbur, Forsyth, and Wooding (1990) assume that all information within 1.2° of visual angle of the fixation center can be perceived. UCIE (Lohse, 1993) assumes that all items around the point of gaze 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, Reingold, & Shen, 2003) assumes that all items are perceived within a "fixation field" normally distributed around the center of fixation, and varies based on the properties of the stimuli.

A challenge when building a comprehensive, predictive model of visual search is to determine which of the many research findings regarding what is perceived in a fixation need to be incorporated into the model, and which can be left out. For example, while some models assume that all items within a given region can be perceived in parallel, no models differentiate between the perception of vertically or horizontally organized objects, though research has shown that the region may be larger in the horizontal dimension than in the vertical (Ojanpää, Näsänen & Kojo, 2002). As another example, Casco and Compana (1999) found that density (but not spatial perturbation) affects search time for simple objects, and spatial perturbation (but not density) affects search time for complex objects. Must a useful model account for this? While a predictive model can be useful without addressing all observed phenomena, it is presently unclear how accurately a predictive model needs to represent what can be perceived in a fixation in order for the model to have theoretical and practical value.

A straightforward and reasonable model of visual search for HCI might assume that all objects in an effective field of view (Bertera & Rayner, 2000) are perceived during each fixation. Many existing models do just that (Barbur et al., 1990; Hornof, 2004; Lohse, 1993). This simplifies the model because it means that *location* is the only object feature that is required to determine which objects are perceived, and an object's position is a feature that can be automatically extracted relatively easily from a physical device to a predictive modeling tool (more easily than color, size, shape, etc.). A reasonable approximation for this region is a radius of 1° of visual angle. Such a region has been used, for example, to successfully explain visual search performance for simple shapes (Barbur et al., 1990) and text (Hornof, 2004).

## Modeling What is Perceived During a Fixation

As shown in Figure 7, people require more fixations for denser text. This trend can be modeled in multiple ways. One way is to reduce the size of the region in which dense text can be perceived. Another is to keep the size of the region constant, but reduce the probability that dense text in this region will be correctly perceived. These two explanations were implemented in two separate models, and the predictions of those two models were compared to the observed data.

FIGURE 7 ABOUT HERE

A reduced-region (RR) model was implemented to test the hypothesis that dense text can be perceived over a smaller region than sparse text can be perceived. Previous research suggests that a straightforward way to predict the observed number of fixations in a search task is to assume that two or three objects are processed per fixation (Hornof & Halverson, 2003). For this model, the EPIC visual-perceptual processor availability function was modified so that two or three words were processed per fixation regardless of density. This was accomplished by processing sparse words that appear within 1° of visual angle of the fixation (consistent with EPIC's default availability function for text) but by processing dense words that appear only within 0.5° of visual angle. This modification resulted in a much better fit for the predicted number of fixations per trial. However, as shown in Figure 7, the RR model still under-predicted the number of fixations per trial in all layouts, with an AAE of 21.1%, and so this words-per-fixation approach was rejected in favor of the probability-of-encoding approach, discussed next.

A text-encoding error (TEE) model was implemented to test the hypothesis that the region in which text can be perceived is a constant 1° of visual angle, and that the *probability* of correctly perceiving text is higher for sparse text than for dense text. To this end, EPIC's perceptual processor was modified so that the probability of correctly encoding text increased with the distance to the nearest neighboring object. This is one of several ways of measuring density, and is somewhat akin to how one might model the flanker effects reported in Bouma (1970).

EPIC's visual-perceptual processor was modified as follows: If an object's closest neighbor is at least 0.15° of visual angle away (sparse text), the probability of the model correctly perceiving the text is 90%. Otherwise, the probability of the model correctly perceiving the text is 50%. These probabilities were chosen because they result in an average of two to three items perceived per fixation across both densities, which previous modeling suggests to be a good estimate of the number of items processed per fixation (Hornof & Halverson, 2003).

The TEE model's strategy includes the following details: Even if the text of an object is incorrectly encoded, that object is nonetheless marked as fixated just like objects that are correctly encoded. This makes it possible for the simulation to sometimes fixate but pass over the target as was observed in the human data. If the entire layout is searched (all objects have been marked as fixated) without finding the target, the model restarts the search by resetting all objects as unfixated.

As seen in Figure 7, with text-encoding errors introduced to the model, the TEE model predicts the number of fixations much better than the RR model. The average absolute errors for the two models are 8.8% and 21.1%, respectively. The number of fixations per trial in the TEE model closely approximates the observed data. 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, permitting the model to “listen to the architecture.” Additional details on the model presented here and its development can be found in Halverson and Hornof (2004a).

The modeling suggests that a useful answer to the question of what can be perceived in a fixation is that all objects in a fixed region are marked as perceived, but that the probability of *correctly* perceiving each object property will vary based on the properties (e.g., density) and, further, that the use of encoding errors is a good method to simulate the challenges associated with perceiving dense objects. For the current task, when all items in a fixed region are perceived in every fixation, the model underpredicts the number of eye movements that the humans need to find the target. When the model is modified to include the possibility of misperceiving text, the model correctly predicts the number of fixations used in each layout.

## 3.3. Where Do the Eyes Move?

The order in which items in a layout are searched—the *scanpath*—may have a large impact on usability. For example, visitors to a web page will sometimes follow the scanpath that the designer intended, and sometimes take a completely different path. A great deal of research has been conducted to determine the factors that influence a user’s scanpath in a visual search task (see Wolfe & Horowitz, 2004, for a review).

Scanpaths are influenced by bottom-up features and top-down strategies. Object features (e.g., color, size, shape, or text) affect the order in which objects are searched. When target features are known, and these features can be perceived in the periphery, this information can guide visual search. Many existing models of visual search use visual features 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 the locations of visual information in the visual environment. Visual search is then guided to the items in order from highest to lowest activation. Guided Search 3 (Wolfe & Gancarz, 1996) adds the additional constraint that objects closer to the center of the fixation produce more activation. The Area Activation model (Pomplun et al., 2003) is similar to Guided Search 2 except that the Area Activation model guides the search process to *regions* rather than items of greatest activation. But if peripherally-visible features or the exact identity of the target is not available, bottom-up features alone cannot guide the scanpath, and top-down strategies are needed.

In addition to bottom-up features, top-down strategic decisions also influence the order in which objects are searched. For example, hierarchical versus non-hierarchical layouts motivate fundamentally different strategies (Hornof, 2004), and the ordering of menu items, either alphabetically, functionally, or randomly, also motivate different strategies (Card, 1982; Perlman, 1984; Somberg, 1987). Search patterns are also motivated by the spatial structure or global contour of the objects to be searched (Findlay & Brown, 2006).

Though many complex factors contribute to the decision of which object to fixate next, in general people tend to move their eyes to objects that are relatively nearby. That is, when the target is not visually salient, saccade destinations tend to be based largely on the proximity of objects to the center of fixation (Motter & Belky, 1998).

Previous cognitive modeling 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. The best-fitting model for the CVC task data in Hornof (2004) uses a strategy that, during each saccade, moves the eyes a few items down each column of CVCs. Although the strategy did a good job explaining the human data for that one task, a more-general proximity-based strategy for selecting saccade destinations is needed for more-general search tasks. Fleetwood and Byrne's model of icon search (2006), for example, moved visual attention to the nearest icon that matched one randomly chosen feature of the target icon. With the goal of determining a general-purpose approach for determining scanpaths, the development of an active-vision model next explores the role of proximity in visual search.

### Modeling the Selection of Saccade Destinations

The model implements a strategy in which saccade destinations are chosen based on each visual object's eccentricity—the distance from the current gaze position—as follows: (a) After each saccade, the eccentricity property of each object is updated based on the new eye position. (b) The eccentricity is scaled by a fluctuation factor, which has a mean of 1.0 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 with text that has not been identified and that are in unvisited groups are marked as potential saccade destinations (search without replacement). (d) The candidate object outside the effective field of view—1° of visual angle from the center of fixation—that has the *lowest* eccentricity is selected as the next saccade destination (this was implemented by introducing a "Least" predicate to EPIC's production system). Figure 2 gives a sense of how this strategy can be implemented in production rules.

In visual search, sometimes the eyes return to locations that have already been searched. To this end, the strategy accommodated an occasional revisit to a group. Participants revisited groups only on occasion, approximately once every one to four trials, usually (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 observed revisits is that people remember the regions that they have explored. Searching groups without replacement, described earlier, provides a straightforward approach to explain this behavior.

Figures 8 and 9 show the predictions made by the fixate-nearby (FN) model as well as those made by the original CVC task model (Hornof, 2004) compared to the observed data. As shown in Figure 8, the fixate-nearby model predicts the mean saccade distances very well, with an AAE of 4.2%, a considerable improvement over the AAE of 43.3% of the original model. As shown in Figure 9, the fixate-nearby model also does a good job of predicting the observed scanpaths. The figure shows the three most frequently observed scanpaths, and how the general purpose fixate-nearby model predicts the observed scanpath frequencies better than the original, rigid location-based model. Additional details on the model presented here can be found in Halverson and Hornof (2007).

FIGURE 8 ABOUT HERE

FIGURE 9 ABOUT HERE

Results from this modeling suggest that people select saccade destinations partly based on eccentricity from the current fixation location. The selection of saccade destinations based on proximity results in a good fit of both the mean saccade distance and the scanpaths that people used in this task. The original CVC model (Hornof, 2004) moves the eyes down a few words on each saccade and predicts saccade distances that are much larger than those observed. Additionally, as can be seen in Figure 8, the original model predicts little difference based on the size of the layout, whereas the fixate-nearby accounts for longer saccades in larger layouts. Further, as seen in Figure 9, the fixate-nearby model correctly accounts for the two most frequent scanpaths.

To explain the eye movement data and to depict the human information processing that is not directly observable, two mechanisms described above were introduced to the fixate-nearby model: (a) noisy saccades to nearby objects and (b) inhibition of group revisits. It might seem that these two mechanisms would interact to produce effects similar to those produced by the encoding errors introduced earlier and may thus help to explain not only the current question of “where do the eyes move” but also the previous question of “what can be perceived by the eyes,” but without the need for encoding errors. In other words, 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 as it was with the recoding errors. However, an exploration of this possibly redundant account of what is perceived in a fixation revealed that removing the text-recoding errors from the current model substantially decreased the accuracy of the predicted number of fixations per trial. The Fixate-Nearby model *without* the encoding errors resulted in an AAE of 14.3% for the fixations per trial, which is not acceptable. Therefore, text-encoding errors were left in the model.

The fixate-nearby strategy used in the model has a couple of benefits for predicting visual search behavior if compared to models whose predictions are based primarily on particular visual structures or saliency of visual features, such as Guided Search (Wolfe, 1994). First, a predictive tool using the fixate-nearby strategy would only need to encode the location information from a device representation. This is beneficial if other properties in the layout are either unknown or difficult to automatically extract from the device representation. Second, the fixate-nearby strategy can be used when bottom-up information alone cannot predict visual search, as can be the case with goal-directed search (Koostra, Nederveen & de Boer, 2006). Third, unlike the original CVC model (Hornof, 2004), the fixate-nearby model does not require a predefined scanpath, making the fixate-nearby model more applicable to a variety of layouts.

Where the eyes move and how to model those processes is a difficult question to answer completely, but a straightforward and useful way to model where the eyes move is based on object proximity. The fixate-nearby strategy works well for models of tasks like those used in the present research in which the salience of objects cannot be easily determined or salience does not vary substantially throughout the interface.

## 3.4. What Information is Integrated Between Eye Movements?

The fourth question of active vision for visual search is what information is integrated between eye movements. For example, when searching for a specific news article, a user may or may not remember which headings have already been searched so that those headings can be passed over for the remainder of the search. A comprehensive model of visual search needs to address how working memory affects visual search.

There are multiple types of working memory that could affect visual search. Research has shown that visual search processes use spatial working memory (Oh & Kim, 2004) but that it does not use verbal working memory (Logan, 1978; Logan, 1979), semantic working memory (Altarriba, Kambe, Pollatsek, & Rayner, 2001), or visual working memory (Woodman, Vogel & Luck, 2001). The spatial working memory used during visual search has been shown to be somewhat coarse (Irwin, 1996). A possible use of a coarse spatial working memory is to help select saccade destinations that are away from previous fixation locations (Klein & MacInnes, 1999).

To avoid re-inspecting objects until after all items have been searched without locating the target (i.e., searching without replacement), most computational models of visual search include some mechanism to remember which items have been inspected. In general, these models do not impose any capacity limitations on such memory but instead assume a perfect memory for objects searched (Anderson et al., 1997; Barbur et al., 1990; Byrne, 2001; Hornof, 2004; Kieras & Meyer, 1997; Pomplun et al., 2003; Wolfe, 1994). The next section identifies the role of working memory in the model of active vision presented in this work.

### Modeling Memory for Examined Objects

The fixate-nearby model proposed and evaluated in the previous section offers an answer to what information is maintained between fixations. The answer was coarse spatial information of fixation locations. As was mentioned, in order to explain the observed eye movement data, the fixate-nearby model was set to maintain the locations of previous fixations. When deciding which group to visit next, the model uses memory of which groups—not individual objects—have been visited. This is consistent with previous research that shows that a coarse spatial memory may be the only short-term memory used in visual search (Irwin, 1996; Oh & Kim, 2004). While the model presented in this research does not rule out the possibility that other information is maintained between fixations, the model explains the observed data quite well by maintaining only a high level overview of what has and has not yet been fixated.

### 3.5. Discussion

The cumulative model presented thus far is offered as a candidate computational model of active vision for visual search in human-computer interaction. Answers to the four questions of active vision have emerged from the process of developing the model, and the model does a good job of explaining the observed eye movement data from two experiments. The active-vision model predicts the 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 process-monitoring strategy 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 further help account for saccades distances and scanpaths. The model synthesizes previous research, explains data, and can now be applied to predict performance for new visual search tasks. The next section validates the active-vision model.

## 4. MODEL VALIDATION WITH THE SEMANTIC GROUPING TASK

This section evaluates the predictive potential of the active-vision model described in the previous section by applying the model to the semantic grouping task (Halverson, 2008).

### 4.1. 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.

Figure 10 illustrates the three variables that were manipulated in the layouts: the semantic cohesion of groups of words, the presence or absence of group labels, and the use or non-use of meta-groups (groups of 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. Meta-groups were indicated by colored regions (gray in Figure 10) that provided a second-level semantic grouping to semantic layouts (such as by associating jewelry with cloth) and provided visual structure to random layouts. Figure 10 shows a layout with semantically-cohesive groups, group labels, and meta-groups. All layouts contained eight groups with five words per group. Eighteen people participated in the study.

FIGURE 10 ABOUT HERE

The results of the experiment show that people's search is guided in part by combining the visual structure with the semantic content of the words. People appear to require only a single fixation to judge the semantic relevance of all objects in a semantically cohesive group, even when the group has no summarizing label. The semantic cohesion of words in a group, it turns out, can to some extent substitute for labels of those groups. The meta-groups did not appear to affect people's behavior.

This experiment provides a rich set of reaction time and eye movement data in a task that is arguably more ecologically valid than the previous tasks on which the model was built, so this should be a good test of the model. One new ecologically valid detail is that the precue always appeared at the location of the target from the *previous* trial, comparable to how when a computer display changes the eyes are often initially positioned at the same location where they were positioned before the change.

## 4.2. Model Validation

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. When words were grouped randomly, the model did a very good job of predicting the search times (AAE = 6.5%), number of fixations (AAE = 3.0%), and saccade distances (AAE = 5.9%). In all three measures, when only considering the random conditions, the model predicted the observed data with accuracies well below an AAE of 10%.

With the exception of saccade distance (AAE = 8.9%), the model did not accurately predict human performance in the semantic conditions (search time per trial AAE = 42.6%; number of fixations per trial AAE = 37.3%). Since the cognitive model had no representation for semantic information, it is not surprising that the model makes more fixations than people, who were observed passing over an entire group with a single fixation that evidently captured the semantic content of the group.

But the model does a good job of predicting the human data when semantic grouping is removed from the layout, and is well within our target margin-of-error of 10%. The active vision model is validated by showing its ability to predict visual search behavior a priori for a task that includes a larger layout, more words, and a different word set. These results suggest that the model would be an appropriate starting place for modeling more complex tasks and more complex stimuli.

The correct and incorrect predictions made by the model in the semantically-grouped conditions provide guidance for future work. That the model correctly predicts saccade distances for both semantic and non-semantic layouts suggests that many details of the model, in this case the basis for saccade destination selection, will continue to be useful and correct in new contexts. 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. This comprehensive model of active-vision for visual search extends our understanding of how people search computer displays and provides a basis for a priori predictive modeling.

## 5. DISCUSSION

This paper presents a computational model of active vision that is a substantial push towards a unified model for predicting visual search in human-computer interaction tasks. Such a model is needed for automated interface analysis tools, such as CogTool (John et al., 2004), which do not yet include fully-developed active-vision subsystems. In that the model is built into a computer program that generates predictions, the model is demonstrated to have achieved a level of completeness and to be sufficient to account for the major processes involved in active vision.

### 5.1. Contributions to Cognitive Modeling

This research moves the field of HCI closer to a 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 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. The model of visual search proposed here accounts for eye movement data, from fixation duration to scanpaths, by employing visual search strategies and constraints which are informed by the eye movement data itself as well as previous research. The model suggests answers to the four questions of active vision, as follows.

Question #1: When do the eyes move? A process-monitoring saccade-initiation strategy accurately predicts fixation durations. The simulated flow of information through perceptual processors, modeled here via the transduction times in the EPIC cognitive architecture, works well to explain observed fixation durations. While other hypotheses of saccade initiation (Hooge & Erkelens, 1996) are not ruled out by this research, the process-monitoring strategy works very well, without additional mechanisms or parameters that would be necessary to implement the other saccade initiation strategies.

Question #2: What do the eyes fixate next? Perhaps most importantly, 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 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.

Question #3: What can be perceived during a fixation? Items near the point of gaze are more likely to be perceived, and different features will be perceived at different eccentricities. The modeling showed that some limitations on the amount of information that can be processed in a fixation are best explained by text-encoding errors rather than by varying the effective field of view. A text-encoding error rate of 10% accurately predicts human performance across a range of tasks.

Question #4: What information is integrated across fixations? At the very least, the memory of previously visited regions is retained. The active-vision model integrates this memory across fixations to guide the search toward unexplored areas and, in doing so, accurately reproduces saccade distances and scanpaths

This research informs the process of building computational models of visual search. The original CVC model (Hornof, 2004) predicted the observed search times well, but did not predict eye movement data as well as the active-vision model. This is not surprising since the original model was not informed by eye movement analysis. This discrepancy suggests a strong need for utilizing eye movement data when building models of visual tasks.

The active-vision model instantiates and integrates previous claims in the research literature (such as, Bertera & Rayner, 2000; Hooge & Erkelens, 1996; Motter & Belky, 1998). The modeling reinforces and refines previous claims, such as Bertera and Rayner's (2000) claim that the effective field of view does not change as a function of density, by showing that density-based text-encoding errors explain text search better than changing the effective field of view as a function of text density. What is more, the tasks used to inform the active-vision model are more ecologically valid than those used by Bertera and Rayner, who used randomly arranged single letters.

## 5.2. Informing the Development of Automated Interface Analysis Tools

An aim of this research is to provide the theoretical underpinnings of visual search needed for automated interface analysis tools by providing a useful method for 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. One way to improve the predictive power of interface analysis tools with respect to predicting users' visual interaction is to enhance existing interface analysis tools with a robust model of visual search based the active-vision model presented here. The active-vision model predicts the visual search of text-based displays with an acceptable level of accuracy for engineering-based models and, as such, will be useful in automated interface analysis tools.

Evidence of the need for an active-vision model of visual search and evidence of the potential impact of such a model is demonstrated by the fact that components of the active-vision model have already been incorporated into automated analysis tools. CogTool-Explorer (Teo & John, 2008, 2010) incorporates components that were described in early reports of this modeling project (Halverson & Hornof, 2007), such as a strategy that uses the distances between objects to decide which object to examine next. The accuracy of CogTool-Explorer's visual search predictions improved after integrating components of the active-vision model. However, CogTool-Explorer and the computational model on which it is partially based, SNIF-ACT (Fu & Pirolli, 2007), do not embrace all aspects of active vision. These tools do not simulate eye movements and do not simulate visual perception with the same fidelity as the active-vision model. For example, in CogTool-Explorer and SNIF-ACT, all visual objects on a web page have equal visual saliency regardless of their location on the page.

## 5.3. Future Directions

While the model 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 model proposed here answers four fundamental questions that should be answered by any predictive model of visual search, but more work is needed to move the active-vision model beyond the search of structured layouts of text.

### Integration of Models of Visual Search

Current models of visual search cannot accurately predict how people will interact with some of the complex visual layouts used in today's computer applications. Individual models exist that separately instantiate the different strategies that people will use when visually searching different layout elements. 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 been realized in cognitive architectures such as EPIC (Kieras & Meyer, 1997) and ACT-R (Anderson et al., 1997). However, the independence of the models within each architecture has a *disunifying* effect if there is no unification of the theory embedded in the individual models. Work is needed to integrate 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 lends itself to the modeling of active vision because it emphasizes the use of the visual-perceptual and ocular-motor processes that are central to active vision, but these processes can be recruited in different ways to answer the same questions.

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 and architectural developments 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 of text is that text can be perceived up to 1° of visual angle from the center of gaze, but this could be replaced with a continuous function in which the availability of text (or some other feature) degrades continuously as a function of eccentricity—a more veridical account of how human perception really works.

Search behavior for some tasks can be produced equally well by relying on either a fixate-nearby strategy (as in the active-vision model) or by relying on the continuous availability functions. Whereas the active-vision model explains where the eyes move using the objects' location, other EPIC models (Kieras & Marshall, 2006; Kieras, 2003) use the continuous availability functions to explain where the eyes move—to objects with available task-relevant features. Further research is required to determine whether both approaches are necessary to predict all scanpaths, how the two methods might be integrated, or whether one approach might subsume the other. Such integration will be useful for extending our active-vision model to a wider variety of visual tasks, and to develop a truly comprehensive model of visual search.

### Integration with Models of Semantic Search

While the active-vision model explained some of the eye movement behavior in the semantic grouping task, the model did not explain how semantics influenced search. Research and modeling has provided much insight into how semantics can guide visual search (Brumby & Howes, 2004; Brumby & Howes, 2008; Fu & Pirolli, 2007). Computational models such as Brumby and Howes's model of the interdependence of link assessment (2004) and Fu and Pirolli's SNIF-ACT 2.0 (2007) accounted for some of the effects of semantics on visual search, but these models use simplified scanpaths and do not account for all aspects of active vision, such as how people select saccade destinations. The integration of these two models with our active-vision model would be a substantial contribution to predictive modeling in HCI.

## 6. 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 a theory of visual search in HCI, and will ultimately provide a means of predicting visual search behavior for 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 better predict visual search behavior. The results from the modeling extend and solidify an understanding of active vision in a computationally instantiated theory that is useful for both future HCI and cognitive psychology research. 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 that will predict that interaction.

## NOTES

*Background.* This article is based on the Ph.D. thesis of the first author.

***Acknowledgments.*** Thanks to Beth Halverson, Richard Young, Dario Salvucci, and an anonymous reviewer for their feedback on this article.

***Support.*** This work was supported by Office of Naval Research grants N00014-02-10440 and N00014-06-10054, and by National Science Foundation grants IIS-0308244; all were awarded to the University of Oregon with Anthony J. Hornof as the principal investigator. The opinions expressed here do not necessarily reflect the views of the funding agencies.

***Authors' Present Addresses.*** Tim Halverson, Air Force Research Laboratory, Mesa Research Site, 6030 South Kent St., Mesa, AZ 85212. Email: thalverson@gmail.com.

Anthony J. Hornof, Department of Computer and Information Science, 1202 University of Oregon, Eugene, OR 97403–1202. E-mail: hornof@cs.uoregon.edu.

***HCI Editorial Record.***

## REFERENCES

Altarriba, J., Kambe, G., Pollatsek, A., & Rayner (2001). Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent spanish-english bilinguals. *Perception and Psychophysics, 63*(5), 875-890.

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.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226, 177-178.

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*, 46-51. Pittsburgh, PA: University of Pittsburgh

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, 190-196. New York: ACM.

Casco, C., & Campana, G. (1999). Spatial interactions in simple and combined-feature visual search. *Spatial Vision, 12*(4), 467-483.

Findlay, J. M. & Brown, V. (2006). Eye scanning of multi-element displays: I. Scanpath planning. *Vision Research, 46*, 179-195.

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*, 124-129. Pittsburgh, PA: University of Pittsburgh

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*, 1860-1864. Santa Monica, CA: Human Factors and Ergonomics Society.

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*, 431-434. New York: ACM.

Henderson, J. M. & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15*(3), 566-573.

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,* 249-256. New York: ACM.

Hornof, A. J. & Halverson, T. (2002). Cleaning up systematic error in eye tracking data by using required fixation locations. *Behavior Research Methods, Instruments, and Computers*, 34(4), 592-604.

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

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Direction in Psychological Science, 5*(3), 94-100.

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*, 455-462. New York: ACM.

Kieras, D.E. (2004). *EPIC Architecture Principles of Operation.* Retrieved from ftp://www.eecs.umich.edu/people/kieras/EPIC/EPICPrinOp.pdf

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*, 423-428. Mahwah, NJ: Lawrence Erlbaum Associates.

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. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

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 IFIP International Conference on Human-Computer Interaction*, 317-321. Amsterdam: North-Holland.

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.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocol. *Proceedings of the Eye Tracking Research and Applications Symposium*, 71-78. New York: ACM.

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*, 79-84. New York: ACM.

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, 950-954. Mahwah, NJ: Lawrence Erlbaum Associates.

Teo, L., & John, B. E. (2010). The evolution of a goal-directed exploration model: Effects of information scent and GoBack utility on successful exploration. In D. D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 10th International Conference on Cognitive Modeling,* Philadelphia, PA: Drexel University, 253-258.

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*, 411-420. New York: ACM.

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 model 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.

Yang, S. N. (2009). Effects of gaze-contingent text changes on fixation duration in reading. *Vision Research, 49*(23), 2843-2855.

## FIGURE CAPTIONS

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

**Figure 2. Example EPIC production rule that illustrates the selection of a saccade destination and prepares an eye movement to that location.**

**Figure 3. A mixed density layout. All angle measurements are in degrees of visual angle. The gray text did not appear during the experiment.**

**Figure 4. A layout without group labels from Hornof's (2001) CVC search task. The gray text did not appear during the experiment.**

**Figure 5. Mean fixation durations observed and predicted by the process-monitoring model (PM) and minimum fixation duration model (MFD) for the mixed density task. The average absolute error (AAE) of the PM model is 10.0% and the MFD model is 65.5%.**

**Figure 6. Flowchart of the production rules for the process-monitoring strategy.**

**Figure 7. Mean number of fixations per trial observed, predicted by the Text-Encoding Error (TEE) model and the Reduced-Region (RR) model for the mixed-density task. The AAE of the TEE model is 8.8% and the RR model is 21.1%.**

**Figure 8. Saccade distance observed in the CVC search task, predicted by the original CVC search task model and predicted by the fixate-Nearby (FN) Model. The AAE of the original model is 43.3% and of the FN model is 4.2%.**

**Figure 9. The most commonly observed scanpaths in the CVC search task for six-group layouts, and how often each path was taken by the participants (Observed), predicted by the fixate-nearby (FN) model, and predicted by the original model. The dashed boxes emphasize the good fits of the FN model.**

**Figure 10. A layout from the semantic grouping experiment, with semantically-cohesive groups, group labels, and meta-groups, annotated with three measurements of visual angle.**

# FIGURES

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

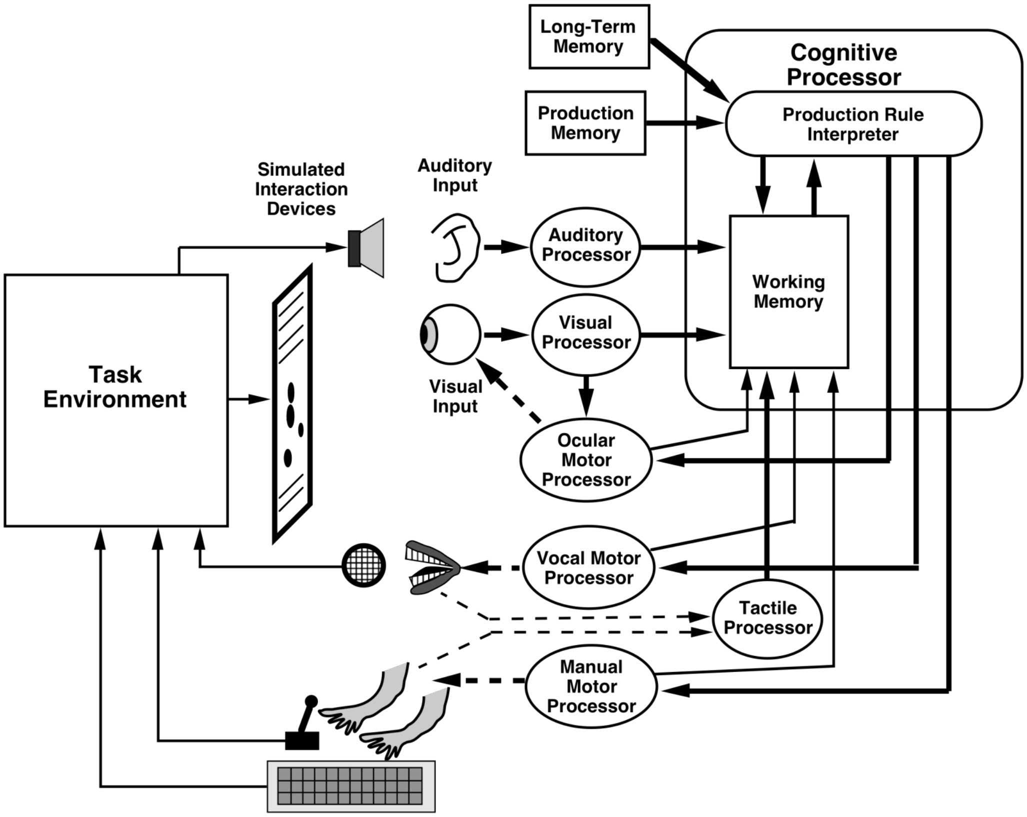
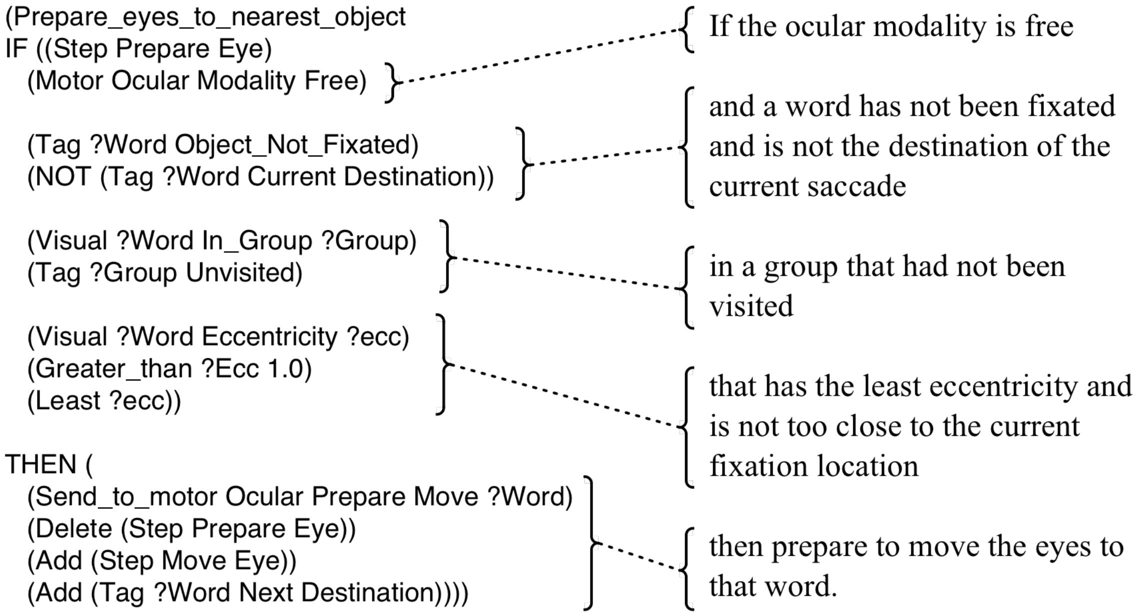


**Figure 2. Example EPIC production rule that illustrates the selection of a saccade destination and prepares an eye movement to that location.**



```
(Prepare_eyes_to_nearest_object
IF ((Step Prepare Eye)
   (Motor Ocular Modality Free)

   (Tag ?Word Object_Not_Fixated)
   (NOT (Tag ?Word Current Destination))

   (Visual ?Word In_Group ?Group)
   (Tag ?Group Unvisited)

   (Visual ?Word Eccentricity ?ecc)
   (Greater_than ?Ecc 1.0)
   (Least ?ecc))

THEN (
   (Send_to_motor Ocular Prepare Move ?Word)
   (Delete (Step Prepare Eye))
   (Add (Step Move Eye))
   (Add (Tag ?Word Next Destination))))
```

If the ocular modality is free

and a word has not been fixated and is not the destination of the current saccade

in a group that had not been visited

that has the least eccentricity and is not too close to the current fixation location

then prepare to move the eyes to that word.

**Figure 3. A mixed density layout. All angle measurements are in degrees of visual angle. The gray text did not appear during the experiment.**

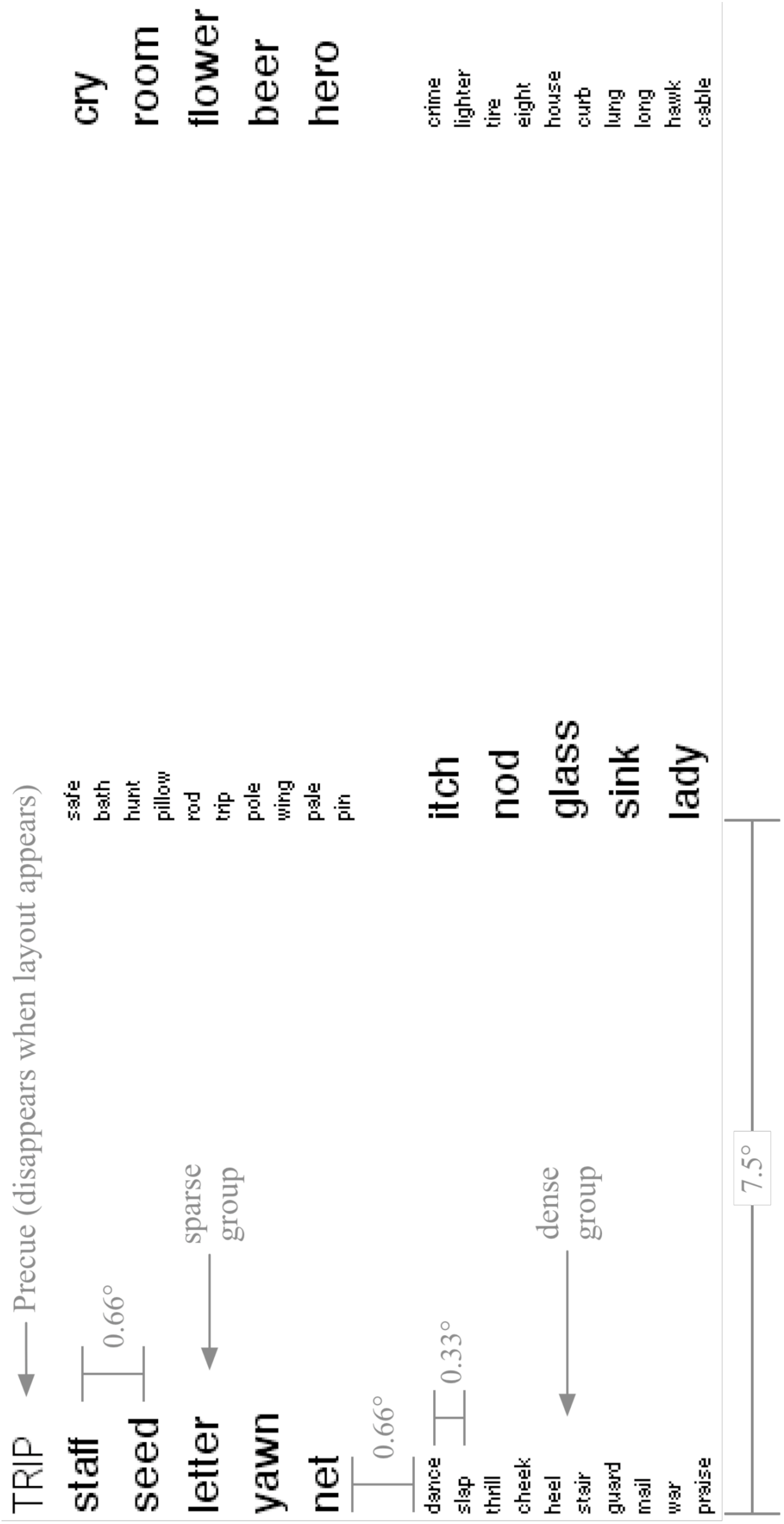


**Figure 4. A layout without group labels from Hornof's (2001) CVC search task. The gray text did not appear during the experiment.**

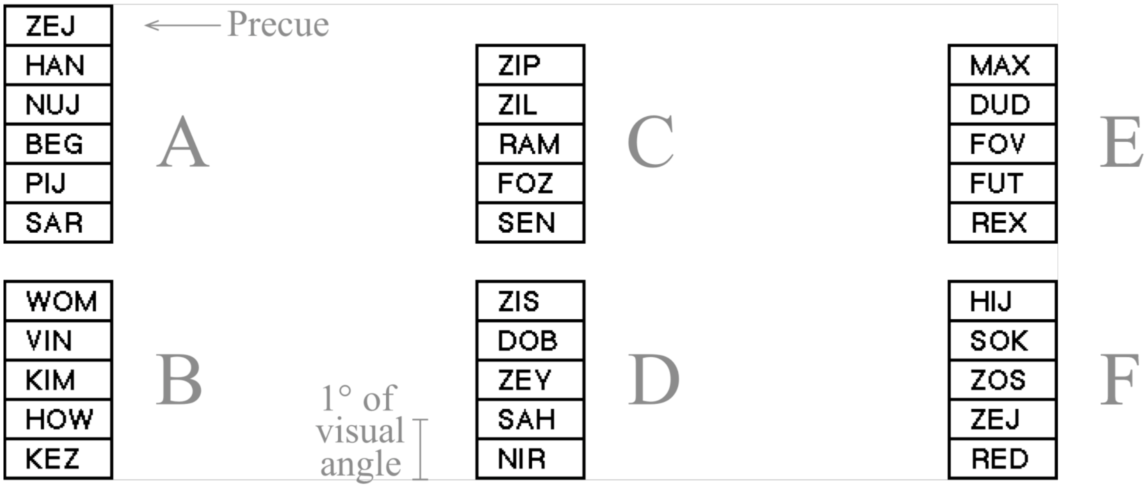


**Figure 5. Mean fixation durations observed and predicted by the process-monitoring model (PM) and minimum fixation duration model (MFD) for the mixed density task. The average absolute error (AAE) of the PM model is 10.0% and the MFD model is 65.5%.**

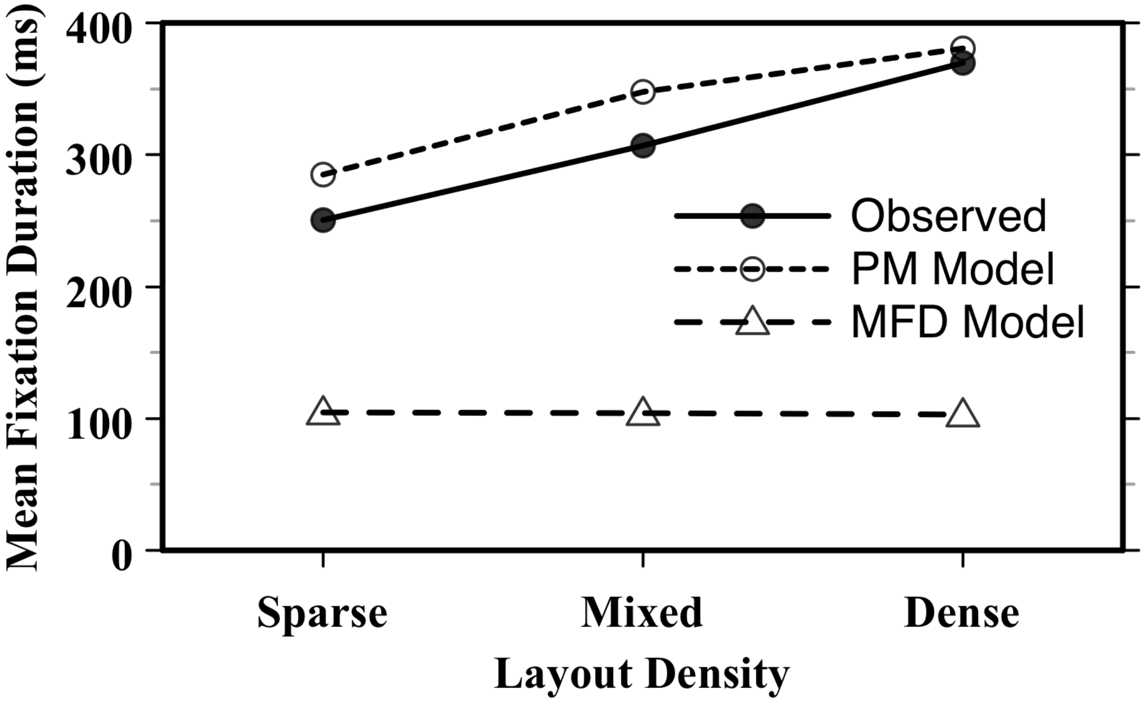


**Figure 6. Flowchart of the production rules for the process-monitoring strategy.**

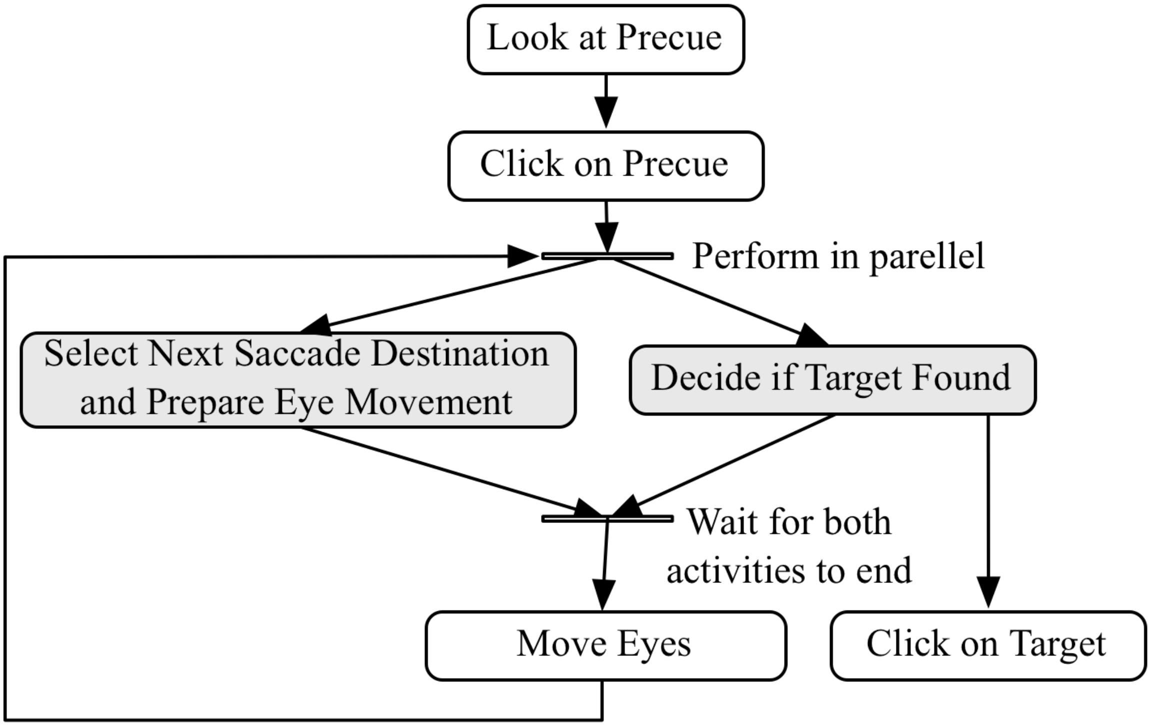


**Figure 7. Mean number of fixations per trial observed, predicted by the Text-Encoding Error (TEE) model and the Reduced-Region (RR) model for the mixed-density task. The AAE of the TEE model is 8.8% and the RR model is 21.1%.**

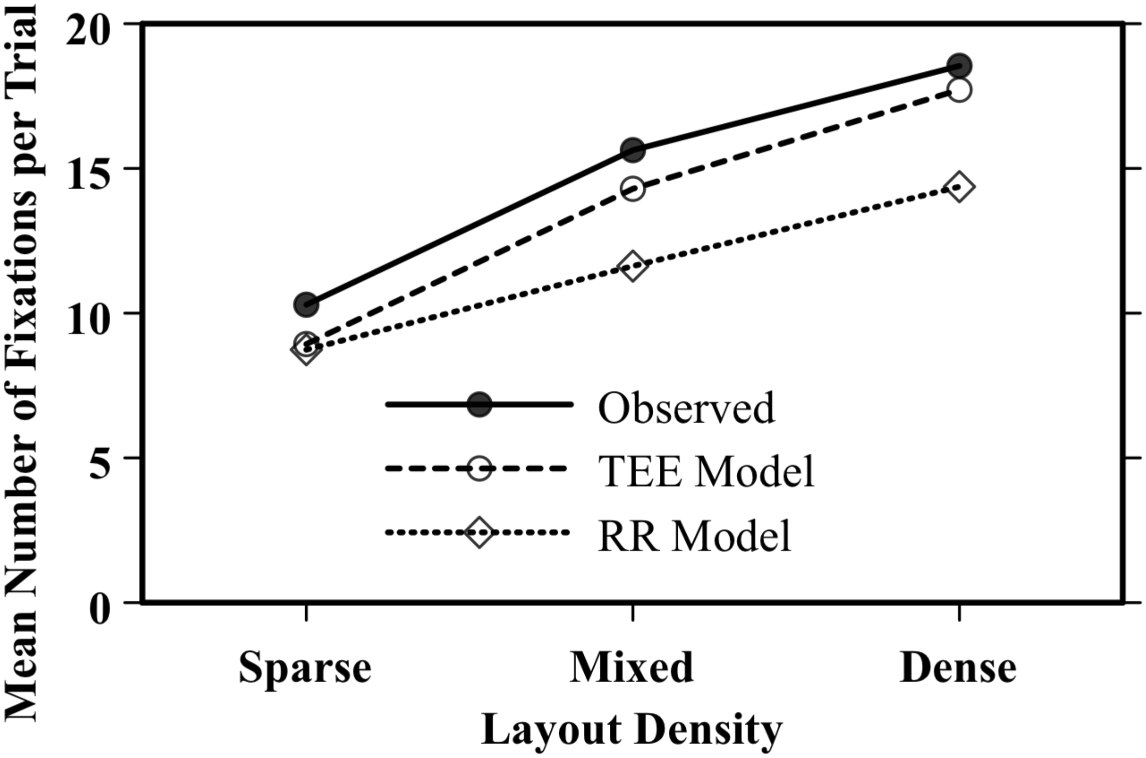


**Figure 8. Saccade distance observed in the CVC search task, predicted by the original CVC search task model and predicted by the fixate-Nearby (FN) Model. The AAE of the original model is 43.3% and of the FN model is 4.2%.**

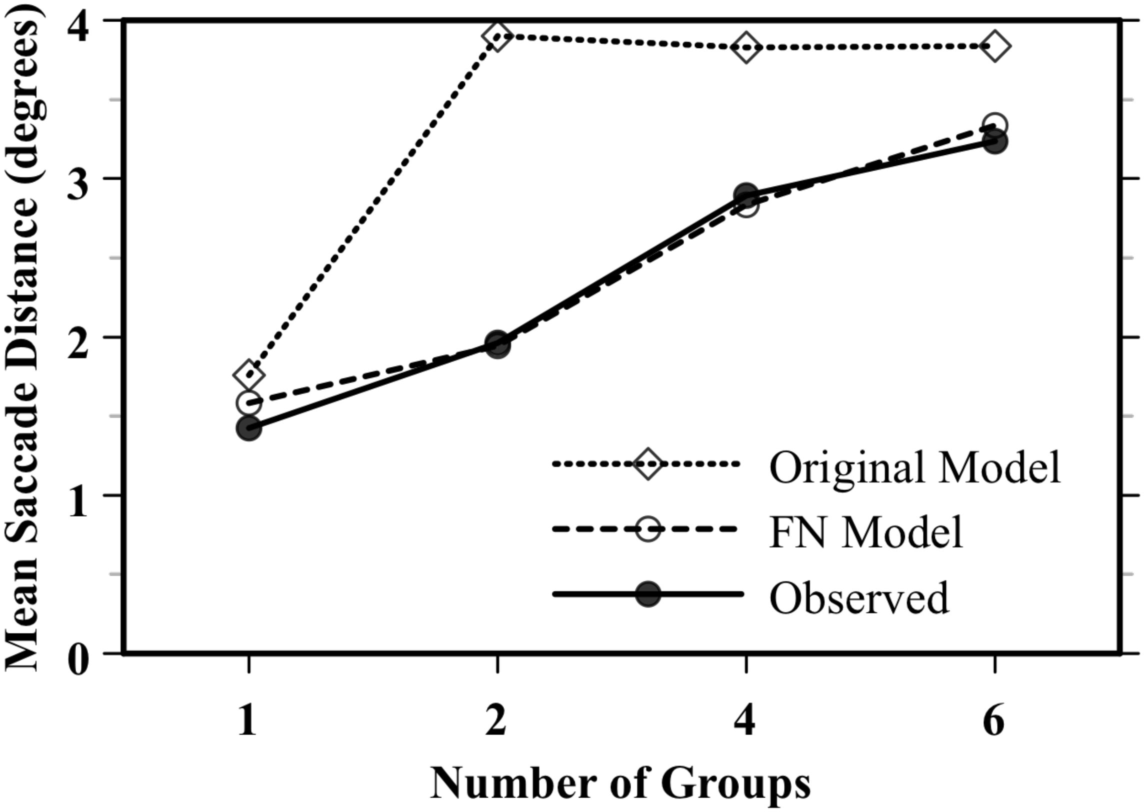
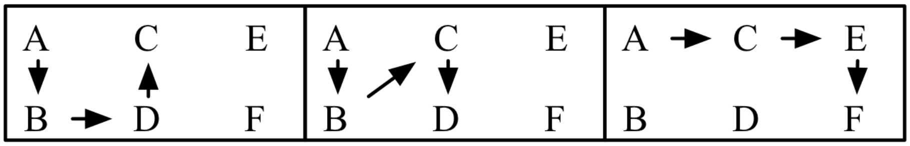
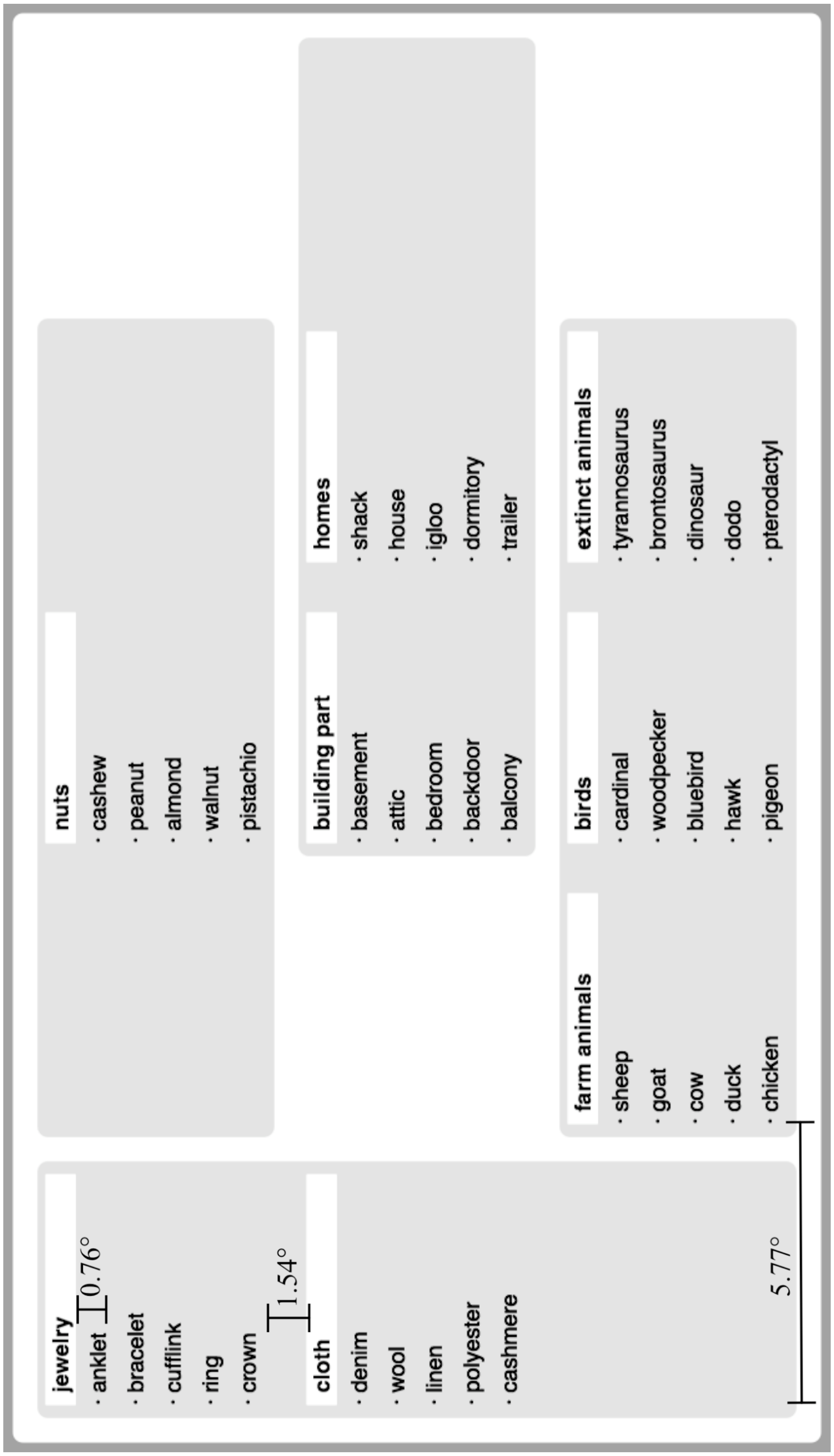


**Figure 9. The most commonly observed scanpaths in the CVC search task for six-group layouts, and how often each path was taken by the participants (Observed), predicted by the fixate-nearby (FN) model, and predicted by the original model. The dashed boxes emphasize the good fits of the FN model.**



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

**Figure 10. A layout from the semantic grouping experiment, with semantically-cohesive groups, group labels, and meta-groups, annotated with three measurements of visual angle.**


jewelry
· anklet
· bracelet
· cufflink
· ring
· crown
0.76°
1.54°
nuts
· cashew
· peanut
· almond
· walnut
· pistachio
cloth
· denim
· wool
· linen
· polyester
· cashmere
building part
· basement
· attic
· bedroom
· backdoor
· balcony
homes
· shack
· house
· igloo
· dormitory
· trailer
5.77°
farm animals
· sheep
· goat
· cow
· duck
· chicken
birds
· cardinal
· woodpecker
· bluebird
· hawk
· pigeon
extinct animals
· tyrannosaurus
· brontosaurus
· dinosaur
· dodo
· pterodactyl