AN "ACTIVE VISION" COMPUTATIONAL MODEL OF VISUAL SEARCH FOR HUMAN-COMPUTER INTERACTION

by

TIMOTHY E. HALVERSON

A DISSERTATION

Presented to the Department of Computer and Information Science
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2008

"An 'Active Vision' Computational Model of Visual Search for Human-Computer Interaction," a dissertation prepared by Timothy E. Halverson in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Computer and Information Science. This dissertation has been approved and accepted by:

Anthony J. Hornof, Chair of the Examining Committee

Date

Committee in Charge: Dr. Anthony J. Hornof, Chair
Dr. Arthur Farley
Dr. Stuart Faulk
Dr. Edward Vogel
Dr. Michal Young

Accepted by:

Dean of the Graduate School

An Abstract of the Dissertation of

Timothy E. Halverson for the degree of Doctor of Philosophy

in the Department of Computer and Information Science to be taken December 2008

Title: AN "ACTIVE VISION" COMPUTATIONAL MODEL OF VISUAL SEARCH FOR HUMAN-COMPUTER INTERACTION

Approved: ____

Anthony J. Hornof

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 dissertation investigates visual search through experiments and computational cognitive modeling. 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 dissertation 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, the answers to which are important to both practitioners and researchers in HCI: What can be perceived in a fixation? When do the eyes move? Where do the eyes move? What information is integrated between eye movements?

This dissertation presents a principled progression of the development of a computational model of active vision. Three experiments were conducted that investigate the effects of visual layout properties: density, color, and word meaning. The experimental results provide a better understanding of how these factors affect human-computer visual interaction. Three sets of data, two from the experiments reported here, were accurately modeled in the EPIC (Executive Process-Interactive Control) cognitive architecture. This work extends the practice of computational cognitive modeling by (a) informing the process of developing computational models through the use of eye movement data and (b) providing the first detailed instantiation of the theory of active vision in a computational framework. This instantiation allows us to better understand (a) the effects and interactions of visual search processes and (b) 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.

CURRICULUM VITAE

NAME OF AUTHOR: Tim Halverson

PLACE OF BIRTH: Iowa City, IA, USA

DATE OF BIRTH: March 31st, 1971

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA

Lane Community College, Eugene, OR, USA

Dickinson State University, Dickinson, ND, USA

DEGREES AWARDED:

Doctor of Philosophy, Computer and Information Science, 2008, University of Oregon

Master of Science, Computer and Information Science, 2003, University of Oregon

Bachelor of Science, Computer and Information Science, 2001, University of Oregon

AREAS OF SPECIAL INTEREST:

Human-Computer Interaction

Cognitive Science

PROFESSIONAL EXPERIENCE:

Cognitive Modeling and Eye Tracking Lab Manager, Dr. Anthony Hornof, September 2006-Current

Graduate Research Assistant, University of Oregon, 2001-Current

Technical Manager and Computer Repair Technician, Galaxy Hardware Publishing 1993-2000
Automatic Tracking Radar Specialist, United States Air Force, 1989-1993

GRANTS, AWARDS AND HONORS:

Henry V. Howe Scholarship, University of Oregon, 2007-2008

PUBLICATIONS:

Hornof, A. J., Halverson, T., Isaacson, A., & Brown, Erik (2008). Transforming Object Locations on a 2D Visual Display Into Cued Locations in 3D Auditory Space. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, New York, 1170-1174.

Halverson, T. & Hornof, A. J. (2008). The effects of semantic grouping on visual search. *Proceedings of ACM CHI 2008: Conference on Human Factors in Computing Systems*, Florence, Italy, 3471-3476.

Hornof, A. J., Rogers, T., & Halverson, T. (2007). EyeMusic: Performing Live Music and Multimedia Compositions with Eye Movements. *NIME 2007: Conference on New Interfaces for Musical Expression*, New York, 299-300.

Halverson, T., & Hornof, A. J. (2007). A minimal model for predicting visual search in Human-Computer Interaction. *Proceedings of ACM CHI 2007: Conference on Human Factors in Computing Systems*, San Jose, CA, 431-434.

Halverson, T., & Hornof, A. J. (2006). Towards a Flexible, Reusable Model for Predicting Eye Movements During Visual Search of Text. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vancouver, BC, Canada, 1428-1433.

Brock, D., McClimens, B., Hornof, A. J., & Halverson, T. (2006). Cognitive Models of the Effect of Audio Cueing on Attentional Shifts in a Complex Multimodal Dual-Display Dual-Task. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*, Vancouver, BC, Canada, 1044-1049.

Halverson, T. (2006). Integrating Models of Human-Computer Visual Interaction. *ACM CHI 2006: Conference on Human Factors in Computing Systems*, Montréal, Québec, Canada, 1747-1750.

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

Halverson, T., & Hornof, A. J. (2004). Local density guides visual search: Sparse groups are first and faster. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting,* New Orleans, LA, 1860-1864.

Halverson, T., & Hornof, A. J. (2004). Strategy shifts in mixed-density search. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, Chicago, IL, 529-534.

Halverson, T., & Hornof, A. J. (2004). Link colors guide a search. *ACM CHI 2004: Conference on Human Factors in Computing Systems*, Vienna, Austria, 1367-1370.

Hornof, A. J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. *Proceedings of ACM CHI 2003: Conference on Human Factors in Computing Systems*, Ft. Lauderdale, FL, 249-156.

Hornof, A. J., & Halverson, T. (2003). Predicting cognitive strategies and eye movements in hierarchical visual search. *Proceedings of the Fifth International Conference on Cognitive Modeling*, Bamberg, Germany, 261-262.

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.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation through grant IIS-0308244 and through the Office of Naval Research through grants N00014-06-10054 and N00014-02-10440 to the University of Oregon with Anthony Hornof as the principal investigator. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants and do not necessarily represent the official views, opinions, or policy of the National Science Foundation or the Office of Naval Research.

I would like to thank my advisor, Anthony Hornof, for the many types of support he provided: intellectual, moral, and so much more. My thanks also go out to my academic “grandfather”, David Kieras. Thank you for your discussions about (and co-creation of) EPIC. I would also like to the thank the research communities that sparked and nurtured my interest: CHI, ICCM, and CogSci. And thanks to my peers that helped in so many ways, specifically: Duncan Brumby, Anna Cavendar, Ishwinder Kaur, Erik Brown, and Andy Isaacson.

To Minerva and all of my family for their support throughout the process:
Beth, my parents, my grandparents, Thom, Andy, Melanie, Lucy, Kristen, Jake,
and so many others.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Visual search is an important part of human-computer interaction (HCI). Users search familiar news web sites to locate new stories of interest. Users search the interfaces of new, unfamiliar desktop applications to familiarize themselves with 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.

Visually searching for information is quite often the fastest and most useful way of finding information in a variety of user interfaces. Functionality such as web search engines or the “Find” command found in many operating systems can be used to find items on a computer screen quickly. However, there are many instances in which visual search is more useful, such as (a) searching among many similar results where it is difficult to specify a search query to locate the desired target, such as examining web search engine results, (b) when an application does not include a find command, such as in video games, and (c) when the exact target is not known by the user, such as when looking for items that match some vague concept or goal. In these cases, if the eyes are used to search instead, fast eye movements can be used rather than slower typing, many visual objects can be evaluated by the user simultaneously, and words can be located that the user may not have generated spontaneously for textual searches.

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. Users encounter many challenges finding the information they seek when visually searching. Figure 1 shows a the home page of a health information web site. In this example layout, if a user is searching for drug interaction information, they may search the menu at the top and encounter many distracting images before they realize they need to perform a text search using the text field below the image of the number two with the orange background. As another example using a simpler interface, Figure 2 shows a page from the popular web site *craigslist.* Visual search of this page is affected both by the layout used (grouping, color, spacing, text size, etc.) and the strategies used while visually searching the page (slow item-by-item, using labels or not, following the columns or not).

Visual search is a particularly fascinating human activity to study because it requires a complex and rapid interplay among three major processes: perceptual, cognitive (decision), and motor. Perceptual processes affect how information from the environment reaches other processes. An example of a perceptual process that affects visual search is the *retinal availability*: The information that can be perceived through the eyes will vary as a function of the orientation of the eyes, because visual acuity is higher in the center of

Figure 1. A section from the Health.com home page. Users encounter many challenges when trying to location information like how two drugs may interact, such as distracting advertisements, headings in many different colors and typefaces, and numerous menus and visual hierarchies. (Source: http://www.health.com/health/ July 24, 2008)

Figure 2. A section from a Craigslist web page where people search for products, services, and jobs. For example, a user may be looking for a used bicycle. The layout and a user's visual search strategy will largely determine how long it takes the user to locate target information. (Source: http://eugene.craigslist.org/ March 6, 2008)

people's field of vision (i.e. the fovea). Cognitive processes affect how information from the environment is used, such as deciding what information in the periphery is most relevant and hence where to move the eyes next. Motor processes result in actions in relationship to the environment, such as orienting the eyes to a new location or clicking on a web page link. It can be difficult to understand and predict the effects and interactions these complex processes have on people's visual search behavior.

Computational cognitive modeling is a very powerful methodology for capturing, asserting, recording, and replaying plausible sets of interaction among the processes at work during visual search. In this dissertation, computational cognitive models are computer simulations of how people perform one or a set of tasks. Cognitive models of visual search have been built to simulate perceptual processes, such as proposals for how the visual features of objects are detected, where visual features are detected, and when visual features are detected. The models simulate cognitive processes such as strategies that people use when conducting visual search in various tasks, such as how people visually search groups of computer icons and how using various devices while driving affects people's visual scanning of the environment. The models simulate motor processes for a range of human motor activities, such as the time it takes to move the eyes or a cursor to an object on the screen.

The most important contribution of computational cognitive models to the field of HCI is that the models provide the 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 (the person using cognitive models) be knowledgeable in cognitive, perceptual, and motoric theories embedded in the tool. Designers of device and application interfaces could use such tools to evaluate their visual layouts, reducing the need for more expensive human user testing early in the development cycle. Potential usability problems in interfaces such as the web page shown in Figure 1 could be identified early, before time consuming human user testing. For example, an automated interface analysis tool could help designers discover that the health information web site shown in Figure 1 may not do a good job of supporting important tasks, such as finding different kinds of drug information. The tool could show likely visual search behavior if a user were to pursue such a task. A user might be likely to miss the small menu item “Drug Finder” near the upper-right corner. If a user does arrive at the large label “4 Ways to Search Conditions, Drugs, and Symptoms” in the middle of the page, that user might be likely to terminate their search on the sub-label “1 Browse Conditions” and never use the appropriate search box with the label “2 Look Up Drug Information.”

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 dissertation (and published earlier (Halverson & Hornof, 2007)). The research presented here is already helping current interface analysis tools to better simulate visual search in HCI tasks. However, CogTool does 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 the necessary processes to 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 using *fixations*. A fixation is the time when the eyes are relatively steady. 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 and the like. Any simulation of active vision must address four questions, the answers of which are important to designers and those interested in HCI. What information in the environment do we process during each fixation? Where do we move our eyes and why? When and why do we move our eyes? What information from the environment do we maintain across fixations?

The goal of this dissertation is to build a computational model of visual search in HCI that integrates a range of theory consistent with the notion of active vision. This research advances the usefulness and applicability of models of visual search based on original research of eye movements in visual search, a synthesis of existing literature, and principled methods for iteratively improving models of visual search based on eye movement data. The aim of the work presented in this dissertation is to improve the overall usability of computer systems by developing fundamental theory and understanding of how users visually interact with computers. Tools for the prediction of user interaction do not yet have an active vision model that can simulate people's visual search behavior. This research is one step towards that comprehensive, active vision model. This dissertation presents a detailed step-by-step principled progression of the development of a computational cognitive model of active vision. The models are explained and detailed with a variety of eye movements to provide answers to the questions put forth by active vision.

This dissertation advances the field of HCI — particularly with respect to computer science — as well as the field of cognitive science. This research benefits HCI and computer science by providing a theory-based foundation for engineering approaches to interface design, such as CogTool (John & Salvucci, 2005), to better predict how computer users visually interact with computers. Additionally, this research advances the field of computer science by improving an understanding of end users, specifically a computational instantiation of how people use their computers. This research also advances cognitive science by providing an instantiation of psychological theory on visual search in a computational model that is a testable integration of that theory.

The remainder of this dissertation is arranged as follows. Chapter II reviews literature on cognitive modeling and visual search that is relevant to a computational cognitive model of active vision for HCI. Chapter III presents three experiments, each of which is aimed at better understanding how people visually search structured, text-based layouts. Chapter IV discusses the development of an integrative computational cognitive model of visual search based on experiments discussed in the previous chapter. Chapter V summarizes the research, identifies key contributions, and suggests future directions.

CHAPTER II

LITERATURE REVIEW

In order to create a model of visual search that is useful to HCI, we must first consider the general premises of such a model. This section provides an overview of relevant literature on visual search and computational cognitive modeling.

This dissertation is concerned with how people visually search displays in everyday, natural tasks. Typically, when people use visual search in HCI the eyes are moved and independent shifts of attention (i.e. covert attention) are not used (Findlay & Gilchrist, 1998, 2003). Since different information is available depending on the orientation of the eyes (Bertera & Rayner, 2000; Findlay & Gilchrist, 2003), the movements of the eyes (as well as 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. Therefore, this dissertation will focus on 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. Others have been developed to predict and explain the effects of specific visual features in a broad range of visual search tasks. The following is a brief overview of relevant models 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. Guided Search predicts that the order in which objects are visually searched is affected by the following: the “strength” of objects’ visual features (e.g. their blueness, yellowness, steepness, and shallowness), the differences between objects, the spatial distance between objects, the similarity to the target, and the distance of objects from the center of gaze (i.e. the eccentricity).

The Area Activation Model (AAM ;Pomplun, Reingold & Shen, 2003) is also a computational model of how visual features direct visual attention. The AAM shares many characteristics with GS, 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.

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.

Understanding Cognitive Information Engineering (UCIE) is a computer model of human reasoning about graphs and tables (Lohse, 1993). UCIE is 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. The time to perceive objects, eye movements, and a limited memory for information provide constraints for the simulation of how people scan graphs and tables to answer questions about the graph or table.

EPIC (Executive Process-Interactive Control) is a framework for building computational models of 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 a variety of 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 or collection of theory that asserts eye movements are central to visual processes, including visual search (Findlay & Gilchrist, 2003). Active vision poses four central questions that would need to be addressed in a model of visual search: (a) *What* can be perceived when the eyes are relatively steady? (b) *When* and why do the eyes move? (c) *Where* do the eyes move next? (d) *What* information is integrated between eye movements?

2.2.1 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? A model of visual search must be able to predict if and when a user can perceive basic and complex features of a visual layout. 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, Forsyth & Wooding, 1990). UCIE (Lohse, 1993) assumes that all items within an unspecified radius are processed, but only the perception of 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.

Another possible assumption about what can be perceived is that the distance between the stimuli and the center of fixation influences what can be perceived. The Area Activation model (Pomplun, Reingold & Shen, 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. While the authors of the Area Activation model argue that the fixation field will vary based on task difficulty, foveal load, experience, density, and heterogeneity, their model does not predict a priori exactly how these factors will affect the field. Rather, the fixation field is estimated experimentally using the number of fixations required to search given stimuli. The problem with this method is that it requires the modeler to collect data for each set of stimuli, and it assumes that the fixation field does not vary across the visual layout even if the properties of objects vary.

Current models of visual search have not integrated all research findings that may be relevant to predicting what is perceived in a fixation during visual search. While some models assume that all items within a given region can be perceived in parallel, no differentiation is made for vertically or horizontally organized objects. Research has shown that the region from which information is used during a fixation may be larger in the horizontal dimension (Ojanpää, Näsänen & Kojo, 2002). As another example, Casco and Compana (1999) found that search time for objects defined by simple features was affected by density and not by spatial perturbation. Contrarily, the search time for objects defined by combined features was affected by spatial perturbation and not by density. While a predictive model need not address all observed phenomena to be useful, more research may be required to determine what can be perceived in a fixation based on the stimuli present.

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, Forsyth & Wooding, 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. Additionally, such a model would require little, if any, additional empirical work. As density does not seem to affect the effective field of view (the region from which information is used in a fixation) in visual search (Bertera & Rayner, 2000), we may also want to restrict our straightforward model to a set region around the center of fixation. 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, Forsyth & Wooding, 1990) and text (Hornof, 2004).

2.2.2 When Do the Eyes Move?

If a model predicts what a user can perceive in an interface within a fixation, the model must also account for when the eyes move. For example, 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. This explanation does not preclude preprogramming multiple fixation durations based on the stimuli encountered. 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; if previous fixations are longer than needed to perceive the stimuli, future fixation durations are shortened. 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. Two of these four explanations of fixation duration, strict process-monitoring and mixed control, require consideration of how long it takes to process objects in order to determine how long fixations will be. Therefore, to fully understand these last two explanations, additional information on the time to process visual objects is required.

Models vary considerably with respect to the how long it takes to visually process objects. Some assume a fixed time per object (Anderson, Matessa & Lebiere, 1997; Byrne, 2001; Wolfe, 1994) or strategy (Kieras, Wood & Meyer, 1997).

Models that assume a fixed time to process objects also tend to assume a short processing time. Processing an object takes 50 ms in Guided Search (Wolfe, 1994) and the same amount of time in models based on ACT-R's Visual Interface (Anderson, Matessa & Lebiere, 1997; Byrne, 2001). When eye movements are considered in Guided Search (Wolfe & Gancarz, 1996), the processing time per fixation is also fairly constant, inspecting four to five items in 200 to 250 ms.

Models that propose a varying time for processing objects do so in a variety of ways. Eye Movements and Movements of Attention (EMMA; Salvucci, 2001a) utilizes two properties of an object to determine the encoding time. The first is the probability of the object appearing in a layout (i.e. the normalized frequency). The second is the eccentricity of the object relative to the center of fixation. UCIE (Lohse, 1993) predicts processing time according to the number, proximity, and similarity of all objects within a limited range of fixation center. Models based on EPIC (Kieras & Meyer, 1997) generally assume a constant time to perceive each property of an object, but these times are determined independently for each feature. So, while the time to perceive each feature is generally constant, the time to perceive all properties of the object will vary with the set of features an object has.

Object processing time is one of the most non-standard properties across different models of visual search. Processing time in most models is a single parameter, with little differentiation for perceiving stimuli of different complexities.

2.2.3 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? For example, in Figure 1, will users first look at the deep blue and light blue menu bars at the top first or will they look at the large image promoting the “Statin Study”?

The path the eyes follow is usually referred to as the scanpath. A great deal of research has been conducted to determine the factors that influence the scanpath in visual search. Research pertaining to the scanpath attempts to understand what factors guide visual attention or the eyes. Understanding the scanpath is seen by many as the core to understanding visual search, and is the focus of many models of visual search.

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. 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, Reingold & Shen, 2003) is similar to Guided Search 2, except that search is guided to regions of greatest activation instead of items.

A great deal of research has been conducted to determine which features can guide where the eyes move in visual search (see Wolfe & Horowitz, 2004 for a review). Based on the strength of current evidence, Wolfe and Horowitz concluded that there are four features that can guide visual search: color, motion, orientation, and size. While other attributes may also guide visual search, there is either contradicting evidence or insufficient evidence to conclude that other features do guide search. Most of the research reviewed by Wolfe and Horowitz has investigated visual search without eye movements. That is, the use of covert attention was required of the participants. So, can such results be used to inform eye movements in models of visual search? Hopefully, yes. Findlay and Gilchrist (2003) argue that shifts of covert attention are associated with eye movements. If the two are associated, some or all factors that affect covert attention may also affect eye movements. However, even if it is assumed that much of the covert attention guidance phenomena can be directly applied to eye movements, it is still unclear to what extent or where in the visual field this guidance information is used. The literature lacks a clear specification of where in the visual field visual features can be used to guide search.

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.

Strategies play a major role in determining saccade destinations. Hierarchical menus have been found to motivate fundamentally different strategies than non-hierarchical menus (Hornof, 2004). The ordering of menu items, either alphabetically or functionally, decreases search time, and therefore may motivate fundamentally different strategies than randomly ordered menus (Card, 1982; Perlman, 1984; Somberg, 1987).

There has been substantial research on factors that influence the destination of the eyes in visual search, and a fair amount of this research is applicable to predictive models of visual search for HCI. Such models will have to consider both intrinsic features of objects in the layout and strategies. Based on past research, models of visual search should at least consider how scanpaths are influenced by motion, color, orientation, and size. It is not clear how the effects of these properties should be represented in the model. A common form of representation in the psychological models is the use of the activation map. However, models need not be limited to such representations as long as the models account for similar phenomena, such as density, distance from the center of fixation, and cumulative effects of multiple features.

2.2.4 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?

Important factors to consider include: (a) What is remembered between fixations? (b) How much can be remembered during visual search? (c) When is information forgotten that is used during visual search?

Models of visual search for HCI need to account for the ways in which memory may 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 and unnecessarily revisit some of 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).

Research suggests that neither visual nor verbal working memory have a major impact on the fundamental visual search processes. Visual and verbal working memories are limited capacity, temporary stores for visual and verbal information. These two memories show little overlap in functionality (Baddeley, 1986). Interestingly, research has shown that when verbal or visual working memory is occupied, visual search remains efficient. When people visually search for a shape while performing a task that is presumed to occupy visual working memory, the rate at which people visually searched was unaffected (Woodman, Vogel & Luck, 2001). A similar result is found when visually searching for a word while verbal working memory is filled (Logan, 1978, 1979). These findings do not mean that working memory does not affect visual search tasks at all. In general, for each modality, people can recall four things on average (Baddeley, 1986;

Luck & Vogel, 1997). If the visual search task requires storing multiple items in memory before search is terminated, limitations on working memory could require that the user terminate search early or forget items for later use.

However, the use of spatial memory (i.e. memory for locations in space), especially for where one has previously fixated, *does* appear to affect visual search. Research has shown that when spatial memory is occupied, visual search efficiency is reduced (Oh & Kim, 2004). A memory for previously fixated locations is also suggested by other research. 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, models of visual search do not incorporate limitations of memory. There are two major exceptions. The first is that many models of visual search assume a perfect memory for objects searched (Anderson, Matessa & Lebiere, 1997; Barbur, Forsyth & Wooding, 1990; Byrne, 2001; Hornof, 2004; Kieras & Meyer, 1997; Pomplun, Reingold & Shen, 2003; Wolfe, 1994). The details of how this memory is implemented varies by model, but in general all the models "tag" each item as it is inspected and then do not re-inspect the objects unless all items have been searched without locating the target. The second exception is that a popular cognitive architecture used to build computational cognitive models, ACT-R (Anderson et al., 2004), includes a rich representation of memory. In short, ACT-R assumes that memory "chunks" have a likelihood of being recalled based on several factors, including the amount of time that has passed since that information was perceived. This limitation has the potential to affect visual search. For example, if visual search takes too long, it may be more difficult to retrieve (or recall) information gathered earlier in the search.

2.2.5 Summary

Active Vision emphasizes the importance of eye movements. Active vision asserts that understanding where and when the eyes move, and how information gathered during eye movements is utilized, are critical for understanding vision and, in particular, visual search. The literature reviewed suggests that no one model of visual search provides answers to all of the questions of active vision. However, every question of active vision is addressed by at least one model. The proposed answers for active vision will be used in this dissertation, along with input from other literature and experimentation reported in this thesis, to help build a candidate active vision model of visual search.

2.3 Specific Visual Search Phenomena

The following three sections discuss research on factors affecting visual search that (a) are relevant to the design of user interfaces, (b) lend themselves well to building computational models, and (c) affect eye movements in very specific ways. These factors are density, color, and semantics.

2.3.1 Density of Visual Objects

One common feature used in interfaces to indicate importance and association is relative density of the visual objects (Mullet & Sano, 1995). Figure 3 shows one example

Figure 3. A section from a Craigslist web page. The two labeled sections illustrate differences in densities. The groups of words are sparse in Region 1 relative to Region 2. The black rectangles and labels were added for illustration.

of a web page that utilizes multiple densities. Most of the users' tasks in this layout will be to locate the appropriate listing category. For example, a user may want to find a boat for sale in Vancouver, Canada. In this web page, some of the groups of words (Region 1) are sparse and show the categories of listings on Craigslist. Other groups (Region 2) are more dense and show geographic regions for which there are listings. This visual layout uses density to not only show association, but also importance.

The density of items in a display is one factor that has been shown to affect visual search. Bertera and Rayner (2000) varied the density of randomly placed characters in a search task. They found that search time decreased as the density of the characters increased. In addition, they estimated that the number of items processed per fixation increased as the density of the items increased. Mackworth (1976) showed similar results in a study in which participants searched for a square among uniformly distributed circles on a scrolling vertical strip. Ojanpää, Näsänen, and Kojo (2002) studied the effect of spacing on the visual search of word lists and found that, as the vertical spacing between words increased (i.e. as density decreased), search time also increased.

Density has also been shown, to a lesser extent, to affect people's visual search strategies with pictorial stimuli. Studies have found that the eyes tend to move to stimuli that are likely to be "more informative." One definition of "more informative" in pictorial stimuli proposed by Mackworth and Morandi (1967) is regions having greater contour. For example, with geometric shapes, angles are considered more informative than straight lines. Yet, it is not readily known how to predict a priori which of two stimuli are more informative. One plausible factor of "informativeness" is local density. It may be that regions with a higher local density are more likely to be searched earlier since they contain more information.

2.3.2 Link Colors

Color is another common feature used in interfaces to indicate group association. For example, a common Web design technique is to vary hypertext link color. One convention is to set unvisited links to blue and visited links to red. Figure 4 shows an example of a task that benefits from differentiated link colors. The idea is that the colors will help users to focus their search on unvisited links, increasing the efficiency of visual search. This convention of differentiating unvisited and visited links has some support from observational studies (Nielsen, 2007; Spool, Scanlong, Schroeder, Snyder & DeAngelo, 1999), but lacks empirical work showing the effect of text color on people's visual search strategies.

The effects of color on visual search is widely studied (Brawn & Snowden, 1999; Christ, 1975; Shen, Reingold & Pomplun, 2003; Shih & Sperling, 1996; Treisman, 1998; Zohary & Hochstein, 1989). However, the visual search literature does not directly address the application of the above guideline. Specifically, there are few if any previous studies of visual search of colored text.

Figure 4. Visited links (yellow) and unvisited links (blue) for categories of children’s book at Powell’s Books.
(Adapted from www.powells.com/ on May 16, 2008.)

Knowing the target color reduces the number of items that need to be searched and thus reduces the time needed to find the target. Previous research has shown that color is available early in the visual system (i.e pre-attentively) and that people can constrain visual search to items with the most informative features (Shen, Reingold & Pomplun, 2000). Therefore, knowing the color of the target of search has the potential to greatly decrease search time (Treisman, 1998).

2.3.3 Semantically Grouped Displays

In addition to being visually grouped, words in a layout can be grouped based on their semantics; that is, words can be grouped based on relationships between the words (e.g. synonyms, hypernyms). Categorical grouping, a type of semantic grouping, is found on many web sites. For example, as shown in Figure 3, the popular classifieds web site craigslist.org connects semantically related items by both proximity and visual cues, such as the grid structure, salient labels, and regions associated with meta-groups.

Previous research has shown how group labels, akin to the word "Canada" in Figure 3, affect the selection of visual search strategies and the execution of the those strategies (Hornof, 2004). In Hornof's research, the participants knew what the exact text of the label and target word for which they were searching. This research found that people use different visual search strategies based on the presence or absence of labels for groups of words. Specifically, when group labels were present, people tended to constrain their initial search to the labels. Perhaps more importantly, Hornof found that the execution of these strategies were substantially different, with labels motivating a more systematic search strategy.

Other research has shown how the semantic information in menu items affects visual search (Brumby & Howes, 2004). Brumby and Howes found that when searching a menu for an item described by a goal phrase not containing that word, people tend to search fewer items (a) when distractor menu items are less similar to the goal than when when the distractors are more similar to the goal and (b) when the target is more similar to the goal than when the target is less similar to the goal. In Brumby and Howes research, the words were not hierarchically related, either visually or semantically.

The previous research has investigated the effects of hierarchical organization and the effects of semantic similarity separately, but they have not addressed how the two phenomena may interact. What happens when the information is hierarchically organized into semantically related sets of words whose relationship is indicated with a meaningful label? How is visual search guided by the semantic content of group labels, or the grouping of the menu items? An experiment presented later in this thesis will investigate how a semantic hierarchy and a visual hierarchy affect users' visual search strategies.

2.4 Computational Cognitive Modeling

Computational cognitive models are computer programs that simulate 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 the behavior of people performing a task. (b) In an *a priori* fashion to predict how people will perform a task. This thesis reports on research that uses cognitive modeling in the former manner, to explain people's behavior.

Building cognitive models to explain users' behavior in a *post hoc* fashion has a rich history. In explanatory modeling, human data is collected and models are built to explain the observed behavior. Such explanatory cognitive models have been used to understand web link navigation behavior (Fu & Pirolli, 2007), driving behavior (Salvucci, 2001b), and time interval estimation (Taatgen, Rijn & Anderson, 2007). Explanatory models are useful in their own right, to expand our understanding of user behavior, but are also useful for informing *a priori* predictive models. For example, the explanatory modeling of driving behavior (Salvucci, 2001b) was used to inform the development of a predictive model of driver behavior while utilizing a cell phone (John, Salvucci, Centgraf, & Prevas, 2004). The models developed in this dissertation are consistent with other explanatory modeling research in that data is collected, the models are built, and lessons learned form the modeling are identified.

Cognitive models are often built using cognitive architectures, as is done in this research. The cognitive architectures instantiate and integrate psychological theory of human perceptual, cognitive, and motor processes in a framework that is used to build cognitive models. The architecture constrains the construction of the models by enforcing capabilities and constraints hypothesized to be similar to those of a human. 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 carryout a task, (b) the instantiated theory embodied in the cognitive architecture, and (c) parameters specified that effect the behavior of the architecture (e.g. the velocity of a saccadic eye movement). While the parameters can be task specific, the majority of the parameters are usually considered fixed across a wide variety of models. Simulations using these 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 modeling and the study of eye movements. The movements of the eyes provide a rich set of data on which models can be built and against which predicted eye movements of the cognitive model can be compared. In this way, the data can provide many constraints on the construction of the models. These include not only the reaction time, but also the number, extent, and timing of the observed eye movements. Therefore, it is beneficial to use a modeling framework that provides facilities for making explicit predictions of eye movements.

2.4.1 EPIC

EPIC (Executive Process-Interactive Control) is a cognitive architecture that instantiates and integrates theory of perceptual, motor, and cognitive constraints and benefits. Figure 5 shows the high-level architecture of EPIC (Kieras & Meyer, 1997). EPIC provides facilities for simulating the human and the task separately. 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 into corresponding visual, auditory, and tactical perceptual processors. Information from the perceptual processors are deposited into working memory. In the cognitive processor, information in working memory interacts with the the cognitive strategy (instantiated in the 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.

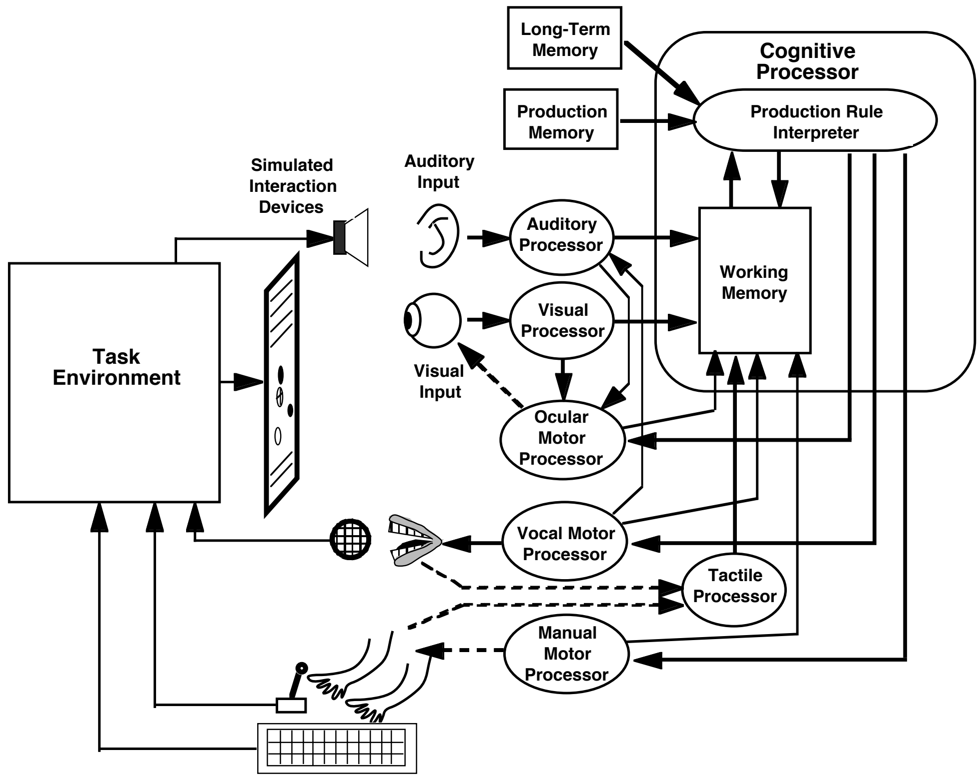


Figure 5. The high-level architecture of the EPIC cognitive architecture (Kieras & Meyer, 1997).

The perceptual and motor processors impose many constraints. Perceptual processors determine what information in the environment are potentially available "downstream" to the cognitive processor. Motor processors determine what information can change in the environment that will again become available to the cognitive processor through the perceptual processors. Particularly relevant to the research in this dissertation are the constraints imposed by the simulated eyes, which are: (a) The retina provides more detailed processing of information in foveal (central) vision. (b) 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 eye movements, called saccades, to gather additional information. The cognitive processor can send commands to the ocular-motor processor to initiate eye movements which take time to prepare and execute.

The strategies used by people to perform a task are instantiated in the production rules. Coordination of the motor processors is a primary responsibility of the strategies. One or more tasks may require the use of one or more processors. However, motor processors can only be controlled by one production rule at a time. For example, the eyes cannot be told to move to two locations at the same time. The timing and coordination of motor actions are vital constraints in EPIC.

EPIC permits multiple production rules to fire in parallel and places the serial bottleneck at the motor and (to a lesser extent) perceptual processors. Thus, an important aspect of modeling with EPIC is instantiating strategies that work with the constraints imposed by the peripheral processors (e.g. the eyes).

EPIC is a C++ programming framework for the Macintosh operating system. EPIC provides an extensive set of C++ classes that an analyst can use to build a program which simulates a user and a computer interface with which the simulated user interacts. The framework represent human cognitive, perceptual, and motor capabilities as discrete, but interacting, processes encoded in object-oriented classes. The design of the framework lends itself well to identifying where in the simulation predicted behavior is originating and thus more easily modify the framework to improve the fidelity of the simulation.

2.4.2 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 eye movements. As shown in Figure 5, the EPIC framework provides an ocular motor processor and visual processors. Within the EPIC framework exists theory that constrains the simulation of (a) what can be perceived in each fixation, (b) when visual features are perceived, and (c) how visual information can be integrated across fixations.

What can be perceived at any given moment is most constrained by EPIC's *retinal availability functions*. 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.

When visual features are perceived, relative to the timing of an object's appearance, is constrained by simulated delays in EPIC's 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 (described above), 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.

What gets integrated between fixations is a factor of memory decay time and the production rules. The perceptual parameters affect how information can be integrated between fixations as follows: When the eyes move away from an object, one or more visual properties may no longer be available. How far the eyes must move is determined by the properties' retinal availability function. When the retinal availability function determine that a feature is no longer available, the visual feature decays from sensory memory and then from visual working memory. Production rules can extend the "life" of visual features and object identity by creating a "tag" memory item before an object decays from visual working memory. However, copying memory items must be explicitly programmed into the production rules, usually based on task dependent criteria.

Where the eyes move is determined by the visual search strategies encoded in the production rules and the contents of working memory. The production rules explicitly state under which circumstances the eyes are moved. When the contents of working memory satisfy a production rule that moves the eyes, a motor movement command is sent to the ocular motor processor, which then initiates an eye movement to the object specified by the production rule.

All told, EPIC is very well-suited as a framework for simulating active vision in the context of HCI tasks. EPIC provides theory of visual-perceptual and ocular-motor processing that are useful for guiding the development of models of visual search in tasks that motivate eye movements.

2.5 Summary

The goal of this dissertation is to build a computational model of visual search that explains a variety of eye movement behavior in HCI tasks. Towards that goal, this dissertation draws on visual search literature, specifically literature related to active vision and previous models of visual search.

An *active vision* approach is adopted in this research to investigate people's visual search behaviors. Active vision is the notion that eye movements and the physical constraints imposed by the eyes must be considered when investigating visual processes. 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 dissertation will explore these issues through cognitive modeling.

Computational cognitive modeling is a useful method for understanding how people perform tasks, and EPIC is a cognitive architecture in which models especially well-suited to an active vision approach can be built. EPIC is a cognitive architecture, a software framework for building cognitive models, that instantiates constraints on the availability and timing of visual information through the simulation of human ocular motor activity and visual perceptual processes.

Now that the theoretical framework and background is in place, the next chapter will present the specific visual search experiments conducted as part of this dissertation work to provide the detailed reaction time and eye movement data needed to guide the development of the models of active vision.

CHAPTER III

EXPERIMENTS

Three experiments were conducted to (a) provide insight into how people visually search computer interfaces and (b) inform the building of computational cognitive models of such tasks. This chapter describes the experiments conducted for this dissertation, the analysis of eye movements recorded from these experiments, and the implications of these experiments for human-computer interaction.

One focus of this research is to understand how the physical arrangement of visual objects affects visual search. Figures 1 and 2 show examples of layouts with physical arrangements that will affect search. In these layouts: (a) the majority of the content are single words or short phrases, (b) the layouts are organized in a grid-like fashion, and (c) the visual properties of the objects that appear in the layout vary. The experiments reported in this dissertation are derived from real-world layouts such as those shown in Figures 1 and 2, but scaled down to provide more experimental control.

Each experiment investigated how a specific design decision affects users' visual search processes as revealed by reaction time and eye movements. The first experiment, mixed density displays, investigated the effects of varying the visual density of elements in a structured layout. The second experiment, link colors, investigated how differentiating potential targets from non-targets using color (as is done with web page links) affects visual search. The third experiment, semantically grouped displays, investigated how both semantic and visual grouping affect people's active vision.

3.1 Mixed Density Displays

An aim of this research is to ultimately predict how people visually search real-world interfaces. The density of content within interfaces, from the world wide web to printed brochures, varies substantially. This experiment is designed to explore how issues of density, discussed in Chapter I and researched by previous authors, should be incorporated into a general purpose, comprehensive, active vision model of visual search for HCI.

Density may be measured as either *overall density* or *local density*. Overall density is the number of items per unit of screen space over an entire layout. Local density is the number of items per unit of screen space within a visually distinct group. The first experiment was conducted to determine how varying local density in a layout affects visual search. Specifically, the effects of mixing of two local densities were investigated. Groups of words of two different densities, which will be referred to as sparse groups and dense groups, were presented alone or together.

It was hypothesized that search time per word would be greater in sparse layouts and that people search dense regions before sparse. As discussed in Chapter II, previous research has shown that densely packed shapes are visually searched faster than sparsely packed shapes. The initial assumption is that this will hold for words as well. Additionally, if people do visually search dense groups of words faster and they are motivated to find the target as quickly as possible while making as few errors as possible, then it is assumed that people will search dense regions first since this would result in finding the target faster on average. These two hypotheses are tested in the following experiment.

3.1.1 Method

This experiment investigated the effect of local density on the visual search of structured layouts. The remainder of this section discusses an experiment that investigates the effects of local density on users' visual interaction. The task is described and the results of an experiment are discussed.

3.1.1.1 Participants

Twenty-four people, 10 female and 14 male, ranging in age from 18 to 55 years of age (M = 24.5, SD = 7.9) from the University of Oregon and surrounding communities participated in the experiment. The participants were screened as follows: 18 years of age and older; a native English speaker; experienced using a computer; no learning disability; normal use of both hands; and normal or corrected-to-normal vision.

The participants' visual acuity, color vision, and stereo vision were verified. Visual acuity was verified using a Runge Near Point Card. The experimenter ensured that the participants had 20/20 vision. The print size of letters on the acuity chart which the participants were required to read were substantially smaller than the text they would later be required to read during the experiment. Normal color vision was verified using the H-R-R Pseudoisochromatic Plates, 3rd Edition. Normal stereo vision was verified using the Stereo Butterfly and Random Dot Test. All eye exam materials were acquired from Richmond Products in Albuquerque, New Mexico.

Participants were financially motivated to perform with high speed and accuracy. The participants were instructed that they should complete each trial quickly while making few errors. Participants were paid $10, plus a bonus that ranged from $0 to $4.54 based on their performance. Speed and accuracy were motivated as follows: The participant was initially given 7¢ to 16¢ for each trial, depending on the complexity of the screen. They lost 1¢ for each second after the trial started until they clicked on the target. If the participant clicked on something besides the target or if they moved the mouse before they found the target, the participant lost all of the money for that trial plus 5¢. The rewards and penalties were explained to the participants to motivate speed and accuracy.

3.1.1.2 Apparatus

Visual stimuli were presented on a ViewSonic VE170 LCD display set to 1280 width by 1024 height resolution at a distance of 61 cm, which resulted in 40 pixels per degree of visual angle. The experimental software ran on a 733 MHz Apple Power Macintosh G4 running OS X 10.2.6. The mouse was an Apple optical Pro Mouse, and the mouse tracking speed was set to the fourth highest in the mouse control panel. The visual stimuli were presented with custom software designed using the C++ programming language, Metrowerk's PowerPlant framework, and Metrowerk's Codewarrior development environment. The software was developed by the author as part of this dissertation work.

3.1.1.3 Eye Tracking

Since we are concerned with understanding visual search processes, the most directly observable and measurable events of interest are the eye movements of the experimental participants. Therefore, eye movements were recorded using an single-camera LC Technologies Eyegaze System, a 60 Hz pupil-center and corneal-reflection eye tracker. A chinrest was used to maintain a consistent eye-to-screen distance of 61 cm.

Eye trackers allow us to directly observe and record people's point of gaze. The participants pupil center and corneal reflection (the point on the eye closest to the camera) are monitored using infrared cameras and specialized video processing software. The vector between the pupil center and the corneal reflection are used to determine where the participant is looking.

3.1.1.3.1 Eye Movement Analysis The analysis and interpretation of eye movements can be challenging. There is an abundance of data, the data can be difficult to segment, and it's not always clear how to best use the data to answer the research questions. To address these issues: (a) Custom software was designed for studying eye data. (b) A well defined subset of the data was identified for analysis.

An extensible application, VizFix, was developed by the author to allow custom analysis of eye movement data as part of the work for this dissertation. The core of

VizFix provides a basic facilities to store, segment, and analyze eye movement data. Perhaps more importantly, VizFix is extendable through plugins that can read different data formats and provide additional, custom analysis for the identification of fixations and the counting, duration, and ordering of those fixations.

The first step in eye movement analysis is to identify meaningful physiological phenomena, fixations for this research, from the stream of gaze position samples provided by the eye tracker. In all analyses presented in this dissertation, fixations are identified using a dispersion-based algorithm (Salvucci & Goldberg, 2000). This algorithm was implemented in VizFix. The dispersion-based algorithm identifies fixations based on temporal and spatial relations between the gaze position samples provided by the eye tracker. In this dissertation, fixations are defined as a series of samples with locations within a 0.5° of visual angle radius of each other for a minimum of 100 ms. To accommodate noise in the eye tracker, this implementation of the dispersion-based algorithm assumes that a fixation continues if one sample occurs outside of the 0.5° radius as long the next sample occurs within that threshold.

Another important step early in the analysis process is to consider the best way to study the fixation data. Both temporal and spatial segmentation of the fixation data is required. Temporal segmentation includes choosing the right subset of fixations to analyze. Depending on the questions to be answered, different constraints must be placed on the selection of data to be analyzed. The spatial segmentation of fixations includes choosing the correct *regions of interest* for each analysis. Region of interest is a phrase commonly used in eye tracking analysis to denote (a) the locations in the visual display that are meaningful for the analysis and (b) the spatial scale at which the analysis will be conducted. For example, the regions of interest may be either individual words or groups of words (spatial scale) and may or may not include the whitespace surrounding the words (meaningful locations). The following are temporal and spatial constraints used in VizFix for the analysis of the eye movement data from the experiments presented in this research:

- The eye movement data analysis is limited to include gaze position samples starting from the first fixation that began after the user had initiated the visual search task and ending with the first fixation that stopped before the participant moved the mouse to click on the target of visual search. That is, only fixations that started after visual search started and fixations that ended before selection started are included in analyses. This constraint was imposed because the fixation that is ongoing when a layout appears may have been used to (a) process the visual stimuli present before the layout appears and (b) guide manual movements to initiate the next visual search task.

- When assigning fixations to regions of interest, a region is considered *visited* if one or more contiguous fixations fall within 1 degree of visual angle of the region. A region is considered *revisited* if that region has been previously visited.

- Unless otherwise stated, revisits are not included in the analyses because it is assumed that the participants' visual search behavior may differ within groups already visited. That is, if a person has previously fixated a group of items, decisions about that group of items may have already been made and may be recalled.

- The final region visited during visual search is not included in analyses because it is assumed that the participants' behavior will differ when a target is found. For example, the target will be fixated longer or more frequently than other items because the cursor must be moved to the target and the target clicked.

3.1.1.3.2 Eye Tracker Calibration Eye trackers need to be calibrated to each user in order to record accurate gaze position data. Adjusting for the unique characteristics of each users' eyes requires the user to look at several points on the screen on which the stimuli will be presented.

The need for calibrations and recalibrations can be a limiting factor in using an eye tracker for data collection. Too many calibrations can interrupt participants' task execution. A solution, implemented for the first time by the authors as part of this dissertation work, is the use of the required fixation location (RFL) technique (Hornof & Halverson, 2002).

The RFL technique provides an objective measure of eye tracker accuracy and can potentially reduce the number of calibrations required. An RFL is a location that the participant must fixate during the execution of the task in order to accomplish the task (e.g. clicking on a target). If the measured gaze location deviates too far from the RFL, a calibration is automatically initiated at the next appropriate time (e.g. between trials).

Other means of verifying eye tracker accuracy tend to be less reliable or more disruptive. The alternatives to using the RFL are: (a) Only a single initial calibration. This is problematic because the calibration can deteriorate over time, resulting in less reliable data. (b) Use the experimenter's subjective evaluation. While the experimenter can usually monitor the eye movements recorded by an eye tracker, it can be very difficult to consistently determine when the eye tracker's accuracy becomes problematic. Such subjective measures can result in unreliable data or unnecessary interruptions to recalibrate during the experiment. (c) Regular interruptions to recalibrate. The experimenter can interrupt the experiment at regular intervals (e.g. at the end of each block of trials) to recalibrate the eye tracker regardless of whether it is truly necessary. Of the three alternatives, this is the most desirable. However, this is an uninformed decision and can result in too few or too many recalibrations.

3.1.1.3 Stimuli

Figure 6 shows a sample layout from a mixed density display. Layouts always contained six groups of left-justified, vertically-listed black words (RGB values of 0, 0, 0) on a white background (RGB values of 255, 255, 255). Groups were sets of words

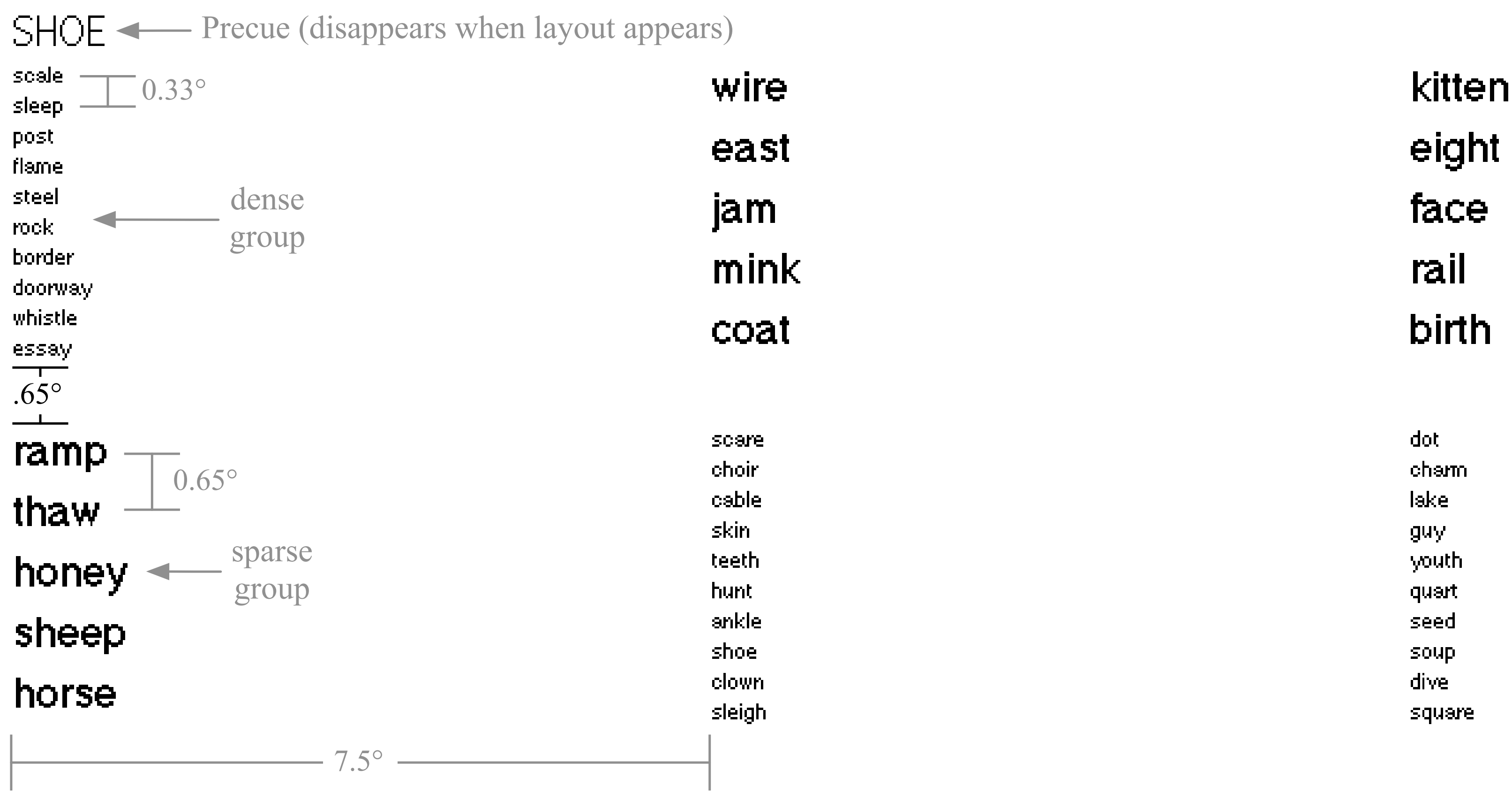


Figure 6. A mixed-density layout. All angle measurements are in degrees of visual angle.

surrounded by white space that was equal to or greater than the distance between the centers of the words in the group. The groups were arranged in three columns and two rows. Columns were 7.5 degrees of visual angle from left edge to left edge. Rows were separated by 0.65 degrees of visual angle.

There were two types of groups, each with a different local density. Sparse groups contained five words of 18 point Helvetica font with 0.65 degrees of vertical angle between the centers of adjacent words (0.45° for word height, and 0.2° for blank space). Dense groups contained 10 words of 9 point Helvetica font with 0.33 degrees of vertical angle between the centers of adjacent words (0.23° for word height, and 0.1° for blank space). Both types of groups subtended the same vertical visual angle.

There were three types of layouts: sparse, dense, and mixed-density. Figure 6 shows an example of a mixed-density layout. Figure 7 shows examples of sparse and dense layouts. Sparse layouts contained six sparse groups. Dense layouts contained six dense groups. Mixed-density layouts contained three sparse groups and three dense groups. The arrangement of the group densities in the mixed-density layouts was randomly determined for each trial. Sparse and dense layouts were identical to the mixed-density layout, with the exception of group densities.

One concern with the experimental design may be that font size and word spacing are conflated. In other words, how can the effects of density be evaluated when both factors are varied? This experiment was designed in part to determine the effect of combining

Figure 7. A sparse layout (top) and dense layout (bottom).

multiple local densities in a single layout. Local density was purposefully manipulated by covarying font size and word spacing. Text size often covaries with local density in real-world tasks, as is seen in the example from the Craigslist web page in Figure 3. Varying just text size or spacing may have removed the effect of visually distinct groups. The number of words per group was varied with local density to keep the height of groups similar so that the visual salience of dense and sparse groups were approximately equivalent.

The words used in each trial were randomly selected from a list of 765 nouns generated from the MRC Psycholinguistic Database (Wilson, 1988). No word appeared more than once per trial. The words from the database were constrained as follows: three to eight letters, two to four phonemes, above-average printed familiarity, and above-average imagability. Five names of colors and thirteen emotionally charged words were removed. The words used are shown in Appendix A.

The target word was randomly chosen from the list of words used in each trial. The participant was precued with the target word before each layout appeared. The precue appeared at the same location every time, directly above the top left word in the layout, in uppercase, 14 point Geneva font – a different font than in the layouts to reduce the effects of visual priming.

3.1.1.4 Procedure

The procedure was as follows. Each participant signed an informed consent form, filled out a demographics questionnaire, received a vision exam, and received instructions for the experiment. The participants were told that they would complete many trials and that each would proceed as follows: The participant should study the precue; click on the precue to make it disappear and the layout appear; find the target word without moving the mouse cursor; and then move the cursor to click on the target.

The participants were financially motivated to not move the mouse cursor until they had found the target. This was done to separate the visual search time from the cursor pointing and clicking time. The point-completion deadline procedure (Hornof, 2001) enforced compliance. Participants practiced meeting the point completion deadline prior to the experiment.

At the start of each experiment, the eye tracker was calibrated to the user. The calibration procedure required the participant to fixate a series of nine points until the average error between the predicted point of gaze and the actual location of the points fell below an error threshold (approximately 0.6 degrees of visual angle). Accurate eye tracking calibration was maintained using the automated required fixation locations (RFL) procedure (Hornof & Halverson, 2002) discussed earlier. If the eye tracker did not detect the user's gaze on the precue before the precue was clicked and yet the participant nonetheless selected the correct target, the eye tracker was recalibrated before the following trial.

The trials were blocked by layout type. Each block contained 30 trials, preceded by five practice trials. The blocks were fully counterbalanced. Trials were marked as errors if whitespace or a word besides the target was clicked on, the point completion deadline was not met, or the eye tracker was out of calibration for the trial. Only correct trials are analyzed.

3.1.2 Results

Search time began when the participant clicked on the precue and ended when the participant started moving the mouse. Eye movement data included in the analyses started from the first fixation (that started) after the precue was clicked and ended with the first fixation (that ended) before the mouse started moving.

The mean search time and eye movement measures for each of the twenty-four participants were analyzed using repeated-measure ANOVAs. An alpha level of .05 was used for all statistical tests. The analyses focused on the experimental manipulation of layout density.

3.1.2.1 Error Rates

Three types of errors were recorded: (a) The participant did not click on the target; (b) the participant moved the mouse before the target was found (i.e. violation of the point completion deadline); (c) the eye tracker calibration was off. As shown in Figure 8, the participants' error rates were low, less than 5 percent, in all conditions. More errors occurred in the high density layouts. An increase in errors was expected in the dense layouts where all of the targets are small. Smaller targets were harder to click on, resulting in more occasions in which the participants would likely click on a nearby target or require more time to click (thus risking a violation of the point-completion deadline). The errors were still relatively low, indicating that the participants were motivated to perform accurately. Further, as will be seen in the next section, visual search took longer

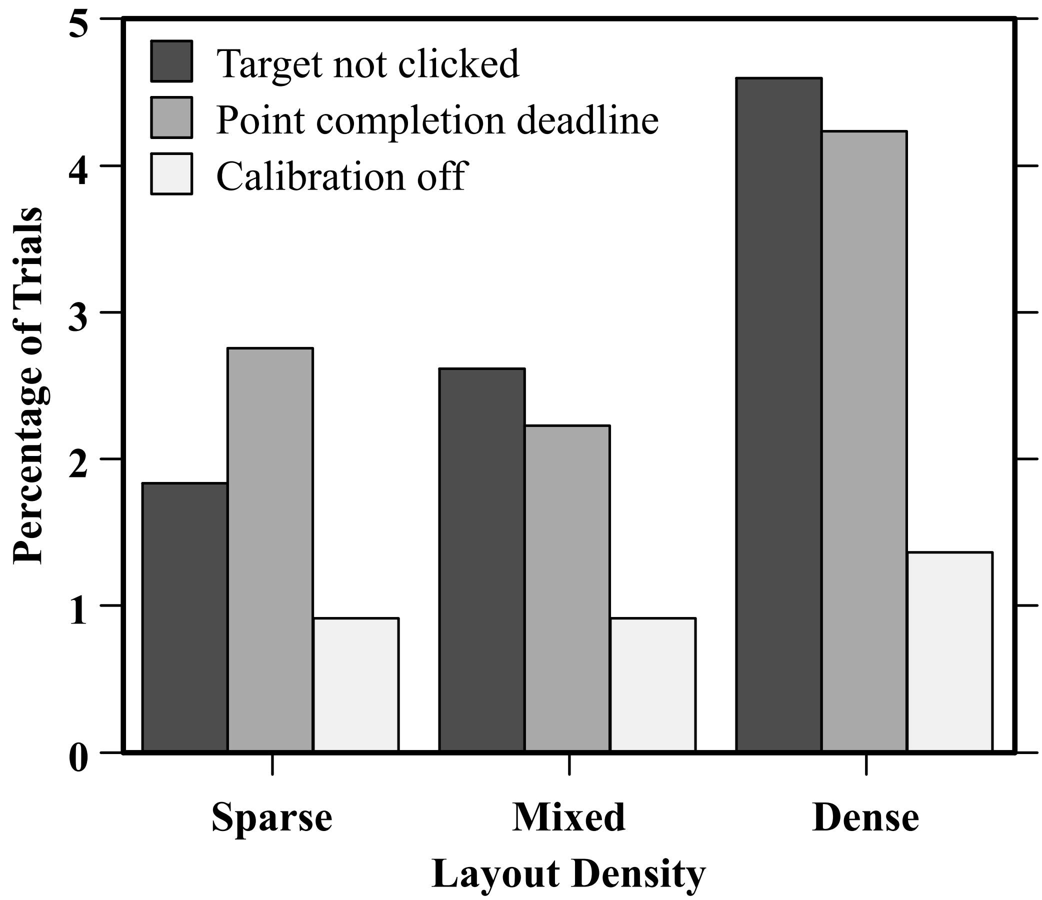


Figure 8. Percentage of trials in which human errors occurred in the mixed-density task, grouped by layout type and error type. Error rates were consistently low. More errors occurred in the dense layouts, which also required more visual search time, suggesting that participants were not trading speed for accuracy.

in the high density conditions. Lower error rates in conditions with faster search times suggest that the participants were not trading speed for accuracy.

Figures 8 also shows that the eye tracker calibration had to be corrected infrequently. The eye tracker had to be recalibrated a total of 25 times, which is less than one calibration per block. This indicates that the participants gaze was reliably detected where it was expected and hence the eye tracking data is reliable.

3.1.2.2 Search Time

As shown in Table 1, participants searched layouts that had fewer dense groups faster than layouts that had more dense groups, $F(2, 46) = 127.80$, $p < .01$. Since dense groups contained more words, the following analyses were conducted after normalizing for the number of words per layout. This was accomplished by dividing the search time per trial by half of the number of words in the layout. This normalization assumes that participants searched half of the words on average. Participants spent less time per word in layouts with fewer dense groups, $F(2, 46) = 13.94$, $p < .01$. More specifically, the participants spent less time per word in sparse layouts compared to mixed layouts, *HSD* = 23.68, $p < .01$; and less time per word in sparse layouts compared to dense layouts, *HSD* = 27.91, $p < .01$; but there was no meaningful difference between the time spent searching per word in the mixed and dense layouts, *HSD* = 4.24, $p = .46$.

The search time was also analyzed as a function of *layout uniformity* (single density vs. mixed density) and target group density. Figure 9 shows the results. Locating a target in a dense group took longer than in a sparse group, $F(1, 23) = 83.87$, $p < .01$. The mean search time for all-sparse and all-dense was no different than the mean search time for mixed-density layouts, $F(1, 23) = 1.03$, $p = .32$. However, there was an interaction between layout uniformity and target group density, $F(1, 23) = 16.87$, $p < .01$. In other words, when the target was in a sparse group, participants found the target faster in all-sparse layouts than in mixed layouts; when the target was in a dense group, participants found the target faster in mixed-density layouts than in all-dense layouts. Furthermore, in

Table 1. Search time per trial and per word, fixations per word, and fixation duration for sparse, mixed-density, and dense layouts. The “per word” measures are normalized based on the number of words in the layout.

	Search Time per Trial (ms)		Search Time per Word (ms)		Fixations per Word		Fixation Duration (ms)	
Layout	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sparse	3125	665	108	49	0.7	0.2	250	22
Mixed	5753	1493	253	62	0.7	0.1	307	49
Dense	7925	1891	265	55	0.6	0.1	370	68

n=24

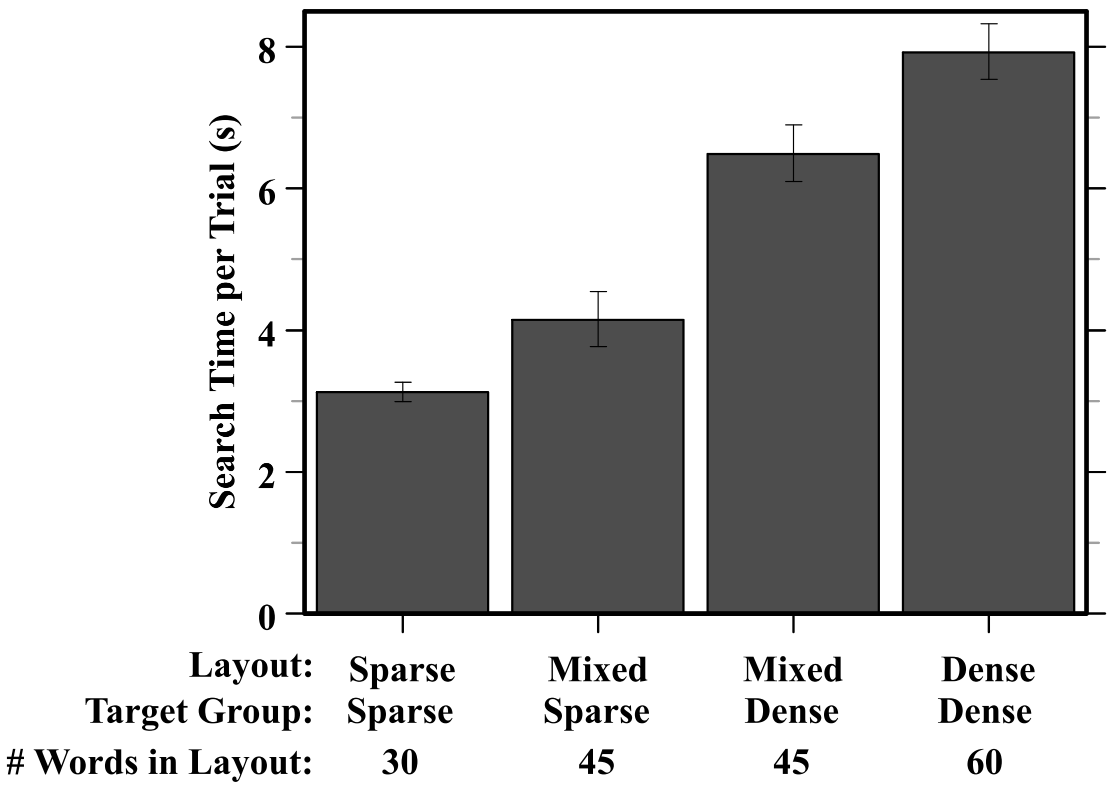


Figure 9. Search time by layout type and the density of the group in which the target appeared. Error bars indicate ±1 standard error of participants means.

mixed density layouts, participants found the target faster when it was in a sparse group, $F(1, 23) = 30.36$, $p < .01$.

3.1.2.3 Eye Movements

As shown in Table 1, participants made slightly fewer fixations per word in layouts with more dense groups, $F(2, 46) = 3.25$, $p = .05$. The participants used fewer fixations per word in the dense layouts than in the mixed layouts, $F(1, 23) = 8.42$, $p = .01$.

The fixation durations were much longer in layouts with more dense groups, $F(2, 46) = 61.82$, $p < .01$. The participants made longer fixations in the dense layouts than in the mixed layouts, $F(1, 23) = 36.01$, $p < .01$, and longer fixations in mixed layouts than in the sparse layouts, $F(1, 23) = 38.11$, $p < .01$.

As shown in Figure 10, participants tended to visit sparse groups before dense groups, $\chi^2(5, N = 24) = 500.04$, $p < .01$. A group was "visited" if one or more contiguous fixations fell within 1 degree of visual angle of the group (group revisits were not included). Differences in the number of group visits in the mixed density layouts were tested by comparing the percentage of visits to sparse or dense groups for the first through sixth group visit.

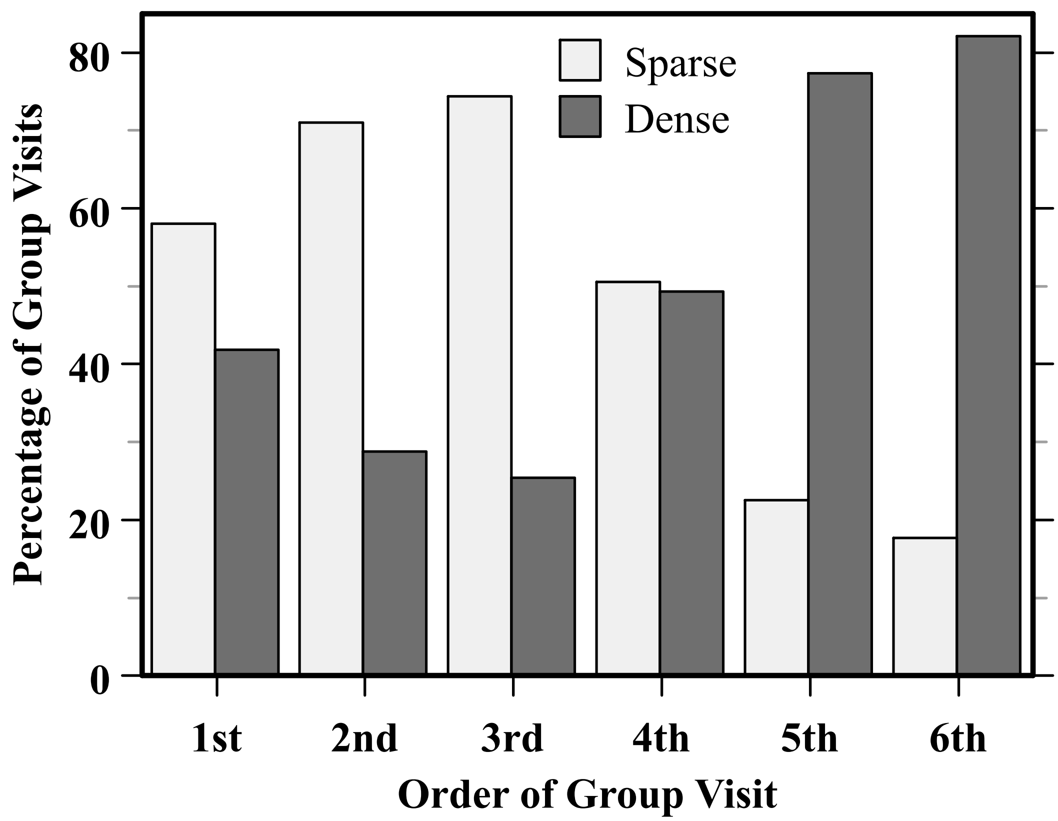


Figure 10. The percentage of group visits in the mixed density layouts that were to sparse or dense groups, as a function of the order in which the group was visited. Sparse groups were more likely to be visited first.

Participants revisited more groups per trial in layouts with more dense groups, $F(2, 46) = 10.50$, $p < .01$. Fewer revisits were made in sparse layouts than mixed layouts, $F(1, 23) = 12.82$, $p < .01$. Although there were fewer revisits per trial in mixed layouts than in dense layouts, the difference was not significant, $F(1, 23) = 2.31$, $p = .14$.

Figure 11 shows the number of fixations per group as a function of *density uniformity* (all groups of the same density or not), density of the group visited, and the order of group visit. There were no meaningful difference in the number of fixations per group between the uniform-density layouts and mixed-density layouts, $F(1, 9) = 2.69$, $p = .14$.

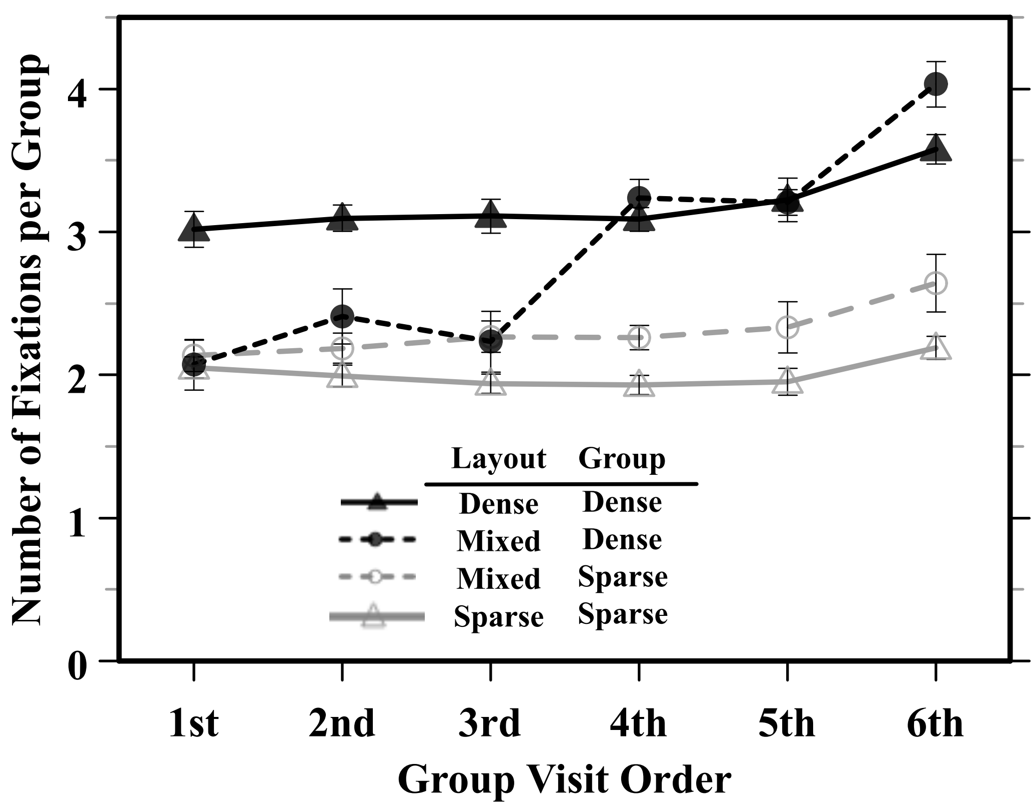


Figure 11. The mean number of fixations per group as a function of density uniformity, the density of the group visited, and order of the group visit. Error bars indicate ±1 standard error of participant means.

Participants used more fixations per group in dense groups than in sparse groups, $F(1, 9) = 112.30$, $p < .01$. Participants used more fixations per group as search progressed, $F(5, 45) = 8.14$, $p < .01$. All interactions were also significant. All of the interactions can be summarized by the three-way interaction of uniformity, target group density, and the order of group visit, $F(5, 45) = 4.52$, $p < .01$. The number of fixations per group increased as search progressed, but much more so for the number of fixations in dense groups in the mixed-density layouts. This interaction is illustrated in Figure 11 by the steeper slope of the black, dashed line.

Figure 12 shows the fixation durations as function of density uniformity, target group density, and the order of group visit. Fixation durations were longer in dense groups than in sparse groups, $F(1, 9) = 139.36$, $p < .01$. Fixation durations tended to be longer for groups visited later than for groups visited earlier, $F(5, 45) = 4.89$, $p < .01$. However, none of the interactions were statistically significant, even though trends similar to those found in the fixations per group analysis can be seen in Figure 12.

3.1.3 Discussion

This study investigates the effects of layout density and variation in density. This experiment reveals active vision strategies people use to search layouts with different densities, as is common in computer interfaces. These results will be used in the following chapter to help guide the development of a computational model of active vision for HCI.

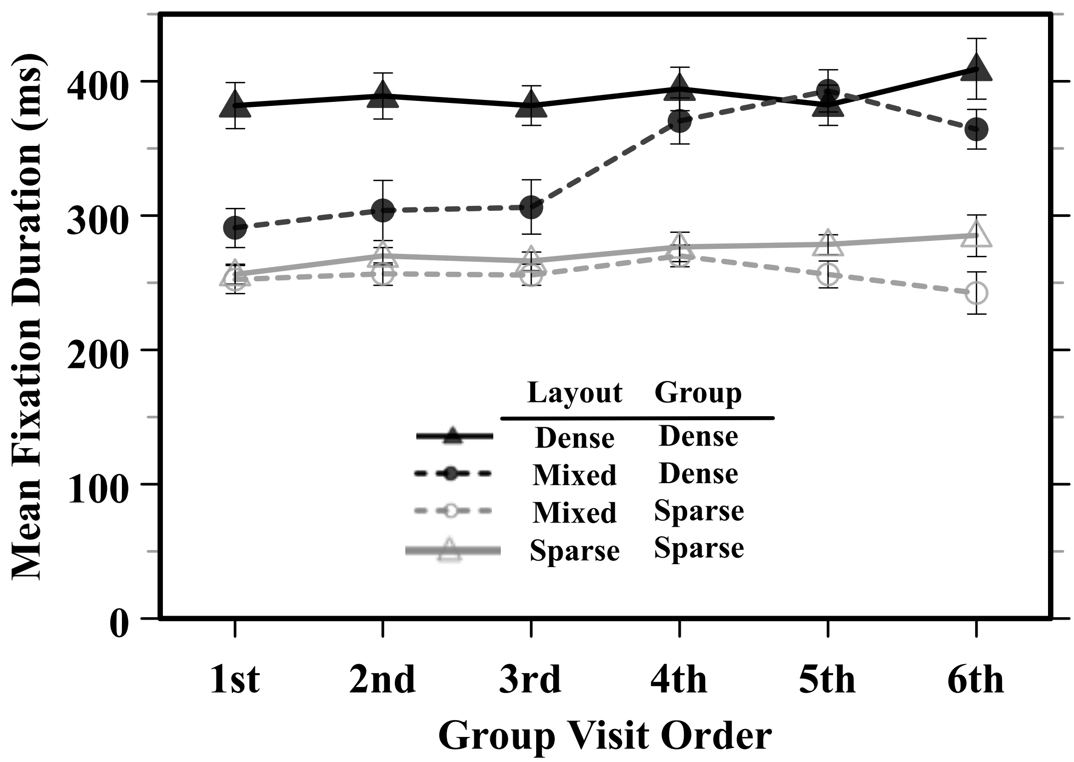


Figure 12. The mean fixation duration as a function of density uniformity, the density of the group visited, and order of the group visit. Error bars indicate ±1 standard error.

The most interesting results were found in the mixed-density condition. It was shown that the targets in sparse groups were found faster than the targets in dense groups, at least in part due to sparse groups being searched earlier.

The results suggest that people tend to search sparse groups first and faster (even when search times are normalized for word count). The search time data reported here demonstrate that people spent less time per word searching sparse layouts. It appears that participants were able to adopt a more efficient eye movement strategy that used slightly more and shorter fixations.

The finding that sparse groups are searched faster is contrary to search time results from some previous research. This result is contrary to the search time results found by Bertera and Rayner (2000) and Ojanpää, et al. (2002) in which the search time per item decreased as the density increased. This discrepancy may relate to the way in which density is manipulated. In the previous studies, the spacing between items was varied, and in the current study, the size of items (i.e. font size) was varied. It may be that although various factors affect local density, they do not all affect visual search of those densities in the same way.

Targets in sparse groups were found faster in part because sparse groups were searched first. People preferred to search sparse groups first. This is not what we expected. We expected dense groups to be searched first, as dense groups contained more information. Targets were found faster in the sparse group of mixed-density layouts, as shown in Figure 9, and the eye movement data also show that the participants tended to look at the sparse groups first. As is seen in Figure 10, it was much more likely for participants to look at sparse groups than dense groups within the first four groups visited. Note that while the first group visited was often a dense group, this is because 89% of all initial fixations were to the top-left group in the layout, and this group was equally likely to be either sparse or dense.

The findings that sparse groups tend to be searched first and faster supports the design practice of using sparse groups of text to attract users' attention to more important information. Information essential for the primary goals should be placed in groups that are sparse relative the other information in the layout so that they are more likely to be found faster. It is shown that sparse groups of words are searched faster and, when presented with dense groups, sparse groups are searched earlier than dense groups. This lends support to the practice of following the typographic conventions of displaying important information such as headlines with larger, more visually salient text.

A good, comprehensive cognitive model must include strategies, and useful analyses will identify these strategies. The data from this experiment provide evidence of such strategies. There was evidence of a shift in participants' search strategy when searching the mixed density layouts. This shift can be seen in Figures 11 and 12 in the darker dashed line that jumps up between the third and fourth group visit. After half of the groups in the layout had been visited, the participants tended to use more and longer fixations to search the dense groups. This transition occurred right around the time that it became more likely for the participants to search the dense groups.

This observed strategy shift suggests that care should be taken when combining densities in a visual layout. If people regularly adopt strategies that are more optimal for sparse text and use these same strategies when searching dense text, this may increase the likelihood of information being missed in dense groups. In general, fewer and shorter fixations are less likely to find a target. The next study examines another common visual characteristic of text in visual interfaces – the color of the text.

3.2 Link Colors

The second experiment, on link colors, is designed to reveal people's visual search strategies and provide general guidance for building computational cognitive models of human-computer visual interaction, specifically for tasks involving colored text. This experiment explores how issues of color, discussed in Chapter II and researched by previous authors, affect active vision in HCI tasks. The experiment investigates how varying the number of items of the same color as the target affects visual search when there are items of a different color present or not. In other words, this study investigates the visual search of structured layouts where only a subset of text, based on the color of text, need be searched. The web design practice of trying to assist visual search by differentiating visited and unvisited links based on color is investigated by varying the ratio of blue and red words, and by including layouts in which red words were replaced by blank space. The target word is always blue. It is hypothesized that search time would increase with the number of blue words and search would be faster with the red words absent.

3.2.1 Method

3.2.1.1 Participants

Twenty-four people, 11 female and 13 male, ranging in age from 19 to 55 years of age (mean = 25.1, SD = 7.9) from the University of Oregon and surrounding communities participated in the experiment. Twenty-two of these participants also took part in the mixed density experiment described above. The same screening criteria were applied.

Participants were monetarily motivated to participate in this study, and to perform with high speed and accuracy. The participants were instructed that they should complete each trial quickly while making few errors. Participants were paid \$10, plus a bonus that ranged from \$2.53 to \$9.20 based on their performance. Speed and accuracy were motivated as follows: The participant was initially given 4¢ to 7¢ for each trial, depending on the complexity of the screen. They lost 1¢ for each second after the trial started until they clicked on the target. If the participant clicked on something besides the target or if they moved the mouse before they found the target, the participant lost all of the money for that trial plus 5¢. The rewards and penalties were explained to the participants to motivate speed and accuracy.

3.2.1.2 Apparatus

The same stimuli presentation and eye tracking computers were used as described above for the mixed density experiment. The visual stimuli were presented with custom software designed using the C++ programming language, Metrowerk's PowerPlant framework, and Metrowerk's Codewarrior development environment. The software was developed by the authors as part of this dissertation work.

3.2.1.3 Stimuli

Figure 13 shows a sample layout. All layouts utilized the same 30 locations for the text stimuli. These locations were divided into six groups of left-justified, vertically-listed words words. The groups surrounded by white space that was equal to or greater than the distance between the centers of the words in the group. The groups were arranged in three

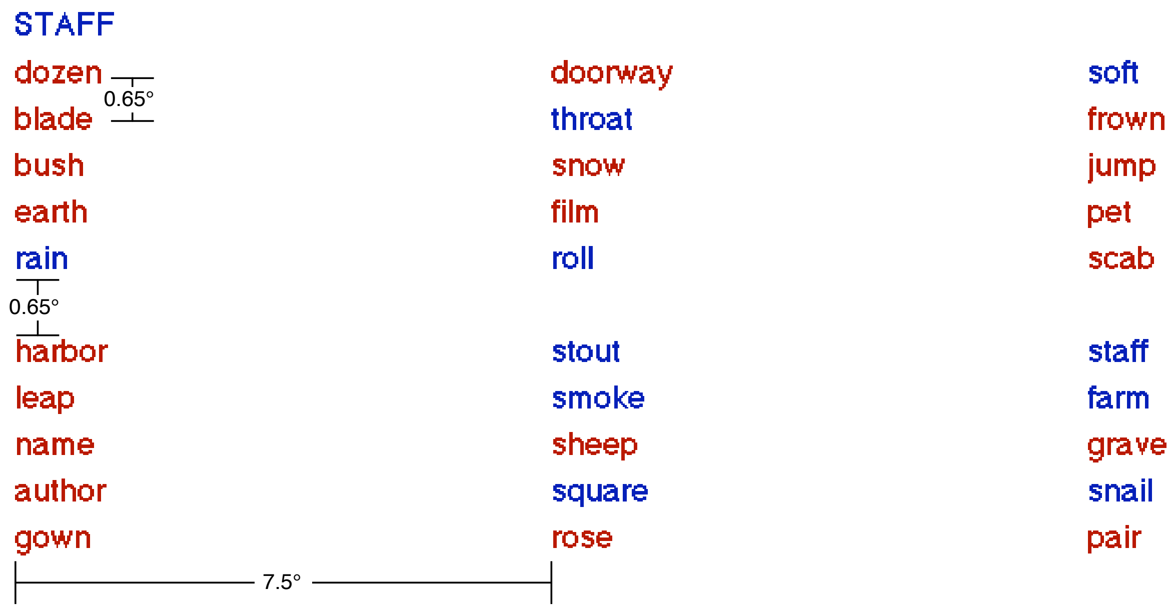


Figure 13. A sample layout from the text color experiment with ten target-colored (blue) words and twenty non-target-colored (red) words. All angle measurements are in degrees of visual angle.

columns and two rows. Columns were 7.5 degrees of visual angle from left edge to left edge. Rows were separated by 0.65 degrees of visual angle. Each group contained five locations for the text stimuli with 0.65 degrees of vertical angle between the centers of adjacent locations (0.45° for word height, and 0.2° for blank space).

The text in the layouts was always the same size and font, but the color of the text could vary. The text was 18 point Helvetica font. The text could either be blue (RGB values of 0, 0, 67) or red (RGB values of 67, 0, 0). The words appeared on a white background (RGB values of 0, 0, 0).

There were seven types of layouts: One layout contained thirty blue words, two layouts contained twenty blue words, two layouts contained ten blue words, and two layouts contained one blue word. In each pair of layouts with the same number of blue words, non-blue positions were either filled with red words as shown in Figure 13, or left blank as shown in Figure 14. The mixed-color layouts were always filled for a total of 30 words. Because the target was always blue, red words could be distinguished as non-targets based on their color alone.

The words used in each trial were randomly selected from a list of 765 nouns generated from the MRC Psycholinguistic Database (Wilson, 1988). No word appeared more than once per trial. The words from the database were constrained as follows: three to eight letters, two to four phonemes, above-average printed familiarity, and above-average imagability. Five names of colors and thirteen emotionally charged words were removed. The words used are shown in Appendix A.

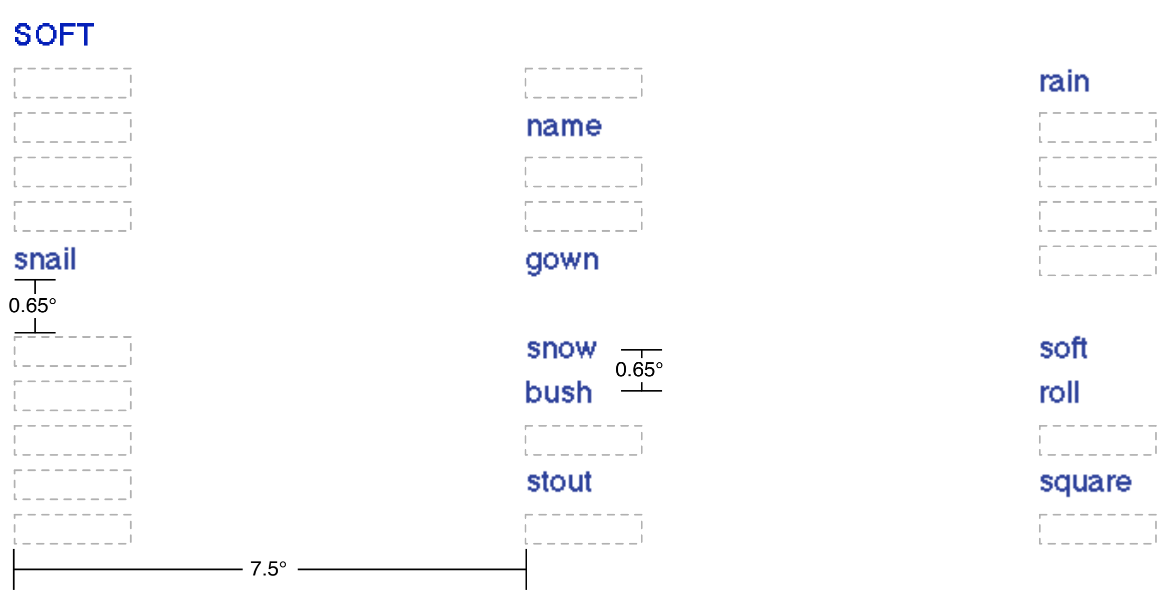


Figure 14. This layout is equivalent to that shown in Figure 13, except that the red words have been replaced with blanks.

The target word was randomly chosen from the list of words used in each trial. The participant was precued with the target word before each layout appeared. The precue appeared at the same location every time, directly above the top left word in the layout, in black (RGB values of 0, 0, 0), all upper case, 14 point Geneva font – a different font than in the layouts to reduce the effects of visual priming.

3.2.1.4 Procedure

The procedure for this experiment was identical to the Local Density experiment procedure described in section 3.1.1.4. In short, the precue appeared, the participant clicked on the precue, searched the layout and clicked on the precue. For participants that took part in this and the mixed density experiment during the same session, the order of presentation for the experiments were counterbalanced.

3.2.2 Results

Search time began when the participant clicked on the precue and ended when the participant started moving the mouse. Eye movements were recorded from the time participants clicked on the precue to when they clicked on the target. The mean search time and eye movement measures for each of the twenty-four participants were analyzed using repeated-measure ANOVAs. An alpha level of .05 was used for all statistical tests.

3.2.2.1 Error Rates

Three types of errors were recorded: (a) The participant did not click on the target. (b) The participant moved the mouse before the target was found (i.e. violation of the point completion deadline). (c) The eye tracker calibration was off. As shown in Figures 15 and

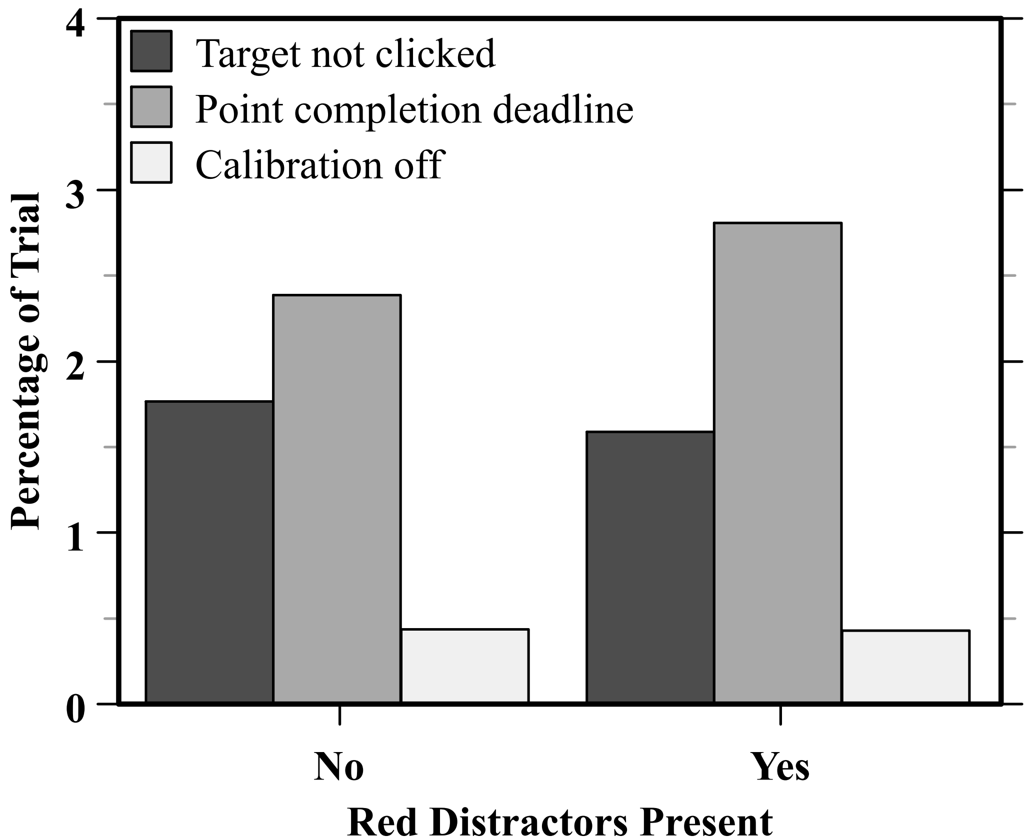


Figure 15. The percentage of errors in the Text Color experiment when the red distractors were present or not.

16, the participants' error rates were low — less than 4 percent — in all conditions. The percentage of point-completion errors tended to increase with the number of blue words in the layout. This may be because with more potential targets in the layouts, the participants were more likely to mistake another word for the target and start moving the cursor towards such a word. However, as will be seen in the next section, visual search time also increased as the number of blue words increases. Lower error rates in conditions with lower search times demonstrated that the participants were not trading speed for accuracy.

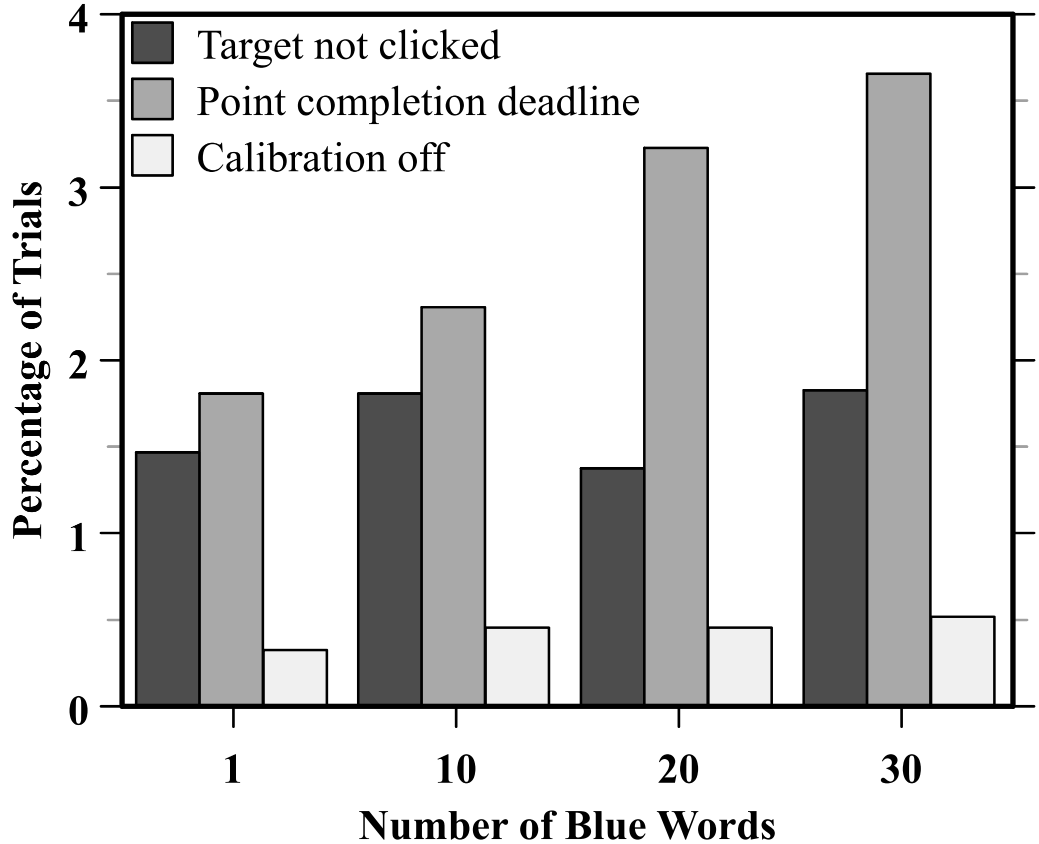


Figure 16. The percentage of errors in the Text Color experiment with 1, 10, 20, or 30 blue words present.

Figures 15 and 16 also show that the eye tracker calibration had to be corrected occasionally, in roughly 0.3% of the trials, but this is not inordinately high. This auto-recalibration was triggered a total of 25 times, on average less than one time per block. This suggests that the eye tracking data is reliable; the participants gaze was reliably detected where it was expected.

3.2.2.2 Search Time

As can be seen in Figure 17 and Table 2, participants found the target faster in layouts with fewer blue words, $F(3, 21) = 378.80$, $p < .01$, and when red distractors were absent,

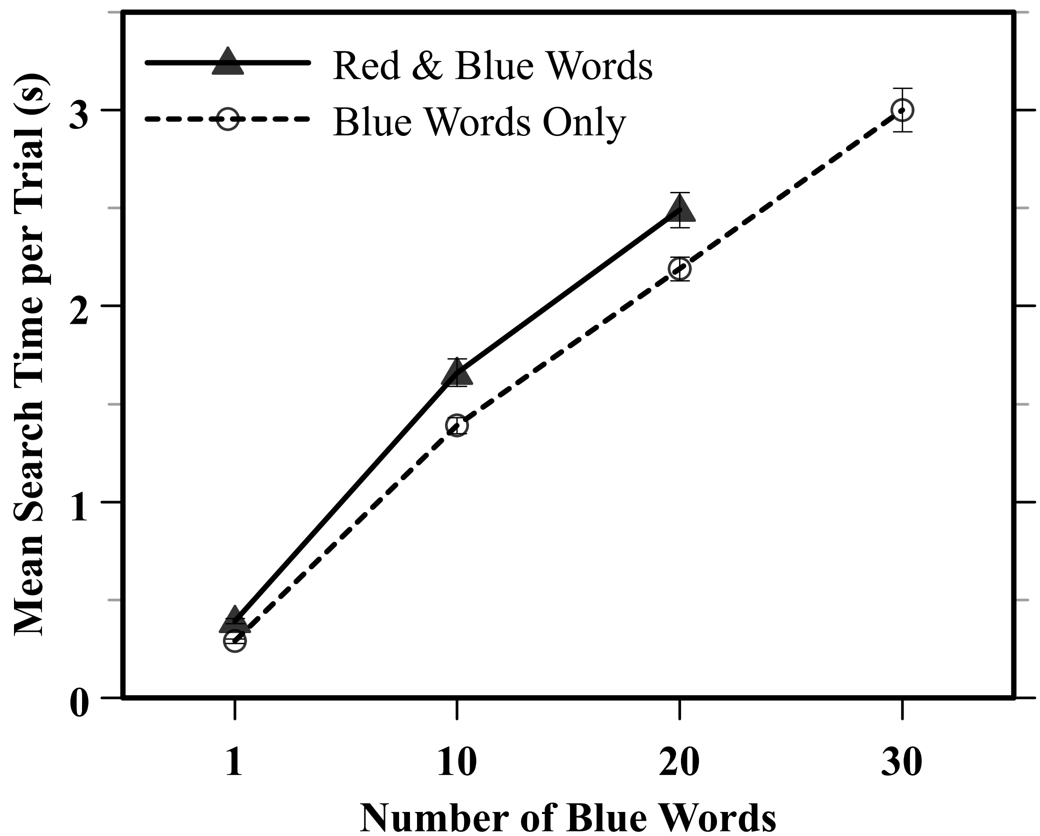


Figure 17. The mean search time per trial in the Text Color experiment as a function of the number of blue words and the presence of red words in the layout. Error bars indicate ±1 standard error.

$F(1, 23) = 32.01$, $p < .01$. There is an interaction between the number of blue distractors and the presence of red distractors, $F(3, 21) = 23.06$, $p < .01$. In other words, the presence of red words slowed search more when more were present.

Since layouts without red distractors contained more words, analyses were also performed after normalizing for the number of blue words per layout. This was accomplished by dividing the search time per trial by half of the number of words in the layout. This normalization assumes that participants searched half of the words on average.

Table 2. Mean search time per trial and per blue word, mean fixations per trial and per blue word, and mean fixation duration for all conditions in the Link Color experiment.

Number of Blue Words	Has Red Distractors?	Search Time per Trial (ms)		Search Time per Blue Word (ms)		Fixations per Trial		Fixations per Blue Word		Fixation Duration (ms)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	No	289	46	289	46	1.7	0.2	1.7	0.2	132	42
1	Yes	393	69	393	69	2.2	0.3	2.2	0.3	151	36
10	No	1387	187	139	19	5.4	0.6	0.5	0.1	200	20
10	Yes	1657	329	166	33	6.1	0.9	0.6	0.1	211	21
20	No	2191	273	110	14	7.7	1.1	0.4	0.1	238	26
20	Yes	2488	457	124	23	8.4	1.5	0.4	0.1	238	26
30	No	3005	544	100	18	9.9	2.0	0.3	0.1	255	23

n=24

As shown in Table 2, participants spent less time searching *as a function of the number of blue words present* in layouts with more blue words, $F(3, 21) = 187.48$, $p < .01$, and when the red distractors were absent, $F(1, 23) = 68.60$, $p < .01$. Moreover, there is an interaction between the number of blue distractors and the presence of red distractors, $F(3, 21) = 23.06$, $p < .001$. Again, it is seen that the presence of red words slowed the search more when there were more red words present, but this time after normalizing for the number of blue words.

3.2.2.3 Eye Movements

As shown in Figure 18 and Table 2, all the main effects and interactions found with search time also appear in the number of fixation data. Participants made more fixations in layouts with more blue words, $F(3, 21) = 379$, $p < .01$. Participants also made more fixations when red words were present, $F(1, 23) = 33$, $p < .01$. Further, the effect of red words was greater when more red words were present, $F(3, 21) = 20$, $p < .01$.

Fixation durations were also analyzed. It was found that, if we account for things like the pop-out effect by removing the layouts in which only one blue word appeared, fixation durations are equivalent across all layouts (all $p > .05$).

Saccade distances and destinations were analyzed. All saccades were more likely to land on a blue word than a red word. Short saccades were more likely than long saccades to land on a blue word rather than a red word. The analysis was done for mixed color layouts as follows: *Short* saccades were defined as those under 7.5 degrees of visual

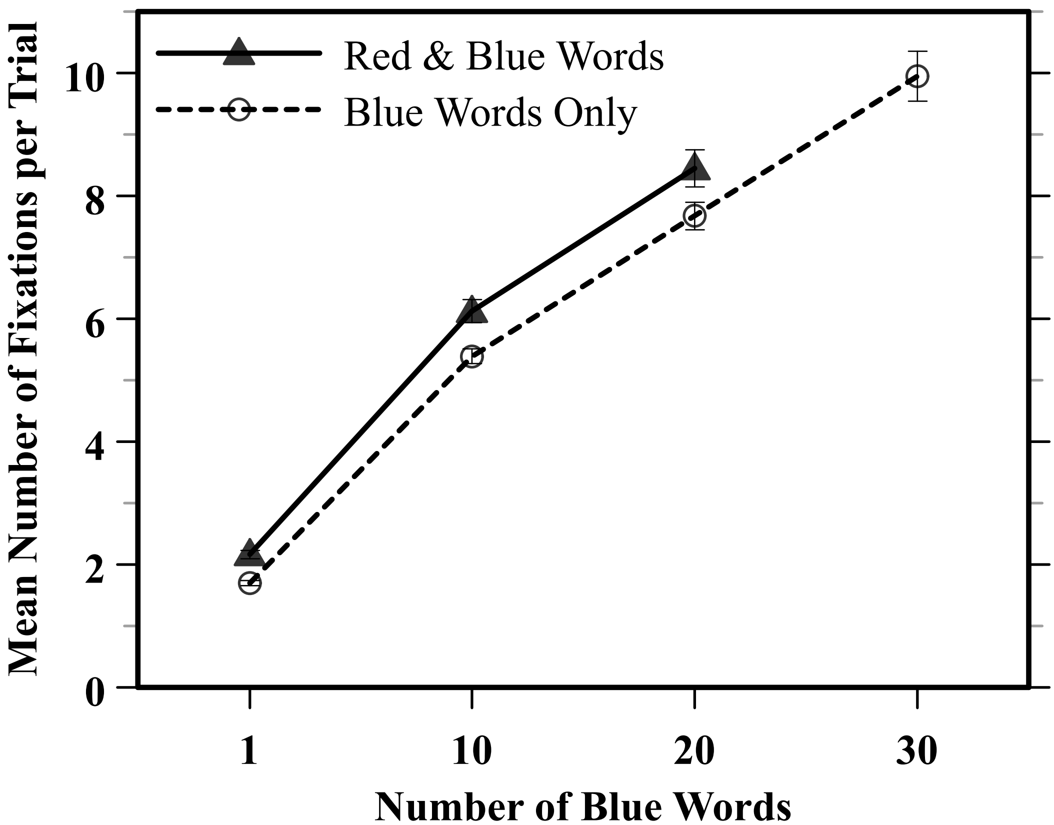


Figure 18. The mean number of fixations per trial in the Text Color experiment as a function of the number of blue words and the presence of red (or non-target-colored) distractors. Error bars indicate ±1 standard error.

angle (which is the distance between two columns of words). All other saccades were classified as *long*. Most saccades were short, $F(1, 23) = 958$, $p < .01$. Most saccades were to blue words, $F(1, 23) = 538$, $p < .01$. Additionally, an interaction of word color and saccade distance shows that the participants were more likely to saccade to a blue word when the saccade was short than when the saccade was long, $F(1, 23) = 359$, $p < .01$. Nonetheless, most long saccades were to blue words, $F(1, 23) = 14$, $p < .01$.

3.2.3 Discussion

This study investigates the degree to which color can be used to guide users' visual search strategies. Participants were presented with visual layouts with a varying number of blue words that they needed to search to find the target. Sometimes the remainder of the layout was filled with red words and sometimes those locations were empty. If people can focus on just the relevant stimuli, based on color, then we would expect equivalent behavior with the red distractors or identically-placed blank spaces.

These findings, for the most part, agree with previous work that suggests that people can ignore task irrelevant stimuli based on color (Poisson & Wilkinson, 1992; Shen, Reingold & Pomplun, 2003; Zohary & Hochstein, 1989). Additionally, this work extends the previous research by using more complex stimuli. Previous research used colored shapes and the like, whereas the current research used colored text.

However, it was also found that the task-irrelevant stimuli (the red words) slowed visual search, even though irrelevant stimuli can in theory be pre-attentively ignored. While not a large effect, the red words did slow visual search. One explanation for more fixations when red words are present is that, just by their presence, the red words provided more visual objects that could be used as the destination of saccades. The eye movement data support this possible explanation in that the likelihood of fixating a blue word rather than a red word decreased for longer saccades. However, the eye movement data also demonstrate that distinct link colors are very useful in guiding a search. This is true because they assist the programming of eye movements to relevant, unvisited links even when they are greater than 7.5 degrees of visual angle away from the current fixation, despite evidence that color (hue) perception is degraded at and beyond this angle.

The results of this experiment are relevant to design guidelines for differentiating the color of web page links based on the whether the linked pages have been visited (Nielsen, 2007; U.S. Department of Health and Human Services, 2006). For tasks in which web users need only search for relevant links to pages that have not been visited, this study shows that the visual search can be made very efficient if visited links are clearly discernible based on color. As the data show, layouts with thirty blue links – akin to Web pages that do not differentiate unvisited and visited links – take longer to search.

However, the finding that the presence of red words slows search time suggests that the guideline to differentiate links by text color might be improved. One possible improvement that could be made is differentiating visited links by luminescence in addition to color. Basic research has shown that color is not easily discernible in the periphery, but luminescence is. As was seen in the eye movement data, link color is useful in the periphery of the display for this task, but a difference in luminescence may increase the benefit of differentiating visited links.

The results of this experiment extend an understanding of how color affects visual search strategies, and informs the development of predictive models of visual search. For example, it was found that people (a) tend to somewhat but not entirely ignore non-target-colored items, and (b) tend to fixate nearby items in part because the target-identifying features are easier to see. Further research may be needed to ensure that these findings hold when the text is surrounded by or embedded in additional content, and to validate the suggestion to differentiate web link further with differences in luminescence. Nonetheless, the strategy components demonstrated in this experiment are excellent candidates for inclusion in predictive visual layout analysis tools.

3.3. Semantically Grouped Displays

A third experiment was conducted to determine how people visually search semantically organized visual layouts. The experiment investigated effects of (a) positioning a target in semantically similar words, (b) giving the groups identifying labels, and (c) further subdividing the layouts into meta-groups using graphic design techniques. It was hypothesized that (a) the use of semantic grouping would be slightly less effective than group labels at facilitating efficient visual search, (b) group labels would speed peoples' visual search, as previous research shows, and (c) subdividing groups by common region would constrain people's visual search patterns.

3.3.1 Method

This experiment investigated the effect of semantic grouping, group labels, and common region on visual search of structured layouts of words. The remainder of this section discusses an experiment that investigates the effects of visual and semantic grouping on users' visual interaction. The task is described and the results of an experiment are discussed.

3.3.1.1 Participants

Eighteen people, nine male and nine female, ranging in age from 20 to 62 years of age (mean = 29.1; SD = 11.3) from the University of Oregon and surrounding community participated in the study. The participant screening criteria, including verification of the participants' vision, were identical to those used in the previous experiment.

Participants were financially motivated to participate in this study, and to perform with high speed and accuracy. The participants were instructed that they should complete each trial quickly while making few errors. Participants were paid $15, plus up to $10 based on their performance. Speed and accuracy were motivated as follows: The participant was initially given 8¢ to 10¢ for each trial, depending on the complexity of the screen. They lost 1¢ for each second after the trial started until they clicked on the target. If the participant clicked on something besides the target or if they moved the mouse before they found the target, the participant lost all of the money for that trial plus 5¢. The rewards and penalties were explained to the participants to motivate speed and accuracy.

3.3.1.2 Apparatus

Visual search stimuli were presented on an AG Neovo X-174 LCD display at a resolution of 1280 width by 1024 height at a distance of 61 cm, which resulted in 39 pixels per degree of visual angle. The experimental software ran on Dual 2GHz PowerMac G5 running OS X 10.4.7. The mouse was a wired Apple Mighty Mouse configured for single-button operation, and the mouse tracking speed was set the fourth highest in the mouse control pane. The visual stimuli were presented with custom software designed using the Objective-C++ programming language, Apple's Cocoa framework, and Apple's XCode development environment. The software was developed by the authors as part of this dissertation work.

Eye movements were recorded using a dual-camera LC Technologies Eyegaze System, a 120 Hz pupil-center and corneal-reflection eye tracker. Details of the eye tracker and eye movement analysis are given in section 3.1.1.3.

Participants completed the automated operation span (OSPAN) task (Unsworth, Heitz, Schrock & Engle, 2005), which was presented using E-Prime 1.1 running on a 3GHz Pentium 4 running Windows XP SP2.

3.3.1.3 Stimuli

Figures 19 and 20 show sample layouts. 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). Figure 19 shows a layout with semantically-cohesive groups, group

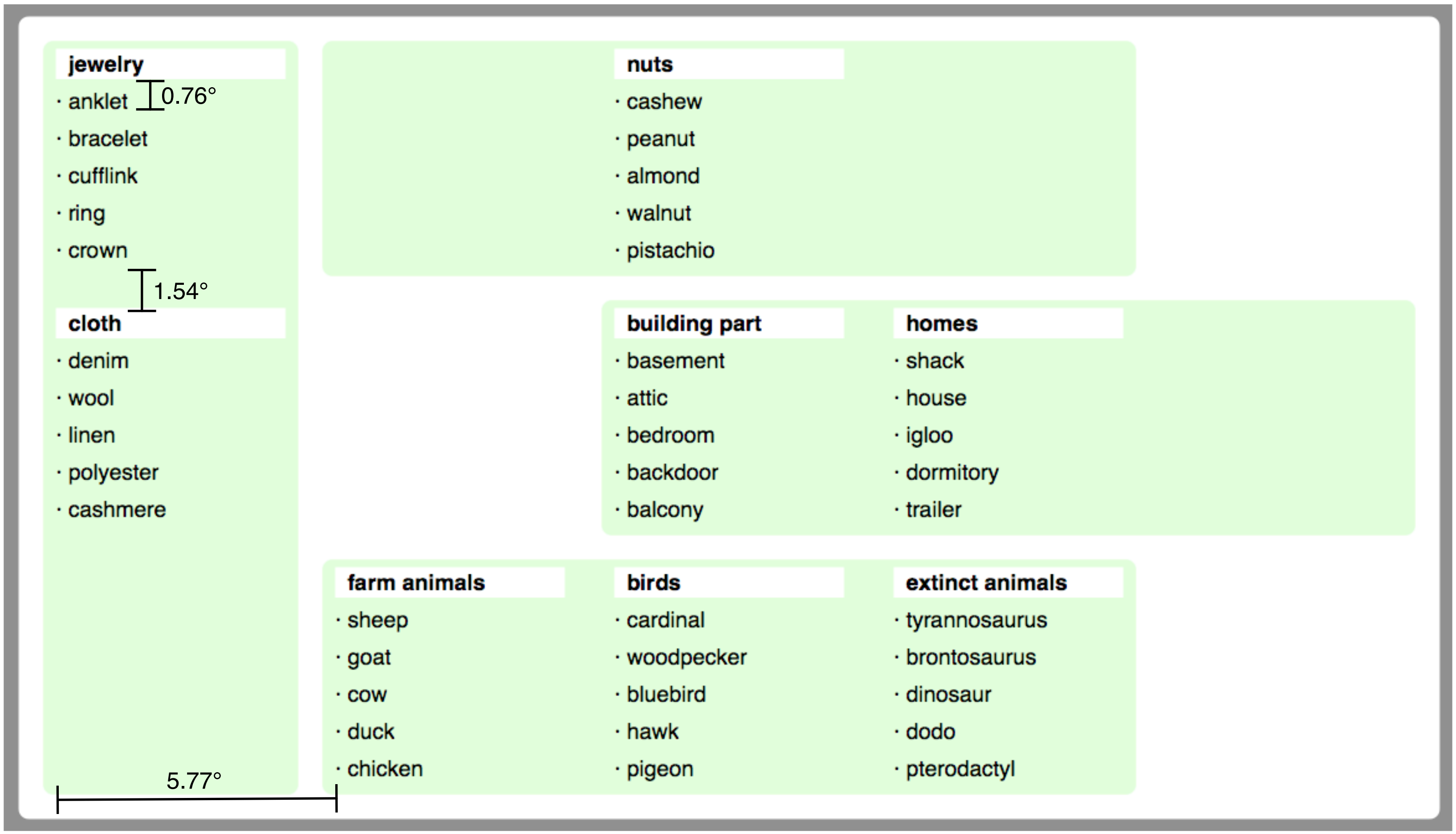


Figure 19. A layout from the semantic grouping experiment, with semantically-cohesive groups, group labels, and meta-groups.

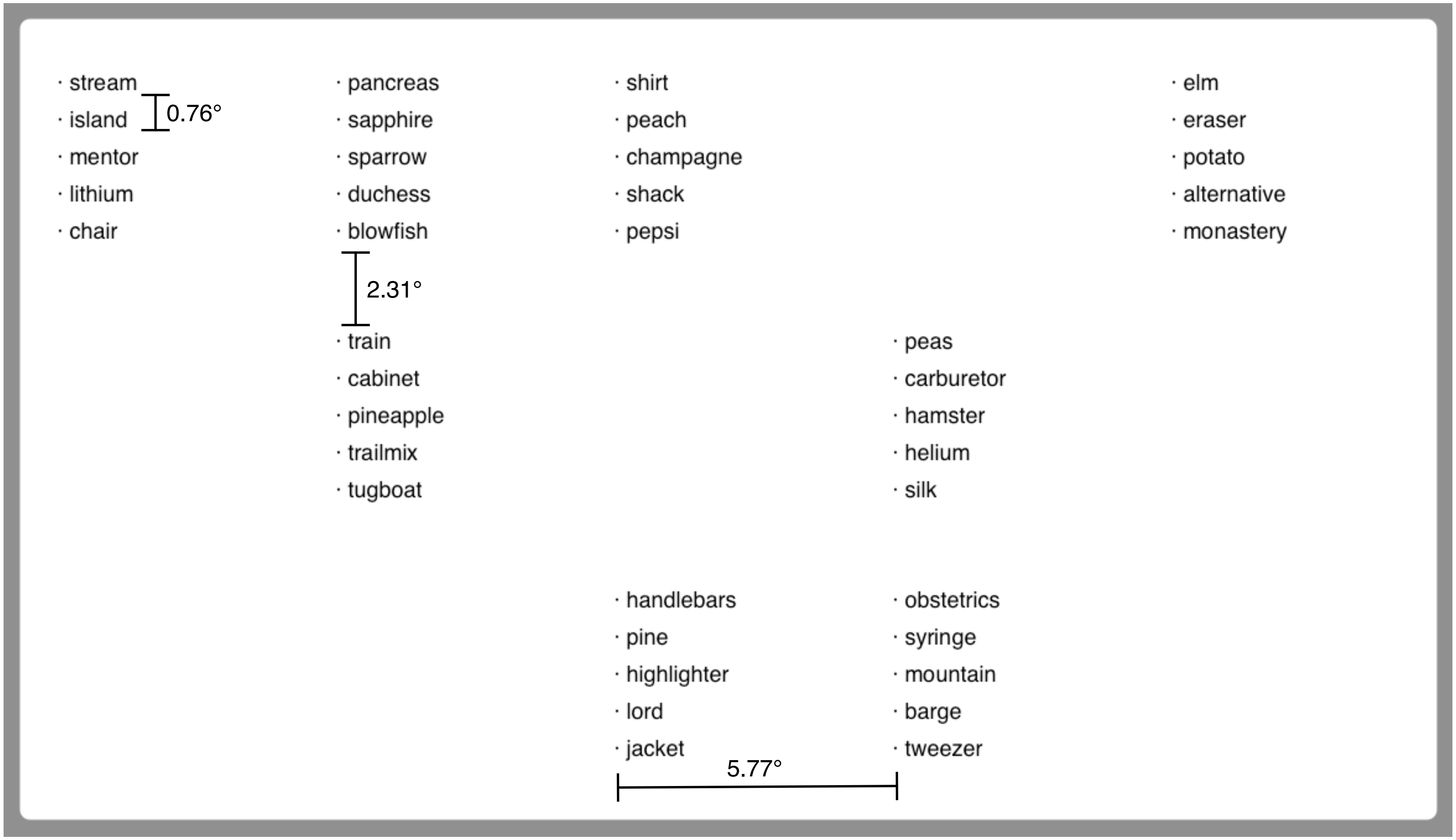


Figure 20. A layout with random-word groups, no group labels, and no meta-groups.

labels, and meta-groups. Figure 20 shows a layout with random-word groups, no group labels, and no meta-groups.

The only levels of variables not combined were randomly grouped words and labels. Labeled, random groups were not used because the nature of the task would change drastically, relative to the task using the other labeled layouts. The labels for the groups would be misleading.

The layouts appeared on a white background (RGB values of 255, 255, 255 and an alpha of 1.0). Layouts were 29° wide by 17° high, centered on the screen. The remainder of the screen was a gray background (RGB values of 128, 128, 128 and an alpha of 1.0). When meta-groups were used, the were indicated with a green region (RGB values of 222, 255, 214 and an alpha of 1.0). When group labels were used, they were placed on a white background that was 4.74° wide by 0.62° high and centered at the top of the group.

The groups were placed in a 3 x 5 grid. Groups are always vertically centered in each grid cell, which are 4.74° wide of visual angle by 4.44° high. Groups were separated by 0.51° of horizontal whitespace and 0.41° of vertical whitespace. This resulted in an inter-group vertical spacing of 1.54° between the baselines of adjacent words for labeled groups and 2.31° for unlabeled groups. The distance between the left edges of horizontally-adjacent groups was 5.77°. The inter-group horizontal spacing varied with the length of the words used in each group.

Layouts always contained eight groups. Each group contained five lowercase words in 18 point Helvetica, black (RGB values 0, 0, 0 and an alpha of 1.0) font with 0.76° of visual angle between the baselines of adjacent words. A bullet character appeared to the left of every non-label word. When labels appeared, they were in 18 point bold Helvetica and appeared above the top word in the group with 0.76 degrees between the baselines.

There were two types of groups: semantically cohesive or random. The words in the cohesive groups always came from the same category. The words in the random groups were pseudo-randomly selected from all category words. In some layouts, the groups were labeled. When the labels were present, the category labels for the words in that group were used. In some layouts, 1 to 4 groups were contained within a common region defined by color. These categories of these groups of words were semantically related (e.g. farm animals, wild animals, and birds).

The words used in each trial were selected from a hierarchical list of words based on categories used in a study of word category norms (Yoon et al., 2004). Sixty-two categories from Yoon, et al. were used. Words were not used that did not meet the following criteria: The entry must not contain non-alphabetic characters or spaces. The word must not be in all capitals. If the word is an initialization, it must be pronounceable (i.e. it must be an acronym). A word could only appear in one category. Variations of the same word could only appear once (e.g. fridge and refrigerator). It must be clear that the word is part of the category without additional context, like the category name.

Categories were not used if there were fewer than eight words in a category or if there was too much similarity with another category, as judged by the author and two colleagues. The remaining categories were formed into super-categories.

The categories were further grouped into 14 super-categories. The formation of super-categories was the result of consensus among the author and two colleagues. Each person divided the groups into super-categories and then discussed categorization until a consensus was reached.

3.3.1.4 Procedure

The procedure was as follows. Each participant signed an informed consent form, a demographics questionnaire was administered, the eye exams were conducted, and the participants were given instructions for the experiment. The participants were told that they would complete many trials and that each would proceeded as follows: The participant was shown and studied a precue of the target. The participant clicked on the precue to make the precue disappear and the layout appear. The participant visually searched for the target word without moving the mouse cursor. The participant moved the cursor to the target word and clicked on it.

The participants were financially motivated to not move the mouse cursor until they had found the target using the point-completion deadline, as discussed earlier.

At the start of each experiment, the eye tracker was calibrated to the user. The calibration procedure required the participant to fixate a series of nine points until the average error between the predicted point of gaze and the actual location of the points fell below an error threshold (approximately 0.6 degrees of visual angle). The automated RFL procedure was not used to maintain calibration accuracy as in the previous two experiments. An eye tracker was used in this experiment that was more accurate than the eye tracker used in the previous two experiments. Specifically, a 120 Hz two-camera upgrade to the LC Technologies system used previously. Preliminary trials indicated that recalibrations were no longer required over the period of time these experiment would take.

The trials were blocked by layout type. Each block contained 40 trials, preceded by five practice trials. The blocks were counterbalanced using the balanced Latin square technique. Trials were marked as errors if whitespace or a word besides the target was clicked on, or if the point completion deadline was not met. Only correct trials are analyzed.

The practice trials were identical to the experimental trials with two exceptions: (a) Only the unlabeled, randomly-grouped, meta-groups-absent condition was used, and (b) the words presented during the practice trials were different from those in the experimental trials. These words were the same as those used in the local density and link color experiments discussed in previous sections. The participants continued to practice until they were comfortable with the task. The number of practice trials varied from 16 to 170 ($M = 51.3$, $SD = 25.2$). Once the practice trials were over, the participants completed the experimental trials. Finally, the experimenter elicited comments from the participants about how they completed the tasks.

Following a successful trial, the precue of the next trial was placed at the target location of the previous trial. Following a trial in which an error occurred, the precue was placed at the center of the display. During the visual search portion of the trial, to prevent the mouse cursor from obscuring any layout item appearing at the same location as the precue, the cursor appeared only when the mouse was moving.

3.3.2 Results

Search time began when the participant clicked on the precue and ended when the participant started moving the mouse. Eye movement data included in the analyses started from the first fixation that began after the precue was clicked, and ended with the first fixation that stopped before the mouse started moving.

Search time and eye movement data were analyzed using mixed-model ANOVAs. The Kenward-Roger correction method was used. An alpha level of .05 was used for all statistical tests. Due to non-normal distributions of the data, a log transform was used on the search time, number of fixations per trial, and fixation duration analyses. All means shown for these data are adjusted means.

The analysis focused on the three experimental factors: the semantic cohesion of groups of words, the presence of group labels, and the use of meta-groups. Because these factors were not fully crossed, all conditions were treated as a single factor in the

ANOVA analysis. The effect of these factors and their interactions were then analyzed with contrast analyses. Besides these three manipulations, a number of additional fixed factors were included in the analysis. These variables were gender, age, OSPAN score, computer experience, the target word, the length of the target word, the target group label, the location of the target group, the location of the precue, the distance to the target, and carryover effects.

The mixed model for each analysis was extended with random variables. The random variables were introduced one at a time and were removed if the Bayesian Information Criterion (BIC) did not decrease by 20 or more. The following random factors were found to contribute substantially to the models: the participants' individual differences, the participants' age, block order, the location of the precue, and the group in which the target appeared.

The data from 11 trials were excluded due to a bug in the software that recorded the reaction time incorrectly for those 11 trials. All 11 trials were from the first three participants' data.

3.3.2.1 Error Rates

As shown in Figures 21, 22 and 23 the participants' error rates were fairly low, less than 7 percent, in all conditions. Slightly more errors occurred when the groups were semantically cohesive or when the groups were labeled. However, all error rates were within 2.4% of each other and there are no clear trends as functions of semantic grouping

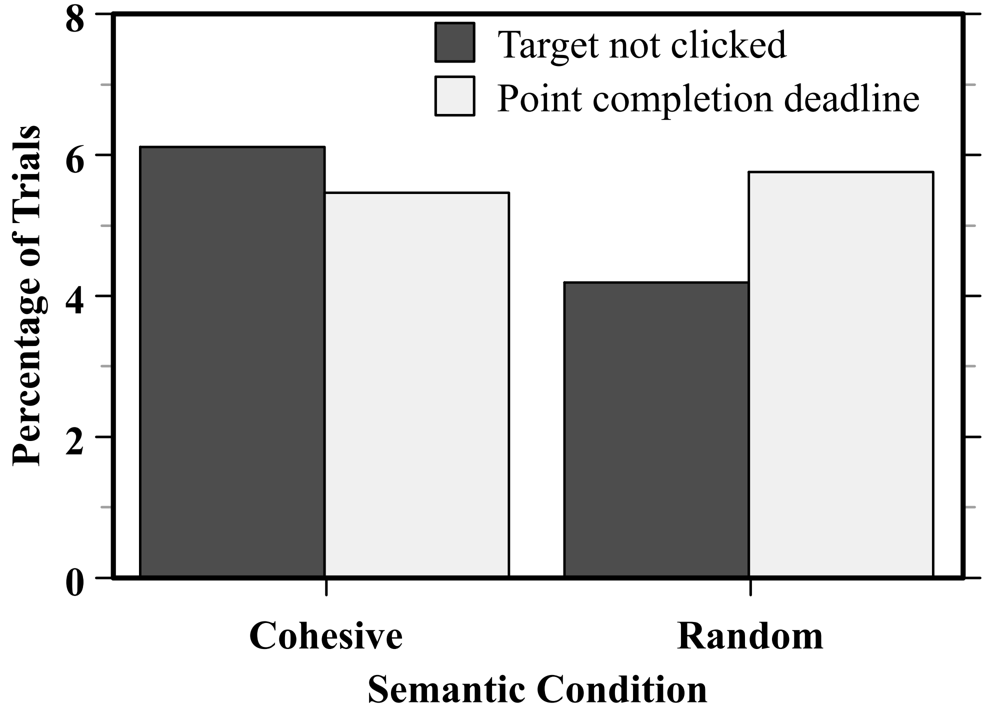


Figure 21. The percentage of errors in the Semantic Grouping experiment as a function of semantically cohesive or random word grouping, and error type.

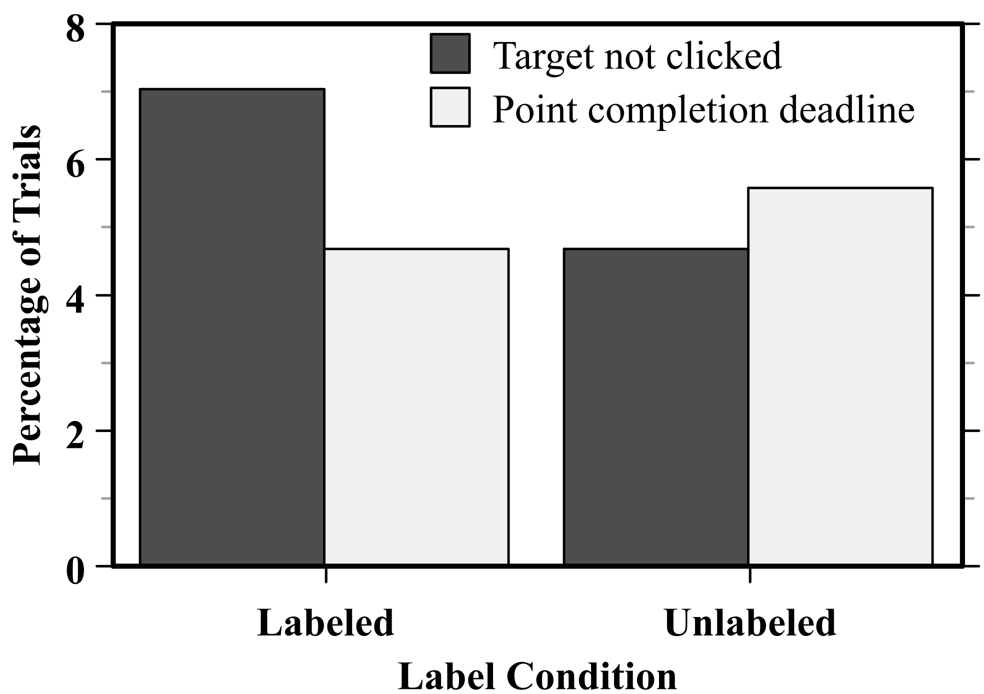
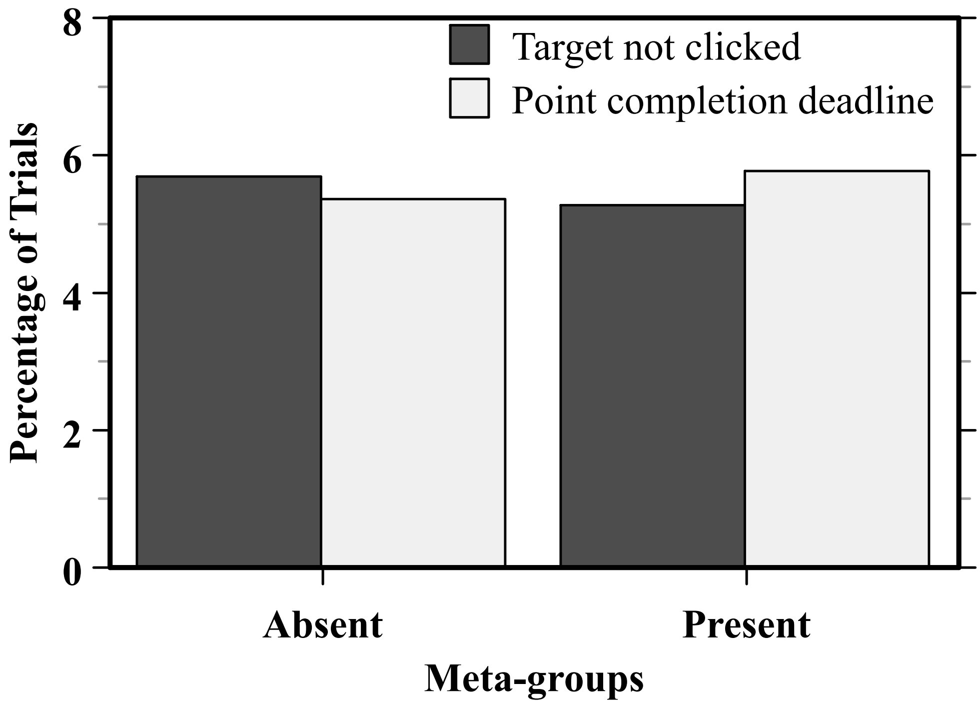


Figure 22. The percentage of errors in the Semantic Grouping experiment as a function of labeled versus unlabeled groups, and error type.



or

Figure 23. The percentage of errors in the Semantic Grouping experiment as a function of the absence or presence of meta-groups, and error type.

label presence. As will be seen in the next section, visual search was faster when the groups were semantically cohesive and when the groups were labeled. However, the consistently low error rates suggest that the participants were not trading speed for accuracy.

3.3.2.2 Search Time

The type of layout affected search time, $F(3, 1497) = 46.39$, $p < .01$, but only the effect of semantic grouping was significant, $t(1474) = 10.06$, $p < .01$. That is, participants tended to take less time to find the targets when the layouts were semantically organized. Figure 24 and Table 3 show the mean search times. The presence of meta-groups did not

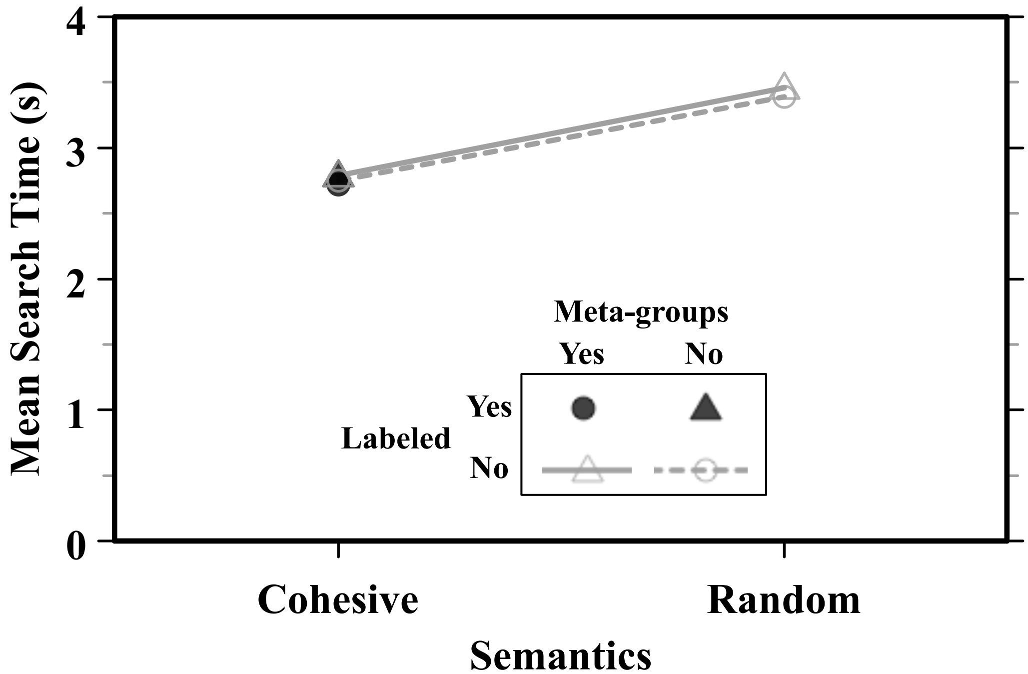


Figure 24. The mean search time per trial in the Semantic Grouping experiment as a function of semantic grouping, group labels, and meta-groups. The standard error is too small for the errors bars to be seen.

affect search time, $t(1503) = 0.98$, p = .33, nor did the presence of the group labels, $t(1503) = 0.23$, $p = .82$.

Participants found nearby targets faster than distant targets, $F(1, 4062) = 305.45$, $p < .01$. Besides the experimental factors, the only factor that affected search time was the distance between the where the participants started searching (at the precue) and where they finished searching (at the target).

No other factors affected participant search time. Search time did not vary with:

OSPAN scores, $F(1, 12.6) = 0.69$, $p = .42$
Gender, $F(1, 12.6) = 0.02$, $p = 0.89$
Age, $F(1, 12.7) = 0.17$, $p = 0.69$
Computer experience, $F(1, 12.6) = 0.14$, $p = .72$

Table 3. Search time, fixations, fixation duration, and saccade distance for all conditions in the Semantic Grouping experiment.

			Search Time (ms)*		Fixations*		Fixation Duration*		Saccade Distance	
Semantics	Labels	Meta-groups	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Cohesive	Present	Present	2724	26	10.2	1.8	217	2	4.26	0.05
		Absent	2724	26	10.4	1.9	218	2	4.24	0.05
	Absent	Present	2750	26	9.8	1.7	221	2	4.27	0.06
		Absent	2794	26	10.2	1.8	221	2	4.29	0.06
Random	Absent	Present	3385	29	12.3	2.0	223	2	4.11	0.05
		Absent	3464	29	12.5	2.0	227	2	4.16	0.05

n=18 *=adjusted means

Carryover from layouts with semantically cohesive group, labels,
and no meta-groups, $F(1, 1472) = 1.14$, $p = .29$
Carryover from layouts with semantically cohesive groups, labels,
and meta-groups, $F(1, 1491) = 0.40$, $p = .53$
Carryover from layouts with semantically cohesive groups, no labels,
and no meta-groups, $F(1, 1525) = 0.04$, $p = .84$
Carryover from layouts with semantically cohesive groups, no labels,
and meta-groups, $F(1, 1482) = 1.60$, $p = .21$
Carryover from layouts with random groups, no labels,
and meta-groups, $F(1, 1527) = 0.27$, $p = .60$

3.3.2.3 Eye Movements

The type of layout also affected the number of fixations per trial, $F(3, 1496) = 32.77$, $p < .01$. Only the effects of semantic grouping was significant. The participants required fewer fixations to find the target, $t(1474) = 9.06$, p < .01, when the groups were semantically cohesive. The presence of meta-groups did not affect the number of fixations, $t(1506) = 1.35$, $p = .18$, nor did the presence of the group labels, $t(1494) = 1.26$, $p = .21$.

The participants' fixation duration was also affected by the type of layout, $F(3, 621) = 5.76$, $p < .01$. Only the effects of semantic grouping was significant. Participants tended to make shorter fixations, $t(443) = 2.36$, $p = .02$, when the layouts were semantically organized. The presence of labels, $t(443) = 1.65$, $p = .10$, and meta-groups, $t(1485) = 0.15$, $p = .88$, had no effect on the fixation durations.

The mean distance of participants' saccade was also affected by the type of layout, $F(3, 1477) = 4.44$, $p < .01$. Only the effects of semantic grouping was significant. Participants tended to make longer fixations, $t(1457) = 3.30$, p < .01, when the layouts were semantically organized. The presence of labels, $t(1477) = 0.36$, $p = .72$, and meta-groups, $t(1485) = 0.40$, $p = .69$, had no effect on the saccade distances.

As shown in Figure 25, a qualitative analysis of the number of fixations per group of words suggests that people behaved differently when groups were labeled and when the groups were semantically cohesive[1]. The participants tended to use just one fixation per

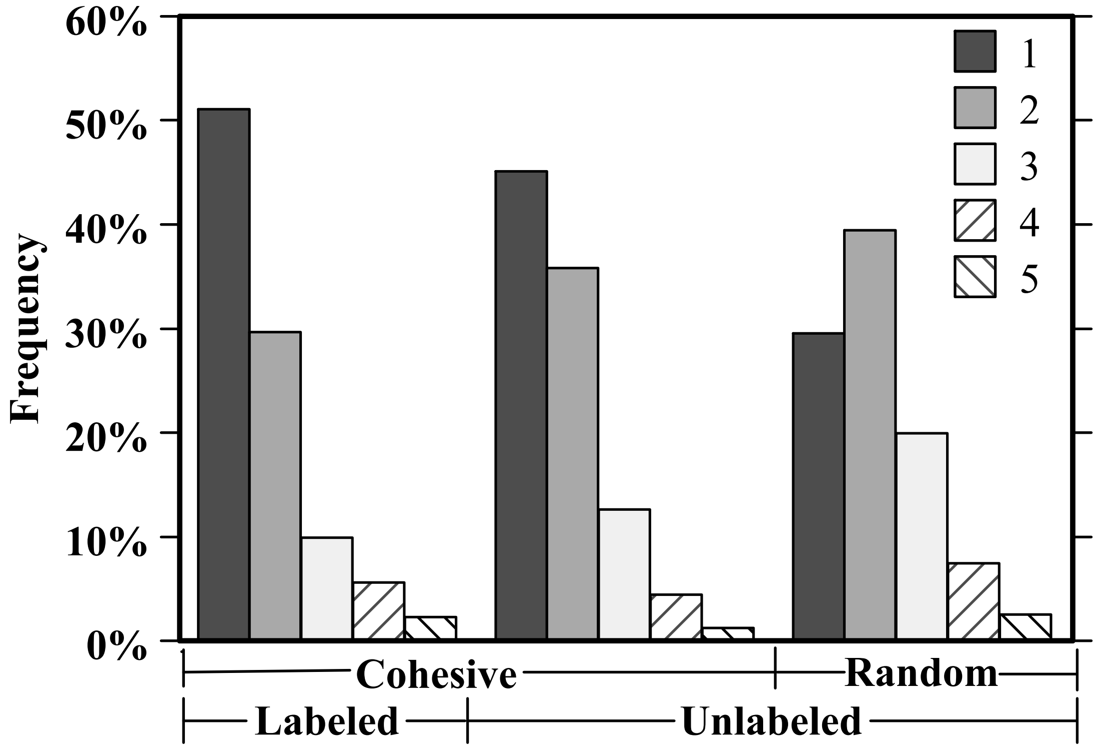


Figure 25. Frequency of the number of fixations per group visit as a function of group type (semantically cohesive and/or group label), and the number of fixations in a group. Notice how one fixation is often enough for a cohesive group, especially in the labeled layouts, whereas two fixations are typically needed for a random group. (Only the data from layouts without meta-groups are shown here; the trends with meta-groups present are similar.

[1] A quantitative analysis could not be performed. The data were "truncated counting data" with a distribution that could not be corrected to normal, which violated assumptions in all analyses of which the author is aware. Any ANOVA analysis performed on the data did not detect the trend seen in Figure 25.

group when the groups were semantically cohesive and two fixations when randomly organized. Within the cohesive groups, participants were more likely to use one fixation when labels were present. Participants were also more likely to make four or more fixations per group when the groups were labeled.

Besides the experimental factors, only thee factors affected the participants' eye movements. These are (a) the distance between the precue and target, (b) the participant's OSPAN score, and (c) a carryover effect from previous blocks. Fewer, $F(1, 4192) = 345.53$, $p < .001$, and shorter, $F(1, 4213) = 167.94$, $p < .001$, saccades were used when the target was closer to the precue. Participants with a higher work memory capacity, as indexed by the OSPAN task results, used shorter fixations, $F(1, 13.6) = 7.06$, $p = .02$. Fixations were longer in blocks that were preceded by blocks with semantically cohesive, unlabeled layouts, whether with meta-groups present, $F(1, 1298) = 3.87$, $p = .05$, or with meta-groups absent, $F(1, 1077) = 8.89$, $p < .01$. Fixations were also longer in blocks that were preceded by blocks with semantically random, unlabeled layouts, with meta-groups absent, $F(1, 1127) = 12.00$, $p < .01$.

No other factors affected participant's eye movements. The number of fixations per trial did not vary with:

OSPAN scores, $F(1 , 12.2) = 0.19$, $p = .67$
Gender, $F(1, 12.2) = 0.00$, $p = .95$
Age, $F(1, 12.3) = 0.01$, $p = .94$
Computer Experience, $F(1, 12{,}2) = 0.11$, $p = .75$
Carryover from layouts with semantically cohesive group, labels,
and no meta-groups, $F(1, 1460) = 0.29$, $p = .59$

Carryover from layouts with semantically cohesive groups, labels,
and meta-groups, $F(1, 1503) = 0.60$, $p = .44$
Carryover from layouts with semantically cohesive groups, no labels,
and no meta-groups, $F(1, 1457) = 0.32$, $p = .57$
Carryover from layouts with semantically cohesive groups, no labels,
and meta-groups, $F(1, 1536) = 1.68$, $p = .20$
Carryover from layouts with random groups, no labels,
and meta-groups, $F(1, 1513) = 0.36$, $p = .55$

Participants' fixation duration did not vary with:

Target distance, $F(1, 4144) = 0.00$, $p = .95$
Gender, $F(1, 13.8) = 1.08$, $p = .32$
Age, $F(1, 12.9) = 3.48$, $p = 0.08$
Computer experience, $F(1, 13.5) = 4.53$, $p = 0.06$
Carryover from layouts with semantically cohesive group, labels,
and no meta-groups, $F(1, 1289) = 0.08$, $p = .78$
Carryover from layouts with semantically cohesive groups, labels,
and meta-groups, $F(1, 1073) = 1.33$, $p = .2499$

Participants' saccade distance did not vary with:

OSPAN score, $F(1, 13) = 1.77$, $p = .21$
Gender, $F(1, 13) = 4.15$, $p = .06$
Age, $F(1, 13) = 0.30$, $p = .59$
Computer experience, $F(1, 13) = 0.01$, $p = .93$
Carryover from layouts with semantically cohesive group, labels,
and no meta-groups, $F(1, 1443) = 0.03$, $p = .87$
Carryover from layouts with semantically cohesive groups, no labels,
and no meta-groups, $F(1, 1438) = 0.00$, $p = 0.99$
Carryover from layouts with semantically cohesive groups, no labels,
and meta-groups, $F(1, 1513) = 2.51$, $p = .11$
Carryover from layouts with random groups, no labels,
and meta-groups, $F(1, 1492) = 0.07$, $p = .79$

3.3.3 Discussion

This study investigates the effects of semantic content and visual indicators of semantic relations on visual search. The data strongly support the hypothesis that people use the structure provided by the semantic content of the words in the layout to guide their search. The unexpected results are what the data suggest about how people use visual indicators of semantic relations.

People search layouts faster when the groups are semantically cohesive. This is not surprising considering that in the semantically cohesive layouts, the meaning of non-targets provide strong cues about the target location, and no similar information is provided in the random layouts. As seen in Figure 25, people are more likely to make just a single fixation to a group in the cohesive layouts. This suggests that people tend to judge the semantic relevance of all objects in a group with that one fixation. This allows the participants to "explore" more of the layout per fixation and thus reduces the number of fixations required to find the target. Conversely, without the semantic content, it is more difficult or impossible to discount an entire group of objects with just one fixation. Both the fixations per trial and saccade distances also support this conclusion.

While the semantic content seems to provide useful information, a first pass of the data suggests that the group labels provided no additional useful information. That is, the results suggest that people use labeled and unlabeled layouts similarly when the groups are semantically organized. This null result would seem to contradict previous research that showed the importance of group labels in users' visual search strategies (Hornof, 2004). This previous finding was supported by results from a task in which no useful semantic information was involved in the search. So, can we conclude from this study that labels are not useful when layouts are semantically cohesive? Almost, but while the labels did not affect the search time or the total number of fixations needed to find a target, the real story is in the detail of the eye movements.

Further analysis of the participants' eye movements supports previous claims of the importance of group labels in visual search strategies. Additionally, the results show how the utility of group labels is extended to group items in semantically organized layouts. Previous research on the eye movements motivated by the presence of group labels found that people tended to use just one fixation per group (Hornof & Halverson, 2003). Comparable results were found here. As shown in Figure 25, when the groups were unlabeled and cohesive, people behave much more like they do when searching labeled groups. The participants tended to make just one fixation, presumably evaluating all words in the group based on the words processed in that one fixation. One way to interpret these results are that people were using any word in unlabeled and cohesive groups as the label for that group.

The eye movement data also differentiate the use of labels and non-labels as semantic indicators. While the semantic grouping had more of an effect than the labels, if we look at the distributions in Figure 25, we can see that people were more likely to use one fixation per group in the semantically organized layouts when the groups were labeled. People were also more likely to make four or more fixations per group when the groups were labeled. It appears as if people had more "trust" in the group labels. That is, people were more likely to discount the contents of groups based on the group label, thus more one-fixation group visits, and more committed to searching a group when they believed the target to be in a group based on the label, thus more four-or-greater fixation group visits.

The results of this experiment extend our understanding of how semantics and group labels affect users' visual search strategies. While previous research has shown the utility of labels or the effects of semantic differences between words in a menu, this research looks at the combined effects of both labels and semantic differences. Unexpected results were found, such as the how the semantic cohesion of words in a group can substitute, to some extent, for labels of those groups.

3.4 Summary

A series of experiments were conducted to better understand how people visually search computer screens and motivate the development of an active vision computational models of visual search. The effects of density, text color, and the semantic cohesion of groups of words were studied. These studied extend previous psychological research. More importantly, the experiments presented in this thesis identify ways in which these factors affect people's human-computer visual interaction.

The local density of groups of words not only affects the speed with which people search words, but also the order in which the groups are searched. It was found that sparse groups are searched first and faster than dense groups. Additionally, in mixed-density layouts when dense groups were searched early, they are searched in a suboptimal manner relative to all-dense layouts. This suggests that interface designers should not only use sparse groups to draw users' attention to important information, but that caution should be used when mixing densities because information in denser groups of words may be missed by the users. This research also suggests that a good computational model of visual search will need to account for the strategies that people use and include appropriate parameters to simulate the difficulty of perceiving denser and perhaps less salient text.

While people are able to limit their search for words based on the color of the words, they evidently cannot completely ignore words of a different color, at least for those words further from the point of gaze. The presence of non-target-colored words slows people's visual search of target-colored words. The results of this research strongly supports the practice of differentiating link status (i.e. visited or unvisited) with peripherally visible color.

Finally, when searching, people can use semantically cohesive structures with or without group labels much the same. When groups of words are semantically related, people can evaluate an entire group in one fixation. This behavior can occur whether labels identifying the category of the group are present or not. Nonetheless, people often use just one fixation to evaluate labeled groups more often than they use just one fixation to evaluate the unlabeled groups. This suggests that group labels can be excluded and users will likely still perform well, as long as the groups are meaningfully grouped . This finding is especially useful when screen real estate is limited, such as on handheld computers, and the absence of group labels would reduce screen clutter.

The results from these experiments informed the construction of an active vision computational cognitive model of visual search. The tasks presented and the data analyses will be particularly useful for modeling visual search in the context of human-computer interaction. Each task utilized structured layouts that approximate real-world interfaces like those shown in Figures 1 and 2. Particularly important for the development of models of visual search, precise eye movement data were collected. Not only was aggregate eye movement data analyzed, but also eye movement data that uncover the strategies people use, like the order in which visual groups in the layout are searched. The next chapter discusses the development of the model using reaction time and eye movement data.

CHAPTER IV

MODELS

This chapter presents a candidate 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 between fixations? The proposed answers to these questions, and the research these answers on based on, will be covered in this chapter.

Throughout this chapter, eye tracking data from two experiments is used to improve the models. A principled approach is proposed for building models of visual search based on a step-by-step improvement of the model using the most appropriate eye movement measurement and model parameters. In this way, the model of active vision is developed, refined and enhanced by accounting for more and more eye movement data.

An aim of this research is to propose a model of visual search that can be used for automated interface analysis tools, like CogTool (John & Salvucci, 2005) or CORE/X-PRT (Tollinger et al., 2005). The criterion for acceptable predictions by the models is a 10% average absolute error (AAE) between the observed and predicted data. A 10% AAE will demonstrate that the models are reasonably accurate for such engineering goals (Kieras, Wood & Meyer, 1997).

All models presented here were built using the EPIC cognitive architecture (Kieras & Meyer, 1997). EPIC lends itself well to developing models of active vision, as it accounts for constraints imposed by the eye and eye movements. As discussed in Chapter II, EPIC is an computational framework written in C++. Some modifications were made to the C++ classes representing EPIC's visual processors during the iterative process of refining the models. These modifications are discussed in the following sections. The production rules for the final model are presented in Appendix C.

4.1 Modeling the Local Density Task

The first task modeled in this research was the local density task presented in Chapter III, section 3.2. The data collected using this task provided sufficient detail to inform the construction of a model of visual search.

The modeling focused on the issues raised by previous research on density, e.g., the number of items perceived per fixation, and other fundamental perceptual and ocular motor issues of visual search. Previous modeling has used data from eye tracking to inform the development of models with respect to the order of search (e.g. Byrne, 2001). However, in this section, the focus is on fixation duration and number of fixations in order to inform the development of other aspects of computation models.

This research 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.

4.1.1 Baseline Random Search Model

The initial model in this research, which will be referred to as the Baseline Model, starts with the a small set of reasonable assumptions. Initial assumptions of the modeling include constraints supplied by the architecture. All of EPIC's perceptual properties were left at established values. The assumptions included: Text centered within 1° of the point of fixation will enter working memory after 149 ms. Saccades took time to prepare, from 50 ms if the previous saccade had the same direction and extent up to 150 ms if the previous fixation had a different direction and extent. Saccades took 4 ms per degree of visual angle (dov).

A couple of initial assumptions were extracted from the literature. First, the model searched "without replacement." That is, any object for which the text had been perceived was 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 were selected at random. There is a scarcity of evidence for where search will proceed in layouts that consist of text. However, 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 included a production-rule strategy that executed the task as follows: The model fixated and memorized the target precue. As soon as the visual search layout appeared, the model started searching for the target. The model moved its eyes to a random word in the layout. As soon as the eyes arrived at the saccade destination, the model initiated the next eye movement to a random object whose text had not entered working memory. If at any time the target was identified, search was terminated, the eyes were moved to the target, and the target was clicked.

4.1.1.1 Predictions

The Baseline Model was overall a poor predictor of human performance. As seen in Figures 26 and 27, the predicted search times and fixation durations are incorrect both in value and trend. Nonetheless, as can be seen in Figure 28, this rudimentary model accurately predicts the observed number of fixations per trial for one condition. While overall this model incorrectly predicts the number of fixations, the prediction for the sparse layouts is quite good.

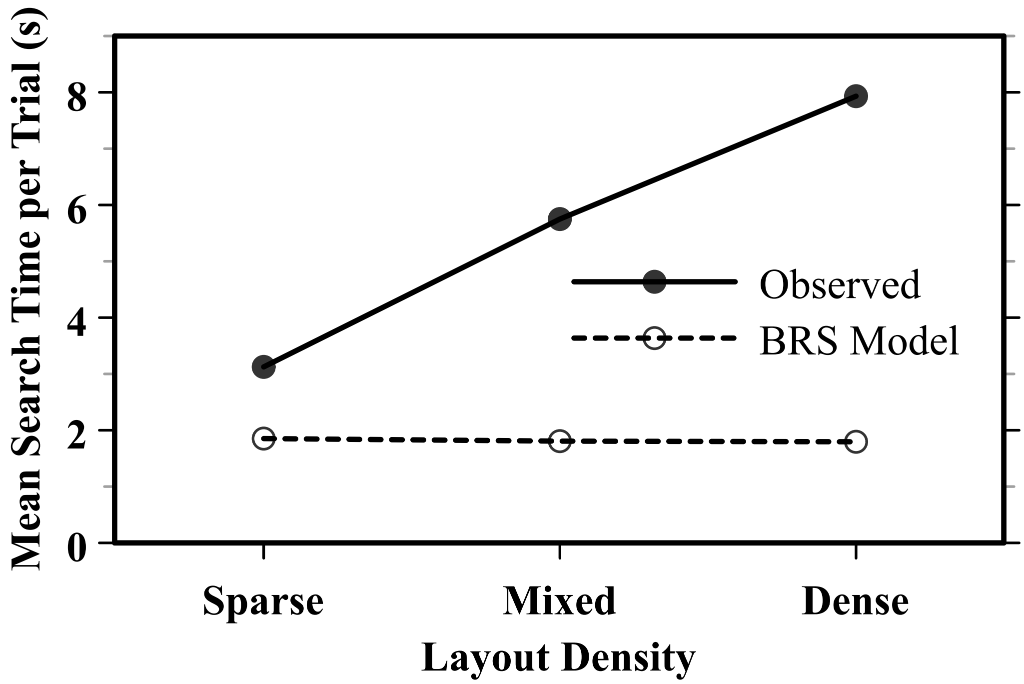


Figure 26. Mean search time per trial observed (solid line) and predicted (dashed line) by the baseline random search (BRS) model for the mixed-density task. Average absolute

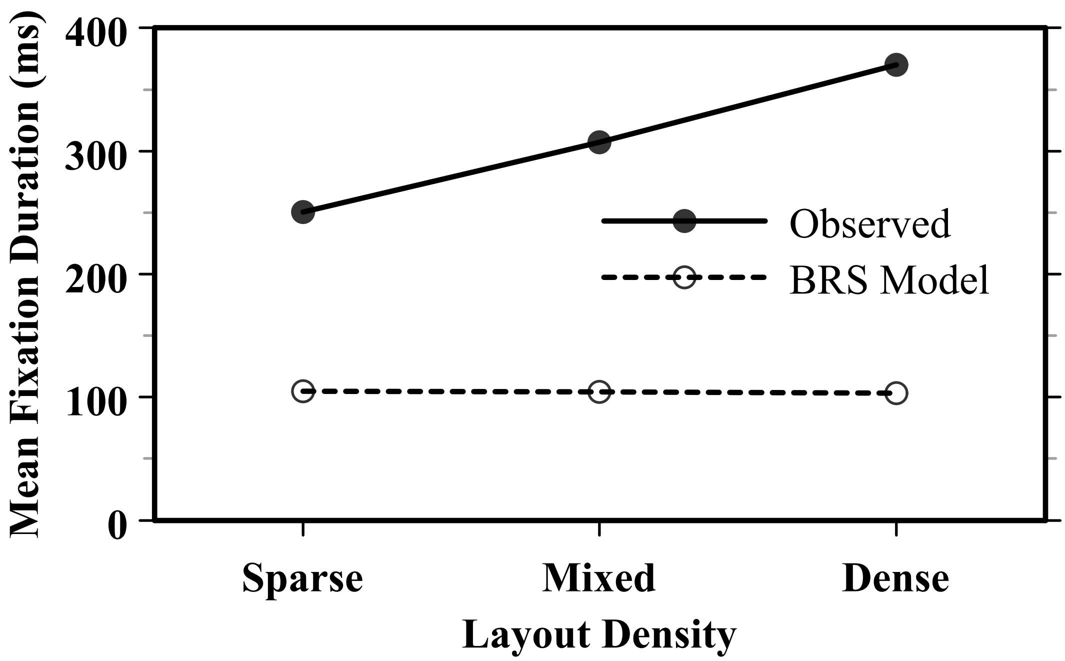


Figure 27. Mean fixation duration observed (solid line) and predicted (dashed line) by the BRS model for the mixed-density task. AAE = 65.5%

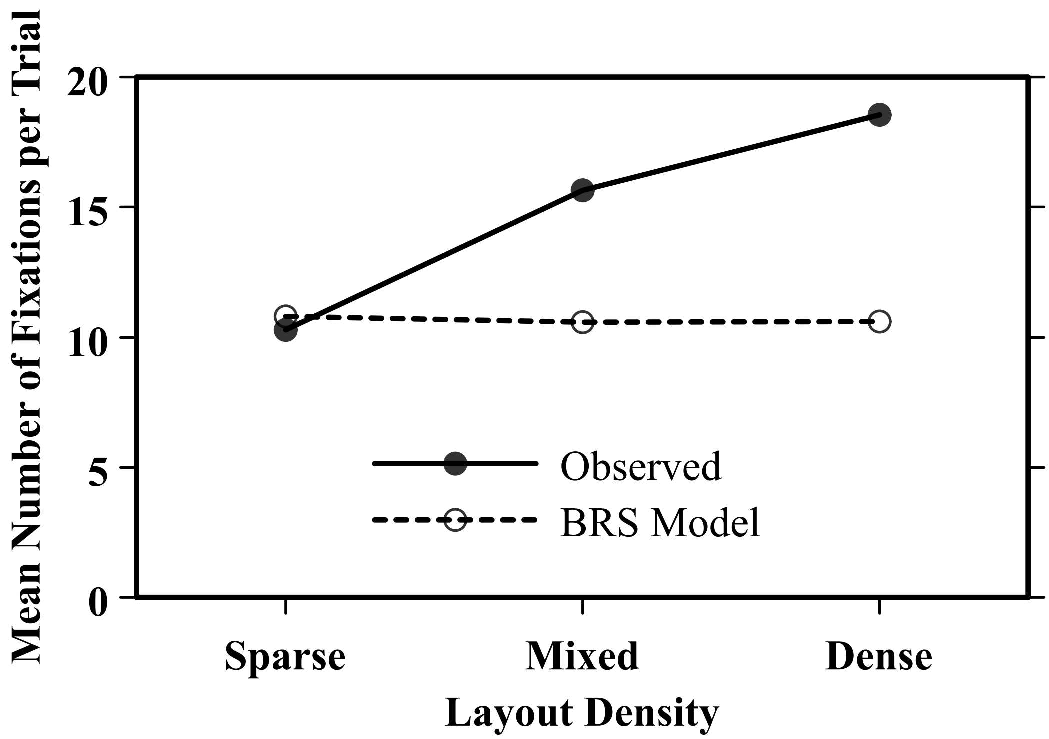


Figure 28. Mean number of fixations per trial observed (solid line) and predicted (dashed line) by the BRS model for the mixed-density task. AAE = 26.7%

4.1.1.2 What Was Learned

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 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. The greatest error was found in the fixation duration predictions, with a 65.5% AAE. Saccades were initiated as soon as they could be by the model, 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 explored the initiation of saccades to improve the model's predictions of fixation durations.

4.1.2 Improving the Predictions of the Fixation Duration

The observed eye movement data from the Local Density experiment presented in Chapter III is used once again to guide the model development. As 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 the fixation durations produced by the model.

One of the things the Baseline Model got wrong was that the model initiated a saccade to the next randomly chosen object as soon as the previous saccade was complete. However, based on the results of the Local Density experiment, people appear to adopt a search process that increased the duration of fixations on smaller, denser text. This could be achieved 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.

The model can be evaluated in the context of the four explanations of fixation duration described by Hooge and Erkelens (1996) and discussed in Chapter II, namely preprogramming-per-trial, preprogramming-per-fixation, strict process-monitoring, and mixed-control. Preprogramming of fixation durations alone does not explain the mixed-density data very well. As shown in Figure 12, the 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. Further, in the mixed-density layout trials, where a change in fixation duration is observed in the dense groups, the change in duration is quite sudden. A preprogramming-per-trial explanation would predict no change in duration during a trial. A preprogramming-per-fixation explanation would likely predict, if anything, a gradual change in fixation duration. Instead, the participants' fixation duration tended to increase dramatically halfway through the search. This suggests a strategy shift and not a change in estimation.

A strict process-monitoring strategy of saccade initiation (Hooge & Erkelens, 1996) explains the mixed-density data better. As shown in Figure 12 in Chapter III, the fixation durations vary as a function of the density of the group fixated, which supports the notion of strict process-monitoring. The increase in fixation duration in the dense groups of the mixed-density layouts may support the mixed-control explanation. However, the mixed-control explanation is less parsimonious than the strict process-monitoring alone. Additionally, 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. While this fit between the theory in EPIC and the process-monitoring does not make process-monitoring “right”, Newell (1990) suggested “listen[ing] to the architecture” to find reasonable solutions. Additionally, instantiating the preprogramming hypotheses of saccade initiation would require additional mechanisms and parameters that are not required with the process-monitoring strategy decreasing the parsimony of the model. A preprogramming saccade initiation strategy might require a theory of time perception (such as Taatgen, Rijn & Anderson, 2007) and to predict saccade time intervals. The introduction of such temporal mechanisms 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 nice mesh of extant theory.

4.1.2.1 Strict Process-Monitoring Strategy Model

The strategy rules were modified to include a strict process-monitoring strategy. Figure 29 shows a flow-chart based on the production rules. This strategy, which shall be called the prepare-then-wait strategy, initiates saccades only after the text property for the current saccade destination becomes available and a decision has been made whether the target has been found or not.

EPIC’s perceptual processor was modified to accommodate a strict process-monitoring, as follows. The default recoding time for text is a constant 100 ms. This was modified when trying to explain the human data. As shown in Table 1, 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 calculate the perceptual time for a feature based on the proximity of adjacent items. If an object’s closest neighbor

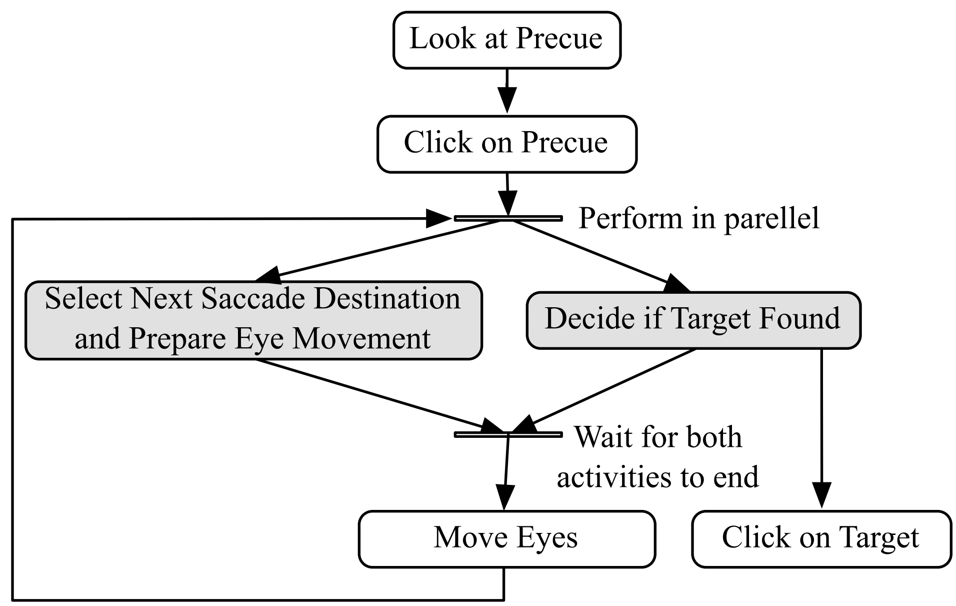


Figure 29. Flow chart of the production rules for an instantiation of the strict process-monitoring strategy.

was closer than 0.15 dov (a dense object), the text recoding time was 150 ms. Otherwise the text recoding time was 50 ms.

4.1.2.2 Predictions

As shown in Figure 30, a large improvement was found in the predicted fixation durations when the model was modified to use a strict process-monitoring strategy. Delaying the initiation of saccades until after the text information had entered working memory and an increased recoding time for dense objects resulted in a differentiation in fixation durations similar to that in the observed data. The predicted data could have been further improved by reducing the text recoding time for sparse objects further, as the majority of the error in the predicted data lies in the sparse and mixed layouts. However, the purpose of this modeling was to approximate the ocular-motor behavior in the

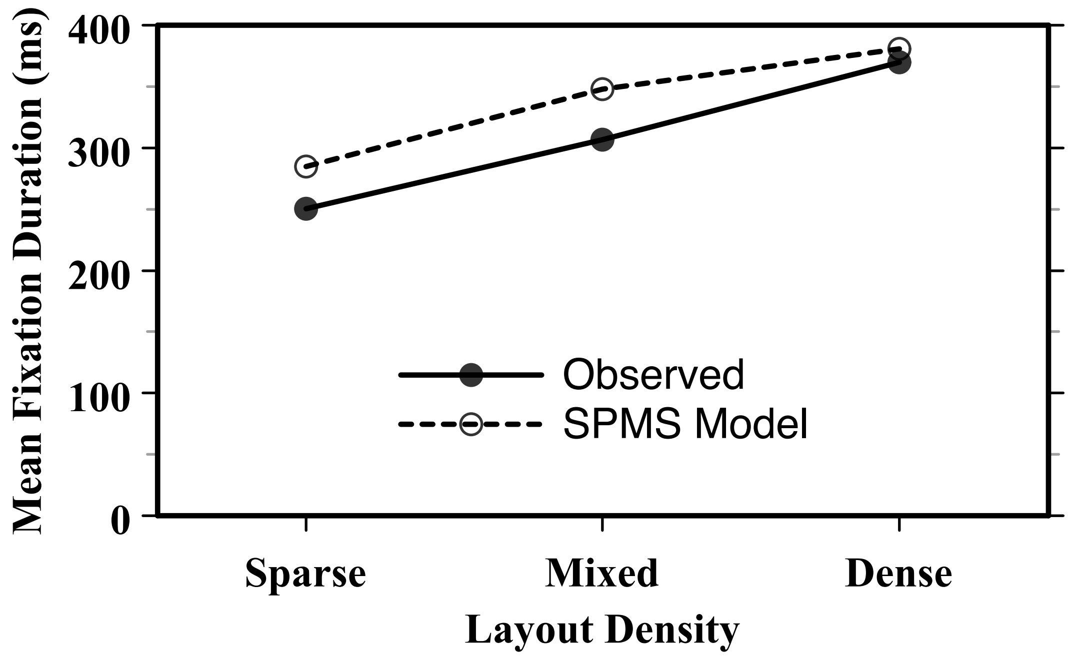


Figure 30. Mean fixation durations observed (solid line) and predicted (dashed line) by the strict process-monitoring strategy (SPMS) model for the mixed-density task. AAE = 10.0%

observed data and meet the criterion for acceptable predictions of 10%, so further fine-tuning of fixation durations was not performed.

The predicted mean search time only improved slightly. As seen in Figure 31, 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 32, the predictions for the number of fixations per trial worsened. The model still does not make more fixations in layouts with dense objects, as is seen in the observed data. Further, the overall mean number of fixations has dropped in comparison to the base model. Detailed traces of the models revealed that the drop in the mean number of fixations was due to the prepare-then-perform strategy. The Baseline

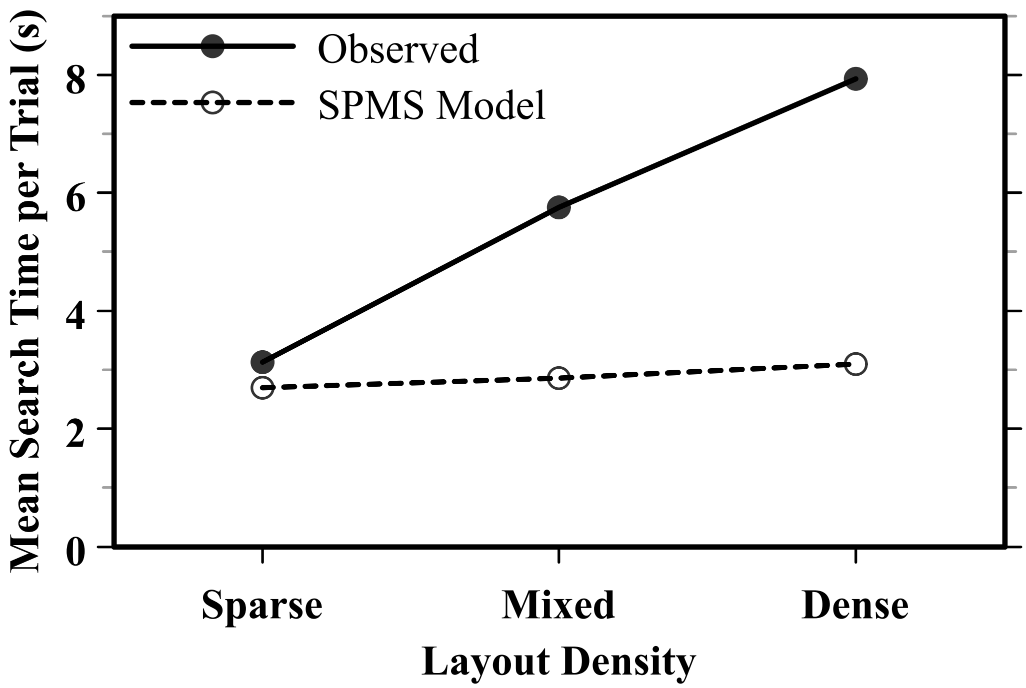


Figure 31. Mean search time per trial observed (solid line) and predicted (dashed line) by the SPMS model for the mixed-density task. AAE = 41.7%

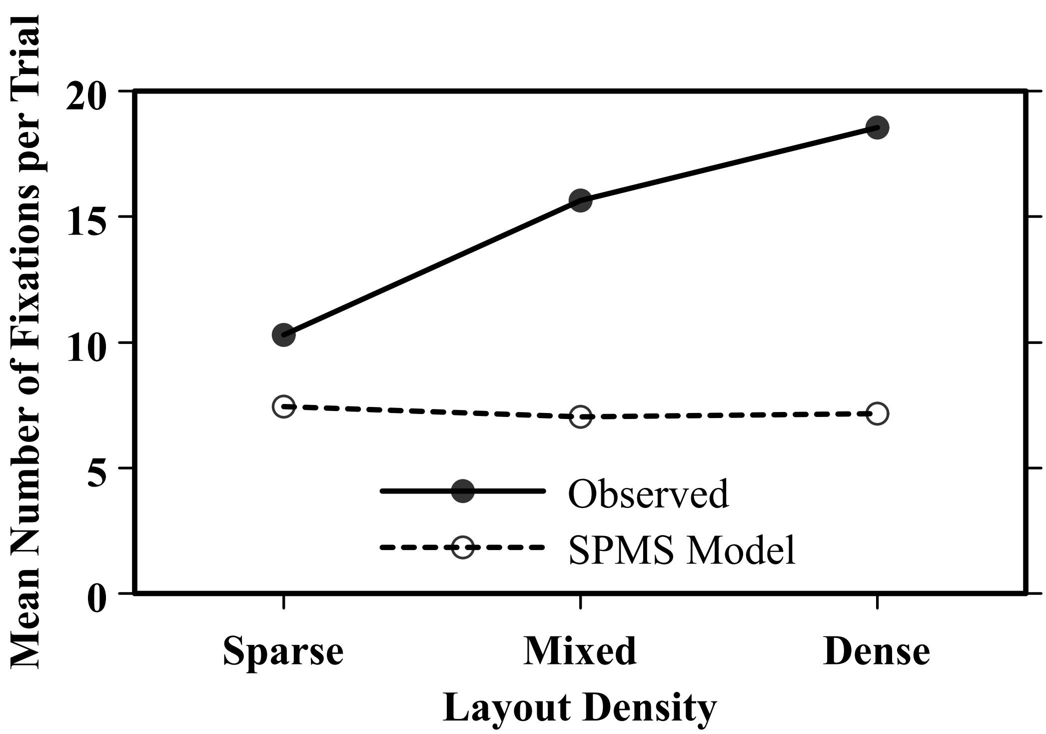


Figure 32. Mean number of fixations per trial observed (solid line) and predicted (dashed line) by the SPMS model for the mixed-density task. AAE = 48.1%

Model initiated approximately three additional fixations after the target had been fixated but before the text property for the target had become available. This resulted in roughly three more fixations per trial than the prepare-then-execute strategy, which inhibited additional fixations until the text property was perceived.

4.1.2.3 What Was Learned

A model 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 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 in this research, 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.

4.1.3 Improving the Predictions of the Number of Fixations

The number of fixations predicted by the model is largely determined by a simplifying assumption about the area in which text is perceived. The assumption was that all text within the fovea (1 dov) is 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 suggests that humans do not do this. People require more fixations for dense text. An increase in the number of fixations predicted for dense objects can be achieved 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.

Adjusting the region in which text can be perceived, such 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). EPIC's default settings already limited sparse words to two or three per fixation. Different region sizes for dense text were tried, and 0.5 dov worked best, resulting in two to three words per fixation in dense text. Perceiving two to three words per fixation, regardless of text density, resulted in a much better fit for the predicted number of fixations per trial. However, as shown in Figure 33, the model was still under-predicting the number of fixations per trial in all layouts, and so this words-

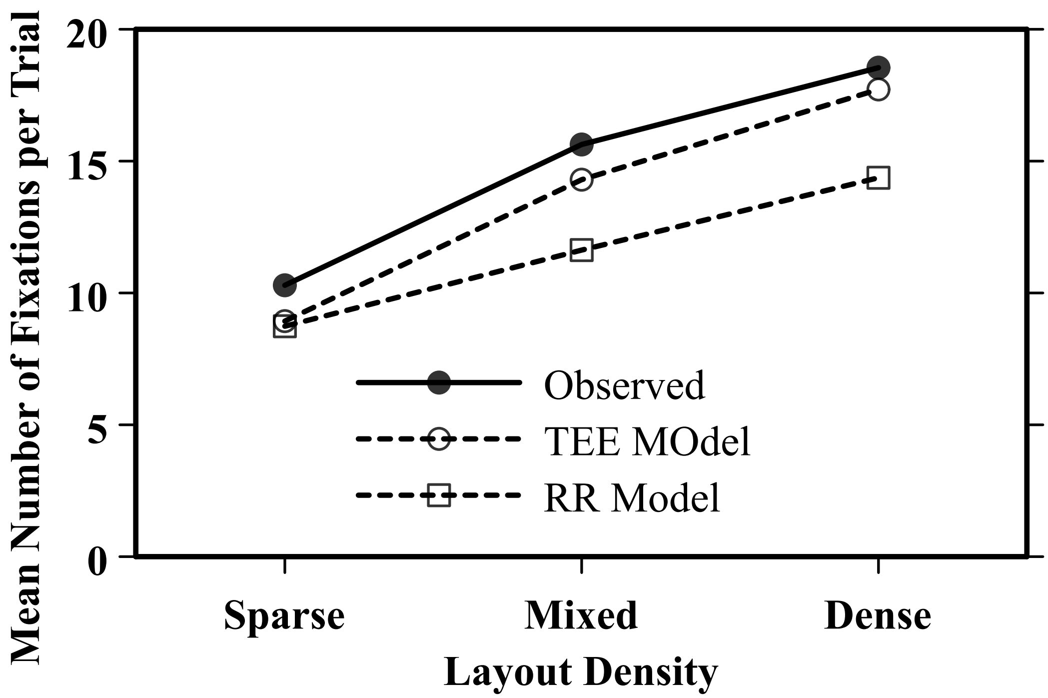


Figure 33. Mean number of fixations per trial observed (solid line), predicted by the Text-Encoding Error (TEE) model (dashed line with circles), and the Reduced-Region (RR) model (dashed line with squares) for the mixed-density task. The AAE of the TEE model is 8.8% and the RR model is 21.1%.

per-fixation approach was passed over in favor of a probability-of-encoding approach discussed next.

4.1.3.1 Text-Encoding Error Model

To adjust the probability of incorrectly encoding text, EPIC's perceptual processor was 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 the screen. If an object's closest neighbor was 0.15 dov away or more (sparse text), the probability of the model incorrectly perceiving the text was 10%. Otherwise, the probability of the model incorrectly perceiving the dense text was 50%. These probabilities were chosen because they would result in two to three items, on average, perceived per fixation across densities, which appeared to be the right number of items per fixation to explain the human data.

4.1.3.2 Predictions

As seen in Figures 33 and 34, 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 data points, number of fixations and search time across all three layouts, were greatly increased by

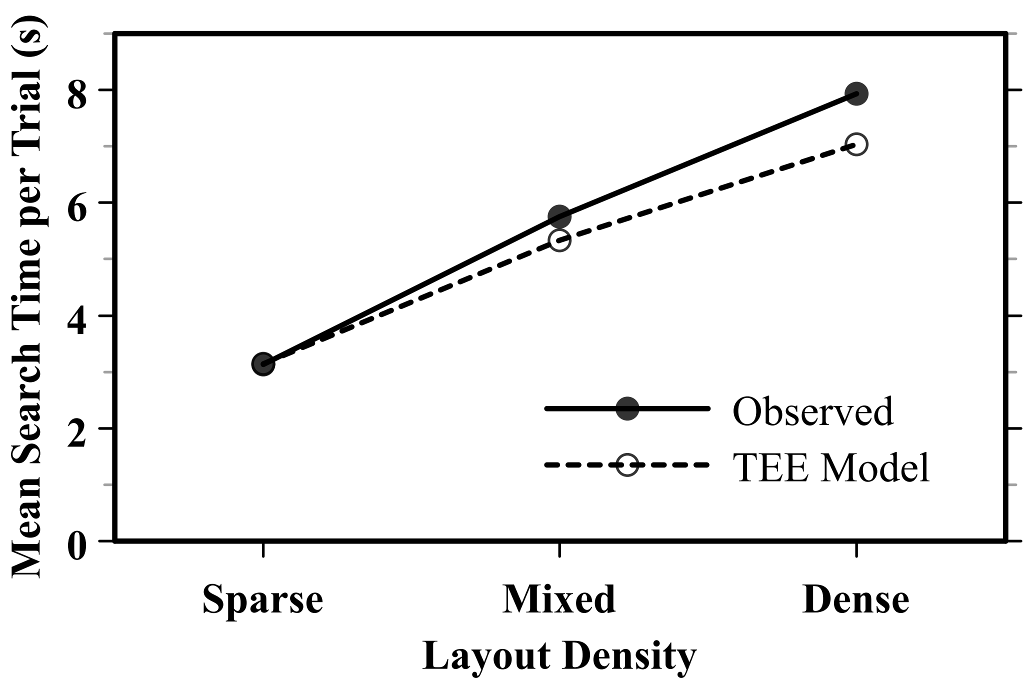


Figure 34. Mean search time per trial observed (solid line) and predicted (dashed line) by the TEE model for the mixed-density task. AAE = 6.5%

adding one perceptual parameter, a 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 exclusively on the features of those visual objects.

4.1.3.3 What Was Learned

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 are perceived in every fixation, the model underpredicts the number of eye movements the humans need to find the target. Reducing the area in which the model could perceive text did not predict the human behavior well. When the model was modified to include the possibility of misperceiving text, the predictions of the number of fixations used in each layout became very good.

4.1.4 Discussion

During *exploratory modeling* (i.e. modeling to investigate how observed human data can be explained), a random search strategy is a reasonable first approximation that allows the analyst to focus on other fundamental ocular-motor activity that affects visual search. If an analyst can initially account for fundamental perceptual and ocular-motor activity with such a parsimonious strategy, the analyst may find it easier to explore other important aspects of the model, like the time to encode and probability of encoding the visual objects, which are unrelated to the order in which the objects are explored.

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. In the model, saccades are initiated as soon as the relevant visual features (i.e. target-identifying features, 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. This suggests that visual features necessary to identify the target will affect the subsequent saccade destination, but that any unnecessary features that may take longer to enter working memory will *not* affect the saccade destination. Confirmation of this hypothesis is left for future research.

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 in that limiting the number of objects perceived per fixation to two to three items helped predict observed search times. The same assumption here helped to predict search time and number of fixations. Note, however, that different kinds of text (e.g. larger font, larger spacing, or longer phrases) might require more fixations per word.

Bertera and Rayner (2000) concluded that the effective field of view, the region in which information is used during a fixation, did not decrease as density increased. The findings here support that conclusion and expand upon it. The task modeled in this research differed from that used by Bertera and Rayner, which used randomly arranged single letters. In the current task, density was manipulated by varying the size of text and spacing (which is arguably more ecologically valid). Still, similar conclusions were reached. Future work is required to study the effects of density where text size and spacing vary independently.

4.2 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, such as those shown in Figures 1 and 2 in Chapter I. The progression towards a model of active vision continues here with the modeling of a second set of data, the CVC (consonant-vowel-consonant) search task (Hornof, 2004). The CVC task is called such because the task used three-letter pseudowords (such as ZEJ, HAN, NUH) that were used to control for word familiarity and other effects. This stage of the modeling primarily focused on two issues — evaluating previous assumptions in the model and refining the model to account for additional eye movement measures.

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. It is useful to return to the CVC search task because there are adequate similarities and meaningful differences between the local density task and the CVC search task. Again, the layouts consisted of text only. However, 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.

The CVC task included layouts *with* and *without* a visual hierarchy. The layouts discussed in this dissertation are those without a visual hierarchy. Figure 35 shows a sample layout from the experiment.

Sixteen people participated in the most recent replication of the study in which eye movement data was collected (Hornof & Halverson, 2003). 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. One-group layouts contained only group A in Figure 35. Two-group layouts used groups A and B. Four-group layouts used groups A through D. In each trial, the entire layout was displayed at the same moment, permitting any search order. The trials were blocked by layout.

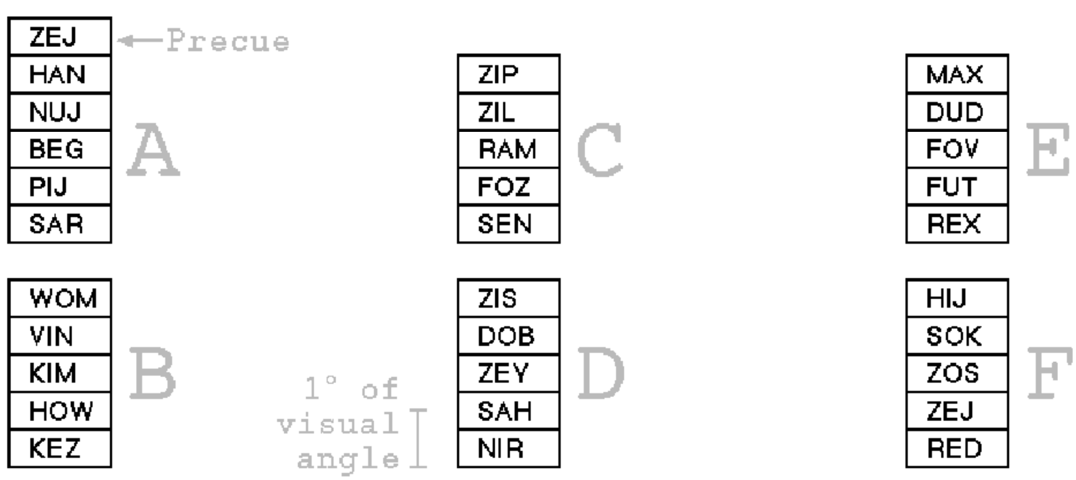


Figure 35. A layout without labels from Hornof’s (2004) CVC search task.

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.

4.2.1 Original CVC Model

Hornof (2004) presented models that predicted and explained the search time data collected from the visual hierarchy task. In the model, the eyes moved down the first column of text, then down the second column, and then down the third. Furthermore, the eyes jumped over a carefully controlled number of items with each eye movement, sometimes missing items on a first pass, which introduced some "noise" into the model and helped explain the human search time data. This selection strategy resulted in a plausible explanation for how people did the task in that the model accounted for the reaction time and a fair number of eye movement measures. It is perhaps impressive that the models correctly accounted for some of the eye movement data in that the models were built without any eye movement data to guide the development of the models.

However, the model's strategy is perhaps somewhat overly tuned to aspects of this one visual task and layout. Components of the strategy, such as the strict use of the three columns, will not be directly applicable to a wide range of visual layouts. The original CVC task model might thus be characterized as somewhat brittle, whereas a more flexible model might be more useful for predicting human performance in a wider range of visual search tasks.

This concern motivated a more flexible model that would predict the eye movements with greater fidelity and in a more general, task-independent manner. The data collected by Hornof and Halverson (2003) are used to further improve on the model of active vision developed in this dissertation.

4.2.2 Improving Saccade Distance

The previous model 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, the location of where the model fixates requires improvement.

One job of the human visual search process 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).

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.

Previous research supports the idea that people tend to fixate nearby objects. Fleetwood and Byrne's model of icon search (2006) shifted covert visual attention to the nearest icon that matched the target icon in some way. The model was very systematic in the sense that it always chose the closest icon that met other criteria that the research emphasized more, namely looking at icons that matched one randomly chosen feature of the target icon. Motter and Belky (1998) investigated where monkeys were likely to detect a target during a fixation and found that the monkey's eyes were more likely to move towards objects just outside the region in which targets could be detected. Additionally, an important finding for the current research was that, although the eyes were more likely to go to nearby objects, they did not always go to the nearest object.

4.2.2.1 Fixate-Nearby Model

The strategy used by the model was modified so that saccades were more likely to land on nearby items, as follows. Rather than searching randomly or following a prescribed search order, as with previous models, the strategy selected 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 is selected as the next saccade destination.

The strategy used by the model was also modified to reduce how often the model would 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 current research takes a straightforward approach to modeling this behavior: A constraint was added to inhibit group revisits until the entire layout had been searched. Without this constraint, the model was much more likely to revisit a group than found in the observed data.

4.2.2.2 Removing Text-Encoding Errors from the Model

In an effort to explain the eye movement data and to depict the human information processing that is not directly observable, two mechanisms have been introduced to the mode: (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 local-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. Encoding errors were previously used to explain the additional saccades sometimes required to re-examine the layout. So that the model does not include two explanations for one phenomenon, the encoding errors were removed.

4.2.2.3 Predictions

As shown in Figure 36, the model predicts the mean saccade distances very well, with an average absolute error (AAE) of 4.2%, a considerable improvement over the AAE of 43.3% in the Original Model. As shown in Figure 37, this model also does a good job of

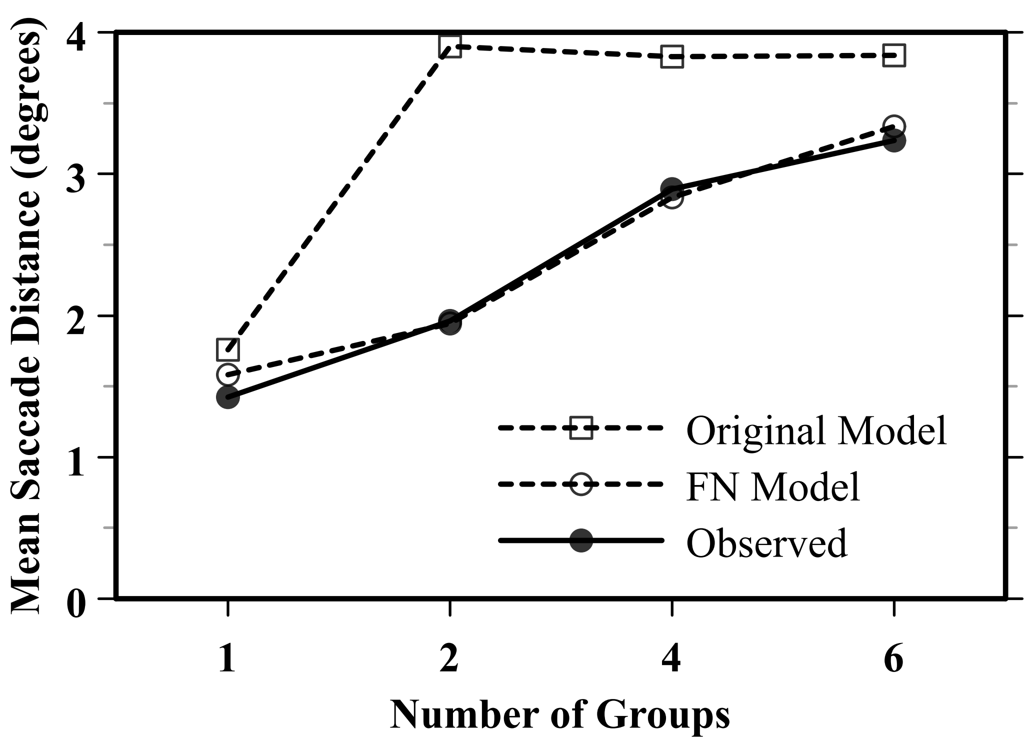
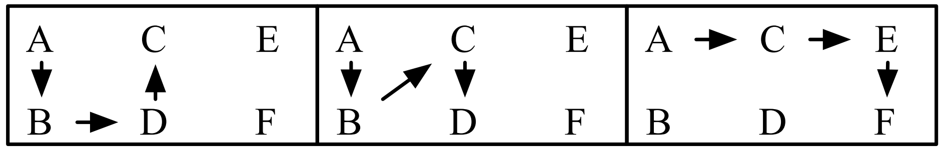


Figure 36. Saccade distance observed in the CVC search task (solid line), predicted by the original CVC search task model (dashed line with squares), and predicted by the Fixate-Nearby (FN) Model (dashed line with circles). The AAE of the original model is 43.3% and of the FN model is 4.2%.



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

Figure 37. 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 the models (original and FN).

predicting the observed scanpaths. The figure shows the three most frequently observed scanpaths, and how the current model predicts the observed scanpath frequencies better than does the Original Model (Hornof, 2001). However, as shown in Figure 38, the predicted number of fixations per trial is not within our intended AAE of 10%, although the predicted number of fixations did improve considerably (AAE = 14.3%) compared to the Original Model (AAE = 37.8%).

4.2.2.4 What Was Learned

Results from this modeling suggest that people select saccade destinations partly based on eccentricity from fixation center. The selection of saccade destinations based on proximity resulted in good fit of both the mean saccade distance and the scan paths that people used in this task. The model with random saccade destinations predicted saccade distances much larger than is seen in the observed data. Additionally, the random selection predicted 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

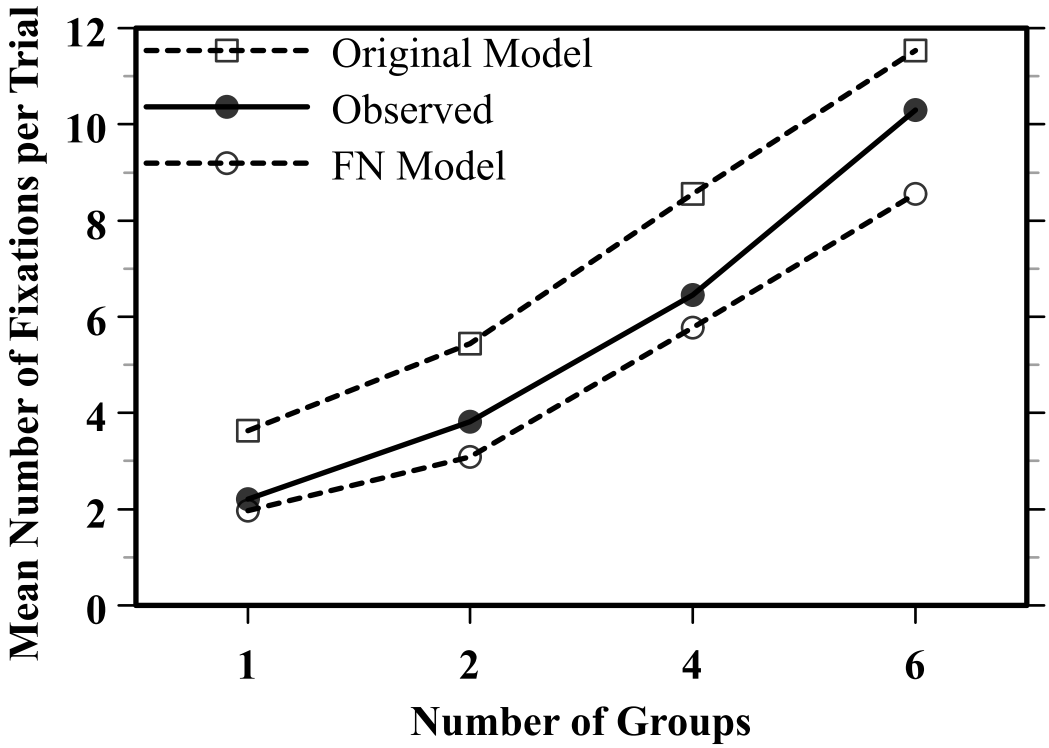


Figure 38. Fixations per trial observed in the CVC search task (sold line), predicted by the original model (dashed line with squares), and predicted by the FN model (dashed line with circles). The AAE of the original model is 37.8% and of the FN model is 14.3%.

saccade distances seen in Figure 36 is accounted for. The observed and predicted saccade distances increase with the size of layout. 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 "nearby with noise" 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, only the location of the layout objects is required. This is beneficial if other properties in the layout are unknown or difficult to extract. 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 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.

While the fidelity of the model improved overall, the predictions for the number of fixations was not acceptably accurate. While the accuracy of the predictions improved and the number of fixations increased with the number of items in the layout, the predicted number of fixations diverged from the observed number of fixations as the layouts grew in size. The model was able to find the target with fewer fixations than the participants did. In the next section, the focus of the modeling returns to the number of fixations.

4.2.3 Revisiting the Predictions of the Number of Fixations

Text-encoding errors were removed from the model presented in the previous section, but the model still underpredicts the number of fixations per trial. It was speculated earlier that text-encoding errors introduced while modeling the local-density task might not be needed because of the changes made in the previous section. However, when text-encoding was removed, the model again underpredicted the number of fixations. Text-encoding errors are reintroduced in order to improve the model’s predictions for the number of fixations per trial.

4.2.3.1 Text-Encoding Errors Revisited Model

The text-encoding error rate was once again set to the previous parameter value. This error rate was changed by 1% increments until the model predicted the mean number of fixations per trial well. A value of 9% provided the best fit for the number of fixations per trial.

4.2.3.2 Predictions

As shown in Figure 39, the text-encoding revisited model predicts the number of fixations per trial very well, with an AAE of 4.2%, which meets our goal of an AAE of 10% or less. The introduction of text-encoding failures improved the predictions.

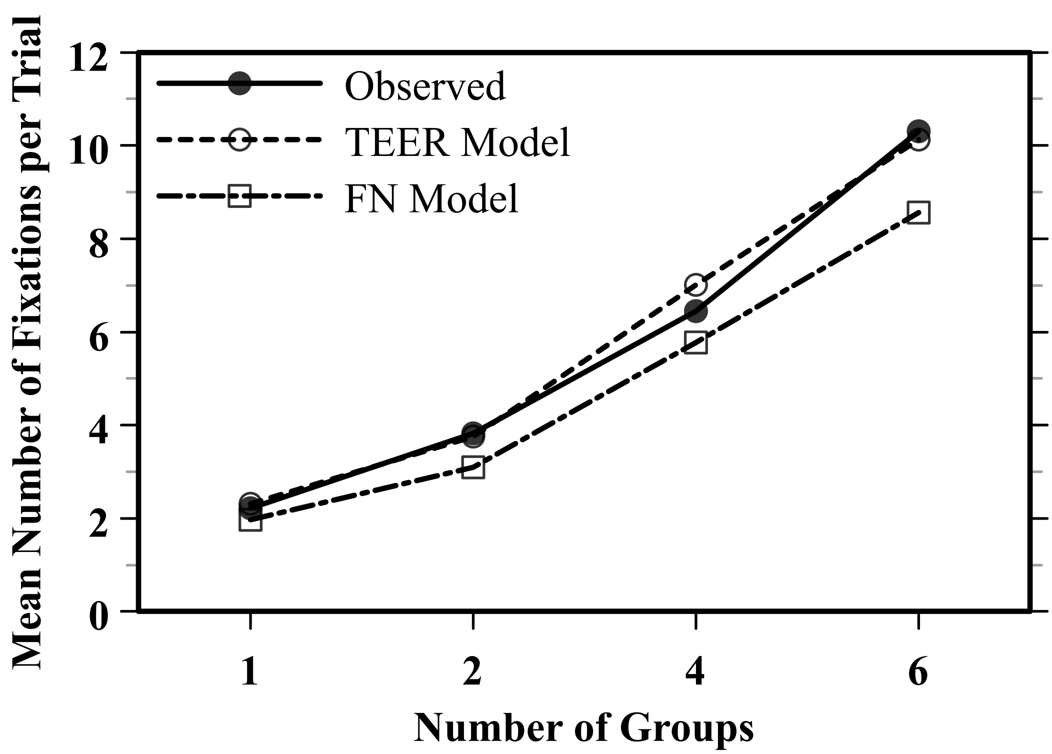


Figure 39. Fixations per trial observed (circles) in the CVC search task, predicted by the Text-Encoding Error (TEER) model (dashed line with squares), and predicted by the FN model (dashed line with circles). The AAE of the TEER model is 14.3% and the FN model is 4.2%.

4.2.3.3 What Was Learned

These results reinforce the findings presented earlier with the local-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. Future research will need to address the possibility of encoding failure rates for non-text stimuli.

4.2.4 Revisiting the Predictions of the Fixation Duration

In the modeling of the local density task, it was found that a strict process-monitoring strategy predicted people's fixation durations well. That is, saccades are initiated as soon as the currently fixated objects are identified.

However, the particular implementation of the strict process-monitoring strategy, the prepare-then-perform strategy, was found to be problematic for two reasons. First, the strategy overpredicts the fixation durations for the CVC task. Second, previous research suggests that the "prepare" part of the prepare-then-perform strategy that was previously implemented as a motor preparation process in EPIC is instead a cognitive process (Kieras, 2003).

4.2.4.1 Process-Monitoring Revisited Model

To address issues identified with the previous implementation of the strict processing 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 process-monitoring strategy largely in two ways: Ocular motor movement preparation is removed from the EPIC architecture and is replaced 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 the execution of eye movements. The motor movement feature preparation times previously included in the model has been attributed to decision processes that are better modeled by selection of saccade destinations in the production rules. 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 based on a set of prioritized rules.

4.2.4.2 Predictions

As shown in Figure 40, 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 seems to predict the users' saccade initiation strategy well. As shown in Figure 41, the model also predicts the observed search time well, AAE = 9.7%.

4.2.4.3 What Was Learned

The strict process-monitoring strategy continues to predict user behavior well, even with a modified implementation of the saccade initiation strategy, a new set of data and different stimuli. While the Original Model predicted the search time better than the

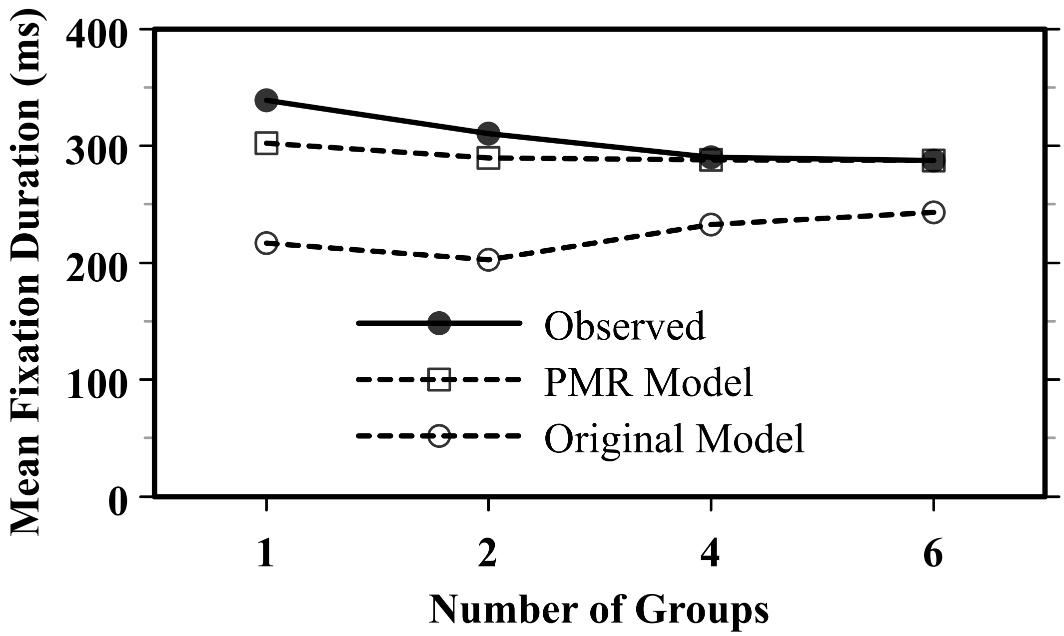


Figure 40. Observed fixation duration (circles) in the CVC search task, predicted by the original model (diamonds), and predicted by the Process-Monitoring Revisited model (squares). The AAE of the original model is 26.5% and the Process-Monitoring Revisited model is 4.6%.

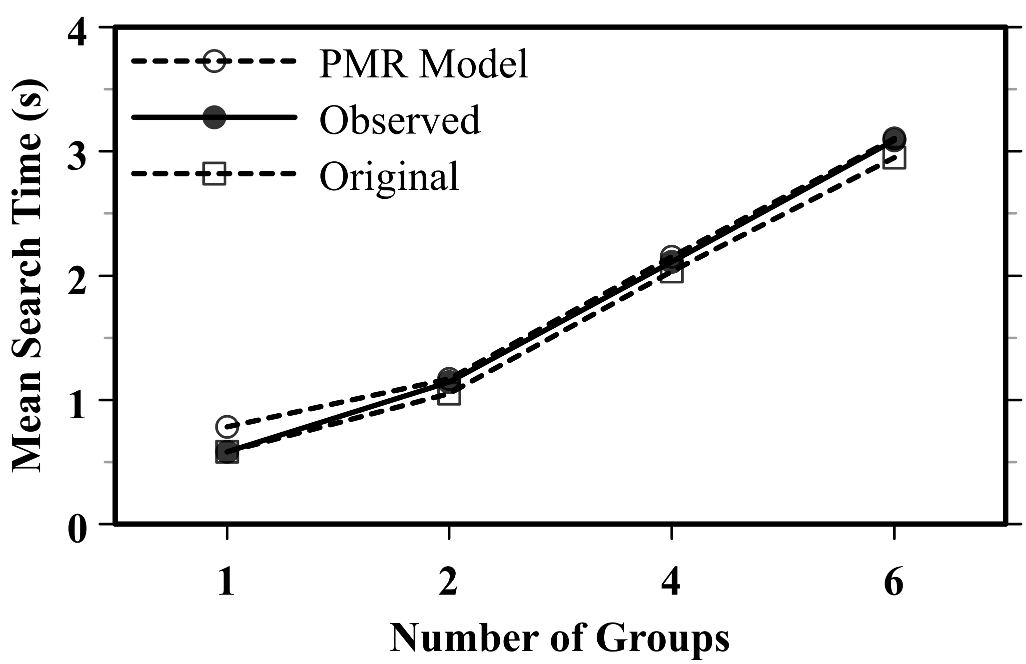


Figure 41. Observed search time (solid line) in the CVC search task, predicted by the original model (dashed line with squares), and predicted by the Process-Monitoring Revisited (PMR) model (dashed line with circles). The AAE of the original model is 4.2% and the Process-Monitoring Revisited model is 9.7%.

active vision model proposed in this thesis, 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 Model.

4.2.5 Discussion

The computational model of visual search proposed by this dissertation 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 strict-processing model to account for saccade durations; (b) text-encoding errors to help account for total fixations; and (c) fixating nearby objects and (d) inhibiting group revisits, both to help account for saccades distances and scanpaths. The model details are motivated by eye movement data and previous research, and can be applied to other modeling research. In the next section, the active vision model is further validated.

4.3 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 of the *a priori* prediction capabilities of the *post hoc* model developed in this research is required. The active vision model that was developed and refined in the research trajectory described in sections 4.1 and 4.2 was next applied to the semantic grouping task discussed in Chapter III. 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. Search time, number of fixations, and fixation duration predictions of the model were compared against human performance for the semantically cohesive and random layouts.

4.3.1 Predictions

As shown in Figures 42, 43, and 44, 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%.

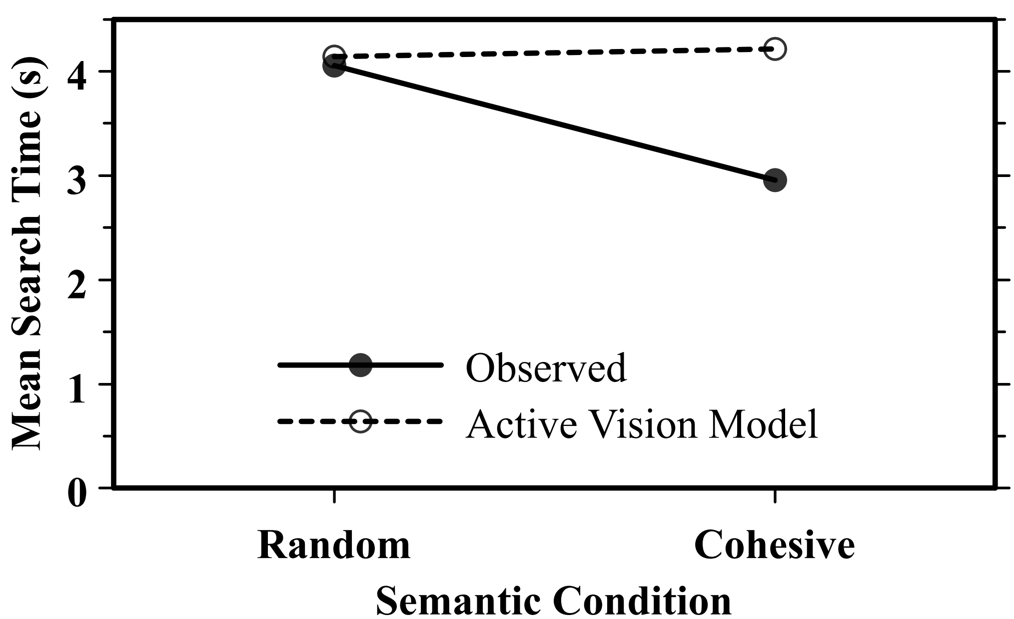


Figure 42. Search time observed in the semantic grouping task (circles), predicted by the Active Vision model (squares). The AAE is 20.7% and for the random layout alone, 6.5%.

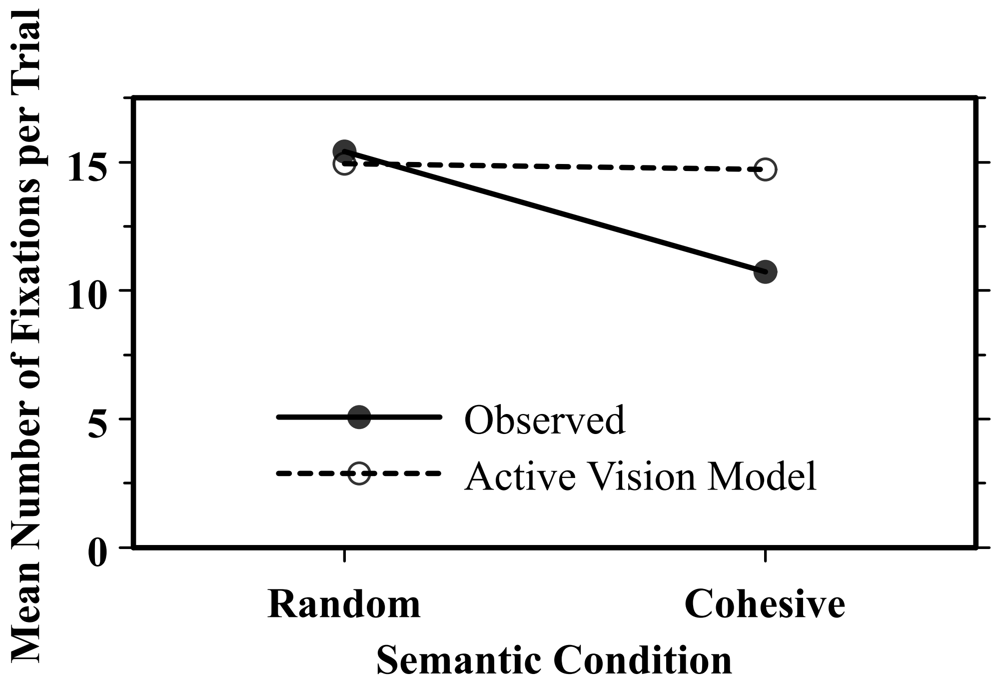


Figure 43. The number of fixations per trial observed in the semantic grouping task (circles), and predicted by the Active Vision model (squares). The overall AAE is 20.2% and for the random layout alone, 3.0%.

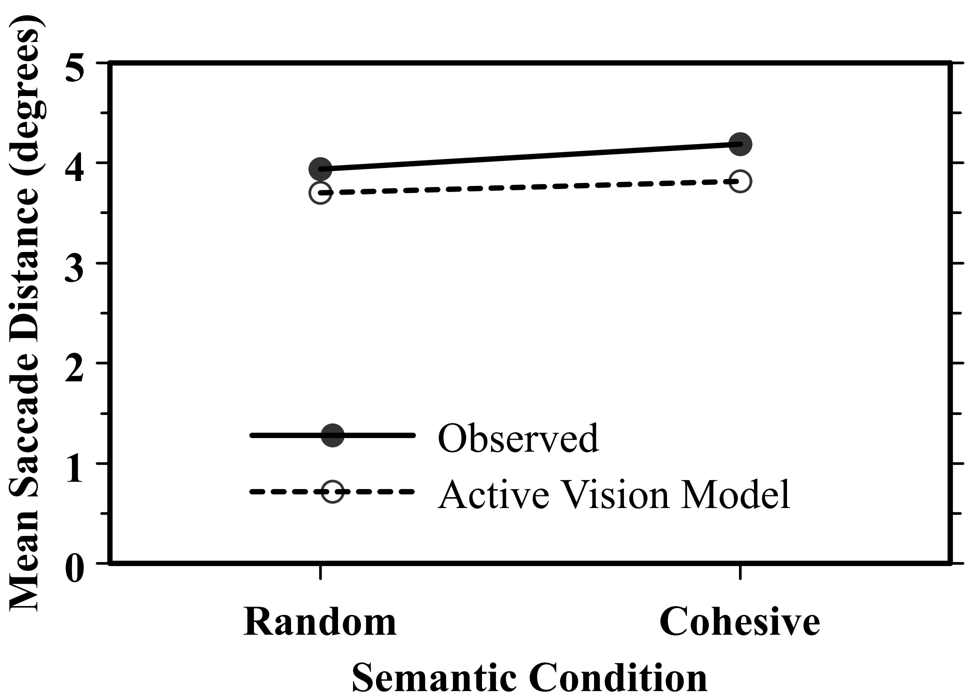


Figure 44. Saccade distance observed in the semantic grouping task (circles), and predicted by the Active Vision model (squares). The AAE is 7.4%.

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 and hence did not utilize semantic information, it was expected that the model would make more fixations in the semantic condition than people who sometimes discounted entire groups of words based on one fixation when the layouts were meaningfully organized.

4.3.2 What Was Learned

The model developed in this dissertation does a good job of predicting visual search performance in tasks slightly different than those it was designed to predict, thus providing some validation for the model. The model predicted human data from the semantic grouping task (for layouts without organized semantic relationships) quite well. The model predicted the search time and eye movement data within our intended 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 some validation for the active vision model proposed in this dissertation. These results suggest that the model would be an appropriate starting place for modeling more complex tasks with more complex stimuli.

The correct predictions and mispredictions made by the model in the semantically-grouped conditions provide guidance for future work. The finding that the saccade distances could be accounted for by the current model suggests that one important aspect of the model, the basis of saccade destination selection, could be utilized for predicting human data from the conditions in which the words were semantically organized. The validations suggests that certain constraints of human information processing are invariant across tasks and that the current model has captured those constraints. The misprediction of the number of fixations required to find the target in the semantically-cohesive condition points to a need for including word-similarity for an even more comprehensive model.

4.4 Summary

The proposed candidate for an active vision model of visual search — and the process of arriving at the model — have implications for representing — and developing — active vision in computational cognitive models. The model of visual search proposed here accounts for a variety of eye movement data, from fixation duration to scanpaths. 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) What can be perceived during a fixation? 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. This research supports the use of text-encoding errors, even for objects very near the center of fixation. (b) When and why are saccades initiated? A strict process-monitoring saccade initiation strategy predicts peoples' fixation durations well. 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 research is able to predict visual search behavior without additional mechanisms or parameters that would be necessary to implement the other saccade initiation strategies. (c) What do the eyes fixate next? 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 predicts people’s eye movement behavior well. (d) What information is integrated between fixations? The memory for the locations previously visited is required between 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 groups 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).

The research reported in this dissertation informs the process of building computational models of visual search in a principled way. The model is (a) based on a variety of eye movement measures, (b) informed by previous research literature on visual search, and (c) 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 discussed in section 4.2 predicted the search time slightly better than the active vision model of visual search proposed in this dissertation.

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 model proposed in this dissertation was informed by a variety of eye movement measurements at every step of the process, which provides more support for the resulting model. The visual search literature provides additional 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 provided 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 reinforced and refined 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 model currently predicts the visual search of text-based displays with an acceptable level of accuracy for engineering based models. An active vision model of visual search based on the research proposed in this dissertation will be useful for automated interface analysis tools. In fact, it has already been demonstrated to be useful for such tools. As evidence of the need for and impact of the research described in this dissertation, some of which has already been disseminated, CogTool-Explorer was recently updated (Teo & John, 2008) to include aspects of the visual search strategies identified by the research reported here. The accuracy with which CogTool-Explorer predicts visual search behavior improved when augmented with principles identified in this research. Future work will be needed to improve on the range of stimuli and task behavior, but the computational model of active vision presented here is already looking to the future.

CHAPTER V

CONCLUSION

This dissertation investigates human-computer visual interaction through experiments and computational cognitive modeling. Three experiments were conducted that investigate the effects of visual layout properties on active vision. Three sets of data (two from the experiments reported here) were accurately modeled in the EPIC cognitive architecture, the results of which extend our understanding of how people visually search computer displays by instantiating a model that addresses the questions put forth by the notion of active vision. Table 4 shows a summary of the experiment and modeling work reported in this dissertation.

5.1 Summary of Empirical Work

The work presented here builds useful theory for human-computer interaction. Three experiments were conducted that further our understanding of how people use active vision to interact with computer displays, specifically text-based layouts. Each experiment investigated the effects of a specific visual design factor. The results from these experiments provide insight for human-computer interaction theory and design practice.

Table 4. A summary of the tasks, observed phenomena, and where the experiments and modeling are reported.

Task	Observed Phenomenon	Experiments	Modeling
Mixed Density	Sparse text searched first and faster	(Halverson & Hornof, 2004c)	(Halverson & Hornof, 2004a)
Link Color	Irrelevantly colored items slow search, especially in the periphery	(Halverson & Hornof, 2004b)	Future Work
CVC Search	Group labels motivate a systematic strategy	(Hornof, 2004)	(Halverson & Hornof, 2006)
Semantic Grouping	Semantically-cohesive groups and group labels allow similar performance but motivate slightly different strategies	(Halverson & Hornof, 2008)	non-semantic layouts only, presented here

The first experiment investigates the effects of varying local density on active vision and finds that people tend to search sparse groups first and faster. Participants move their eyes faster, using shorter fixations, when searching sparse groups relative to dense groups and move their eyes to sparse groups first. Interestingly, when the layouts are of mixed-density, regardless of whether the the group being searched is sparse or dense, the participants initially search in a manner similar to when all-sparse layouts are searched. However, the sparse groups tend to be searched first and when searching dense groups later in a trial, the participants tend to adopted a strategy similar to that used in the all-dense layouts that utilize more and longer fixations. Perhaps users have learned over time that larger fonts, as used in the sparse groups, indicate headlines or headings. Regardless, these findings suggest that designers should use sparse text for important information that the users need to find earlier.

The second experiment investigates the effects of text color on active vision and finds that people occasionally look at irrelevant text identified by color. In comparison to layouts where irrelevantly-colored text is absent, visual search is slower when irrelevantly-colored text is present. Visual search is slowed further as the ratio of irrelevantly-colored to relevantly-colored text increases. Further, the larger the eye movement, the more likely the participants are to look at an irrelevantly-colored item, as is expected with eye movements planned using the reduced resolution and hue sensitivity of peripheral vision. However, even for the very large eye movements, participants are more likely to fixate relevantly-colored text, suggesting a degraded but not absent use of color in the periphery. These findings provide a theoretical and empirical basis for recommendations on the use of link color: First, we have an active vision explanation for precisely how the differentiation of visited and unvisited links can benefit a user; the visited links can largely be ignored as the eyes tend to moved to nearby items where the color is more readily attained. Second, the differently-colored, unvisited links cannot be completely ignored; unvisited links should be removed to improve efficiency if layout consistency is not required for other reasons.

The final experiment investigates the effects of semantic cohesion and group labels on active vision. When groups of words are semantically cohesive, people appear to judge the relevance of semantically grouped words with one fixation, much like is seen when people search layouts in which the groups are labeled. Semantically cohesive or labeled layouts allow people to find the target faster by discounting more objects per fixation than if the layouts are randomly organized. Interestingly, the use of the additional semantic information does not increase the time required to evaluate the objects in each fixation. These active vision findings have direct relevance to HCI as follows: When the space available within an interface is severely restricted (e.g. handheld displays), removing group labels that indicate category will not necessarily put the user at a significant disadvantage. If provided with sufficiently cohesive grouping, users can navigate such layouts as efficiently as if they groups were labeled.

5.2 Summary of Modeling Work

This dissertation presents an active vision model of visual search that accounts for a wide range of eye movement data. Human data from two tasks were used to develop the model and the data from these tasks were accounted for in the modeling. The Text-Encoding Error Model accounted for the local-density task that was presented in the empirical work section. This modeling showed that fixation durations can be explained using a process-monitoring strategy to predict when people move their eyes. Additionally, simulating a small percentage of eye movement patterns that result from misperceived stimuli (i.e. encoding errors) is a useful, straightforward way of explaining what people perceive with each fixation. The Process-Monitoring Revisited Model explained the eye movement data from the CVC search task (Hornof, 2004) with much greater fidelity than had been done previously. The model showed that *where* and *how far* people move their eyes can be explained well by a model that prefers nearby saccade destinations.

The resulting active vision model developed and tuned throughout this thesis was validated by predicting the observed eye movements from the semantic-grouping task presented in the empirical work section. This active vision model was able to predict human performance very well for the conditions in which the words were randomly organized. As expected, the model did not predict the effects of semantic grouping well, as the model does not yet have a notion of semantic relatedness. However, some aspects of the observed search behavior in the semantically organized layouts were predicted well, suggesting that the model will be useful in future modeling efforts that will account for additional factors, like semantics.

The model was incrementally improved based on eye movement analysis and psychological literature. The eye movement data analysis reported in this thesis is more detailed than previously reported in the literature to inform the development of cognitive architecture-based models of visual search (e.g. Fleetwood & Byrne, 2006). The application of the eye movement analysis was further supplemented by established results in the psychological literature, such as hypotheses of saccade initiation (Hooge & Erkelens, 1996), evidence for a constant effective field of view across stimuli density (Bertera & Rayner, 2000), and evidence for saccades tending to be directed to nearby locations (Motter & Belky, 1998).

5.3 Contributions to Cognitive Modeling

This dissertation moves the fields of HCI and cognitive science 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 constraints proposed in this research worked well to predict people's ability to locate a target of visual search. The model addresses the question about what can be perceived during a fixation by showing that 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. It was found that a text-encoding error rate of roughly 10% helps to accurately predict how quickly people can find a target.

This thesis provides support for the use of a strict process-monitoring saccade initiation strategy in computational models of visual search. The modeling is relevant to the issue of when saccades are initiated in that it shows how a relatively straightforward set of assumptions regarding visual information and ocular motor processing, as built into a cognitive architecture, lends itself quite well to explaining, and perhaps thus supporting, a process-monitoring explanation of saccade initiation. This thesis extends existing theory within EPIC to instantiate the process-monitoring strategy. EPIC embraces the notion of transduction time in the various perceptual processors, which is the time for information to move through a processor and become available to later stages. This thesis modifies the instantiation of this theory in EPIC by varying the transduction times of visual properties based on the value of other visual properties or relationship to other visual objects. This dissertation shows how the timing of this transduction can be used to explain the observed fixation durations.

Perhaps most important to the prediction of visual search for applications to human-computer interaction, this dissertation provides a detailed active vision model for explaining *scanpaths*. The modeling supports the idea that proximity is an important factor in predicting where people move their eyes. 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.

This thesis shows one important way in which memory for locations may be integrated between eye movements. The only memory that affects, to any large extent, the performance of the proposed model is the memory for previously fixated locations. This need for memory is restricted to those items currently being searched and those regions (i.e. groups) that have been previously searched.

The modeling in this dissertation is a candidate computational model of active vision. Each of the major questions of active vision proposed by Findlay and Gilchrist (2003) are addressed in this dissertation: *What* can be perceived in a fixation? *When* do the eyes move? *Where* do the eyes move? And, what information is *integrated* between eye movements? Addressing each of these issues resulted in a visual search model that will be useful to further research in predicting and understanding user behavior in HCI.

5.4 Future Directions

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

5.4.1 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 (Newell, 1973), 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" (p.

305). His vision of a unified theory of cognition (UTC) has to some extent been realized in cognitive architectures used to create cognitive models like the one in this dissertation. However, the independence of the models instantiated in the architectures can have a decentralizing effect if there is no unification of the theory embedded in the individual models. Therefore, future work is required to integrate across multiple models, including models from different cognitive architectures. One future extension of the research in this dissertation is to investigate methods for integrating the model proposed here with other disparate models.

5.4.1.1 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, like 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 appropriate theoretical exploration. For another, a wide variety of tasks need to be simulated before a truly comprehensive model can be developed.

An active 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 — 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 straightforward: text can be perceived out to 1º of visual angle. Availability functions are necessary to accurately describe visual search behavior. Kieras' recent research with availability functions has explained a range of results from different visual search experiments. The resulting models support the idea that visual properties, like color and shape, are available according to a quadratic function (an equation involving two or more variables raised to the second power or less) based on the eccentricity from the center of fixation and size of the object. The models Kieras constructed using these quadratic availability functions select saccade destinations at random from the objects available according to the quadratic availability functions used. Figure 45 shows possible quadratic, linear, and constant availability functions as a function of eccentricity only.

Both Kieras's availability functions and the nearest-with-noise strategy proposed in this thesis can be used to explain people's saccade selection behavior in different tasks. Further research is required to determine whether *both* are necessary to predict observed scanpaths in visual search, or how the two methods may be integrated, or whether one strategy subsumes the other. This thesis and other research (Findlay, 1997) has shown that when people are searching for objects differentiated by color, people are more likely to fixate on target-colored object regardless of the distance to the object. However, there is sometimes a preference for nearby objects independent of the identification of the

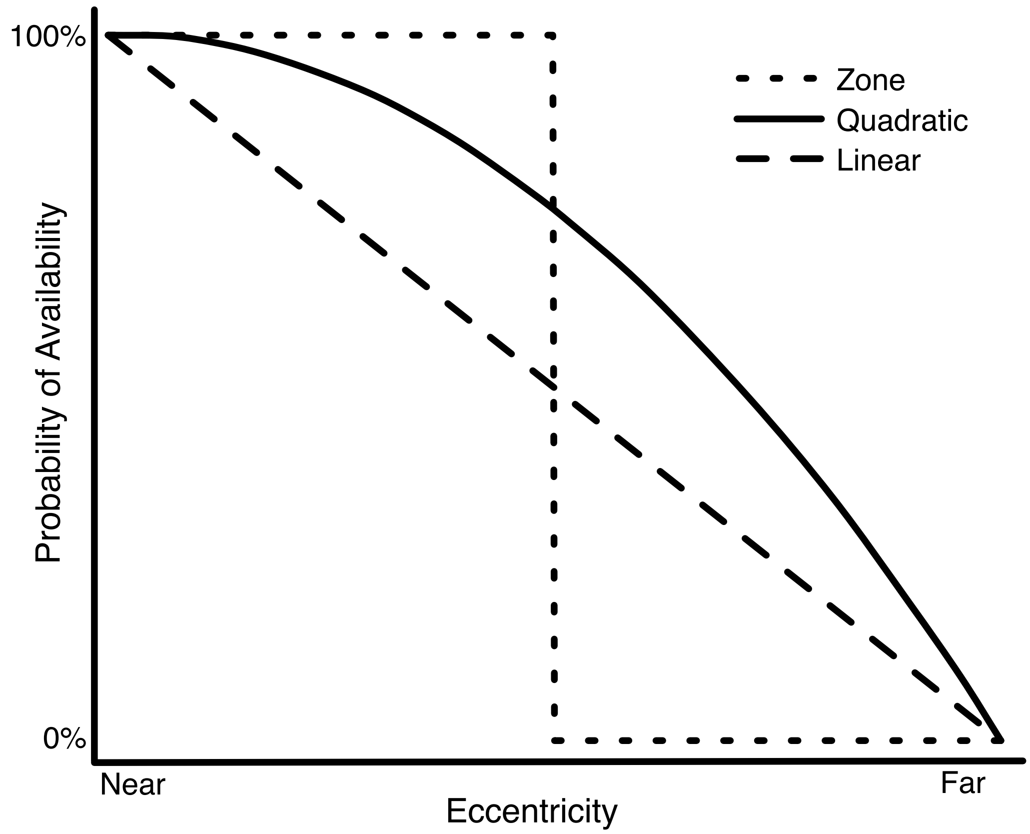


Figure 45. Theoretical plots of availability as a function of eccentricity. As objects move further from the point of gaze, visual features become less available. *Zone* functions are all or nothing. *Quadratic* functions tend to decrease slowly for nearby items and rapidly for distant items. *Linear* functions decrease uniformly.

objects' visual features (ibid.). These complementary findings may support the need for both mechanisms. While differences remain between the modeling presented in this thesis and other models of visual search in EPIC, these differences are reconcilable through integration and additional empirical investigation. Future research will investigate integrating the availability functions proposed by Kieras with the saccade selection strategy proposed in this research. Such integration will be useful for extending the model proposed in this thesis to simulate active vision for a wider variety of tasks.

5.4.1.2 Integration with Models of Semantic Search

The active vision model proposed in this thesis will be improved to account for the effects of semantics on visual 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 by Brumby and Howes (2004; 2008) has provided much insight in to how the semantics can guide visual search. Brumby and Howes (2008) investigated the effect of word meaning on the visual search of menus. Menu items may be semantically related to one another to a lesser or greater degree. Additionally, the target menu item may be semantically related to the search goal to a lesser or greater degree. Both the semantic relationship between menu items and between the target and goal can affect visual search. They found that people tend to search fewer items when distractor menu items are less similar to the goal and when the target is more similar to the goal. Further, people tend to revisit smaller and smaller groups of menu items as visual search progresses before selecting a menu item to click on. A model was constructed to explain the findings in the ACTR cognitive architecture (Brumby & Howes, 2004). The model used an *interdependence of link assessment* search strategy that accounts for the perceived semantic distance between menu items and the target word. The strategy contains three key elements: (a) If the perceived semantic distance between a menu item and the target is close enough, mark the menu item as a potential target. (b) If the perceived semantic distance between the menu item and the goal is even higher, select that menu item as the target. (c) If the currently fixated menu item is not the target, make another eye movement to an unassessed menu item or to a menu item previously marked as a potential target.

It would be useful to integrate the Interdependence of Link Assessment Model with the model proposed in this thesis, as the models have complimentary strengths. The model proposed in this dissertation performs visual searches for exact targets. The semantic content of the text being searched does not influence the model's visual search processes as would be the case for people. The Interdependence of Link Assessment Model has accounted quite well for the influence of the semantics of text on people's visual search processes. Conversely, the Interdependence of Link Assessment model uses an over-simplified scanpath that searches from top-to-bottom in a menu one item at a time. The model presented here does quite well at predicting how people select saccade destinations. The integration of these models would benefit predictive modeling in HCI tasks greatly, as both the location and the content of computer layouts are important factors of screen design.

5.4.2 Informing the Development of Automated Interface Analysis Tools

The aim of all of this research is to provide theoretical underpinnings 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. 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 CogTool: (a) Improve the predictive power of the model presented in this dissertation and (b) enhance CogTool with a more robust model of visual search based on this model. This entire dissertation has focused on improving the model. Regarding the second goal, some progress has already been made, but more is needed. Teo and John (2008) have enhanced CogTool-Explorer (an extension to CogTool) to include some aspects of the research presented here. 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 extremely limited simulations of visual perception. For example, all visual objects on a web page have equal visual saliency regardless of location on the page. CogTool-Explorer needs a greater integration with the current model presented in this dissertation to more accurately simulate human visual-perceptual and ocular-motor processes in order to more accurately predict human visual search performance.

5.5 Conclusion

To better support users and predict their behavior on potential, 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 experimental results presented in this thesis provide a better understanding of how text density, color, and word meaning affect human-computer visual interaction. The computational cognitive modeling that built upon those experimental results 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 by instantiating the theory in a computational model. This instantiation allows us to better understand (a) the effects and interactions of visual search processes and (b) 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.

APPENDIX A

WORDS USED IN EXPERIMENTS 1 AND 2

age	bean	box	cat	cold	dart	dye
aisle	bear	boy	cattle	collar	date	earth
alley	bed	bra	cave	cone	dawn	east
angle	bee	brain	cell	cook	day	edge
ankle	beef	brake	chain	copper	decay	egg
ant	beer	brat	chair	cork	deck	eight
ape	beet	bread	chalk	corn	deep	elbow
apple	beetle	breath	charm	corner	deer	end
arm	bell	breeze	chart	cotton	desk	essay
army	belt	brick	cheat	couch	dial	face
arrow	bet	bridge	cheek	count	diet	faint
art	bill	broil	chin	court	dime	fall
ash	bin	broom	choir	cousin	dinner	fan
atom	birch	brush	cider	cow	dirt	farm
aunt	bird	bubble	cigar	crawl	ditch	fat
author	birth	bump	circle	cream	dive	father
autumn	blade	burn	city	crime	dog	feet
baby	block	burner	clash	cross	doll	felt
back	blouse	bush	clean	crow	dollar	fence
bag	blush	butter	clock	crowd	dome	fight
ball	board	button	cloth	crown	door	figure
band	boat	cable	cloud	crumb	doorway	film
bang	body	cafe	clown	cry	dot	filth
bank	boil	cake	club	cube	down	fire
bar	bone	calf	coach	cup	dozen	fish
bark	book	camp	coal	curb	drain	flag
basin	boot	cape	coast	curler	dream	flame
bass	border	car	coat	curve	dress	flare
bat	boss	card	coffee	cut	drug	flash
bath	bottle	case	coil	dad	drum	flea
battle	bow	cash	coin	daisy	duck	float
beach	bowl	cast	coke	dance	dust	flood

floor	guest	ice	lane	mat	nod	pie
flower	guide	inch	lap	match	noodle	pig
flush	gun	ink	laugh	mate	nose	pile
flute	guy	iron	lawn	meal	note	pill
fog	hair	itch	lawyer	meat	nun	pillow
foil	half	jail	lead	medal	nurse	pin
food	hall	jam	leader	men	oak	pine
foot	ham	jar	leaf	metal	ocean	pint
fork	hammer	jaw	leak	mile	office	pipe
form	hand	jeep	lean	milk	oil	pit
fox	harbor	jelly	leap	mine	organ	plain
frame	hat	jersey	leather	miner	ounce	plane
frog	hate	jet	leg	mink	oven	plate
frown	hawk	jewel	lens	mirror	page	play
fruit	head	job	letter	mist	paint	plug
fun	heap	jog	lever	mold	pair	plum
fur	heart	joke	life	money	pale	poet
gang	heat	joy	lift	moon	palm	point
gas	heel	judge	light	moose	pan	poison
gate	height	juice	lighter	moth	pants	pole
ghost	help	jump	limb	mother	paper	pond
gift	herb	jury	limp	motor	parcel	pony
gin	hero	kettle	line	mouse	park	pool
girdle	highway	key	lion	mouth	party	pope
girl	hill	kick	lip	movie	pass	pork
glass	hobby	kid	liquor	mud	paste	post
glove	hockey	king	load	mug	pea	pot
goal	hoe	kiss	lock	muscle	peach	pound
gold	hog	kitten	locker	nag	pear	pour
golf	hole	knee	long	nail	pearl	powder
gown	home	knife	loop	name	pedal	praise
grape	honey	knight	love	narrow	pen	prayer
graph	hood	knob	lump	neck	penny	prize
grass	hook	knuckle	lung	needle	people	puddle
grave	horn	lad	mail	nerve	pepper	pump
grief	horror	lady	male	nest	person	pup
grip	horse	lake	man	net	pet	puppy
groan	house	lamb	map	news	phone	purse
group	hunt	lamp	maple	nickel	pick	quart
guard	hurt	land	march	night	pickle	queen

race
rail
rain
ramp
rat
rear
rent
rib
rice
riddle
rim
ring
riot
rise
river
road
robber
rock
rod
roll
roof
room
root
rope
rose
rough
round
rubber
rug
ruler
rum
sack
safe
sail
salt
sauce
saucer
saw
scab
scale
scar
scare
school
sea
seam
season
seat
seed
self
sewer
shadow
shallow
shape
shark
shed
sheep
sheet
shell
ship
shirt
shiver
shock
shoe
shoot
shop
shore
shot
shout
shovel
shower
sign
singer
sink
skate
ski
skin
skirt
skull
sky
slap
sleep
sleeve
sleigh
slice
slide
slip
slope
slush
smack
smash
smell
smile
smoke
snail
snake
sneeze
snow
soap
sob
soccer
sock
soda
sofa
soft
soil
song
sore
sound
soup
south
space
spade
spark
spear
spice
spoke
spool
spoon
spray
square
staff
stain
stair
star
state
steak
steam
steel
stem
step
stew
stick
stone
stool
stop
store
storm
stout
stove
straw
sugar
suit
sum
summer
sun
supper
surf
sweat
sweep
sweet
swim
table
tail
talk
tank
tap
tape
tar
tea
team
tear
teeth
tent
terror
test
thaw
thick
thief
thread
three
thrill
throat
throw
thumb
tide
tidy
tie
tiger
tin
tip
tire
toad
toast
toe
ton
tongue
tool
tooth
top
touch
tough
town
toy
track
trail
train
trash
tray
tree
tribe
trick
trip
truck
tube
tune
tunnel
turtle
twig
uncle
valley
van
vein
voice
vote
voter
wage
waist
walk
wall
war
wash
watch
water
wave
wax
wealth
weather
weed
week
weight
well
whale
wheat
wheel
whistle
wide
wife
wig
wild
win
wind
wine
wing
wink
wire
wolf
womb
wood
wool
work
worker
world
worm
wrap
wreck
yard
yawn
youth
zero
zipper
zoo

APPENDIX B

WORDS, CATEGORIES, AND META-GROUPS USED IN EXPERIMENT 3

The words, category labels (italicized), and super categories (bold) used in experiment 3 are listed below. Note that the super categories were not shown and the category labels were not italicized in the experiment's layouts.

animals

farm animals
cow
pig
horse
sheep
goat
lamb
ox
rabbit
bull

wild animals
lion
tiger
bear
elephant
wolf
boar
fox
deer
cheetah
zebra

birds
robin
cardinal
eagle
bluebird
sparrow
parrot
hawk
pigeon
canary
woodpecker

rodents
rat
mouse
squirrel
gerbil
hamster
opossum
chipmunk
bat
beaver
gopher

tropical fish
piranhas
angelfish
blowfish
clownfish
seahorse
barracuda
stingray
starfish
sunfish
swordfish

food1

bread
rye
pumpernickel
sourdough
challah
roll
pita
croissant
bagel

toast
biscuit

dairy
milk
cheese
yogurt
butter
cream
eggnog
buttermilk
cheddar
kefir
feta

meat
beef
pork
steak
veal
hamburger
ham
venison
ribs
salami
roast

vegetables
carrot
lettuce
broccoli
corn
celery
tomato
cucumber
potato
peas
onion

fruits
apple
orange
banana
pear
grape
strawberry
peach
kiwi
mango
pineapple

food2

alcohol
beer
wine
vodka
rum
gin
whiskey
champagne
tequila
liquor
scotch

beverage
water
juice
coke
coffee
tea
lemonade
punch
pepsi
sprite
shake

candy
chocolate
gum
licorice
sucker
mints
caramel
taffy
skittles
jawbreaker
snickers

condiments
salt
pepper
sugar
vanilla
ketchup
lemon
barbeque
mustard
vinegar
tabasco

attire

clothing
shirt
pants
socks
underwear
hat
sweater
jacket
skirt
shorts
jeans

cloth
cotton
silk
polyester
wool
rayon
linen

nylon
satin
denim
cashmere

footwear
sandals
boots
shoes
slippers
sneakers
loafers
moccasins
pumps
clogs
skates

cosmetics
lipstick
blush
mascara
eyeliner
foundation
powder
rouge
perfume
lotion
gloss

jewelry
necklace
ring
bracelet
earring
watch
anklet
brooch
tiara
cufflink
crown

entertainment

dance
ballet
swing
tango
waltz
disco
macarana
mambo
lambada
samba
polka

music
rock
rap
classical
jazz
country
alternative
blues
hiphop
folk
reggae

instruments
piano
flute
drum
saxophone
trumpet
violin
guitar
clarinet
oboe
tuba

singing
soprano
alto
bass
tenor
baritone
falsetto
operatic
contralto
mezzo
countertenor

media

reference
encyclopedia
dictionary
thesaurus
journal
almanac
atlas
textbook
index
phonebook
handbook

reading
magazine
newspaper
pamphlet
novel
brochure
fiction
comic
essay
book
mystery

writing
pen
pencil
marker
crayon

chalk
highlighter
paper
ink
eraser
typewriter

communication
telephone
letter
talk
phone
fax
internet
telegram
mail
radio
pager

buildings

buildings
school
skyscraper
hospital
restaurant
museum
mall
hotel
warehouse
prison
bank

building part
annex
atrium
attic
backdoor
balcony
basement
bathroom
bedroom
ceiling
chimney

religious building
church
temple
synagogue
mosque
cathedral
chapel
shrine
monastery
convent
tabernacle

homes
house
apartment
condominium
hut
dormitory
mansion
shack
igloo
trailer
townhouse

medicine

medical specialty
pediatrics
gynecology
surgical
cardiology
orthopedics
obstetrics
oncology
urology
ophthalmology
dermatology

diseases
cancer
herpes
leukemia
hepatitis
alzheimers
diabetes
tuberculosis
malaria
hiv
parkinsons

organs
heart
liver
lung
kidney
stomach
brain
intestine
pancreas
spleen
uterus

body part
leg
arm
hand
head
foot
toe
finger
neck
shoulder
chest

face part
nose
eyes

mouth
cheeks
ears
lips
eyebrows
forehead
chin
eyelashes

transportation

auto parts

engine
wheel
carburetor
tire
brake
muffler
transmission
battery
radiator
axle

bicycle parts

seat
chain
spokes
pedal
gears
handlebars
frame
horn
reflector
bell

vehicles

car
truck
bus
van
motorcycle
train
airplane
jeep
limousine
moped

boats

sailboat
yacht
battleship
submarine
rowboat
tugboat
cruiseliner
canoe
speedboat
barge

boat parts

sail
mast
bow
stern
deck
hull
rudder
oar
anchor
cabin

furnishings

major appliances

refrigerator
stove
dishwasher
dryer
oven
television
freezer
stereo
furnace
computer

small appliances

toaster
blender
microwave
mixer
juicer
timer
beater
breadmaker
crockpot
coffeemaker

bathroom fixtures

sink
toilet
bath
shower
mirror
light
faucets
cabinet
fan

furniture

chair
couch
table
bed
desk
sofa
dresser
lamp
loveseat
ottoman

chemistry

precious stones
- diamond
- ruby
- emerald
- sapphire
- opal
- amethyst
- pearl
- topaz
- garnet
- jade

metals
- gold
- silver
- steel
- iron
- copper
- aluminum
- platinum
- tin
- lead
- bronze

chemical elements
- oxygen
- hydrogen
- nitrogen
- helium
- carbon
- sulfur
- sodium
- lithium
- potassium
- phosphorus

energy source
- solar
- hydro
- wind
- electricity
- gas
- coal
- oil
- nuclear
- thermal

earth

landscapes
- mountain
- valley
- hill
- volcano
- canyon
- glacier
- plateau
- cave
- cliff
- island

waterways
- river
- ocean
- lake
- stream
- canal
- sea
- pond
- creek
- channel
- brook

trees
- oak
- pine
- maple
- birch
- redwood
- willow
- elm
- cedar
- palm
- spruce

storm
- thunder
- rain
- hail
- hurricane
- tornado
- windstorm
- lightning
- blizzard
- sandstorm
- sleet

titles

relatives
- aunt
- uncle
- cousin
- brother
- sister
- mother
- father
- niece
- grandma
- grandpa

nonrelatives
- friend
- boyfriend
- teacher
- girlfriend
- acquaintance
- boss
- enemy
- neighbor
- mentor
- roommate

military title
general
sergeant
lieutenant
captain
private
colonel
major
corporal
commander
officer

occupation
lawyer
engineer
accountant
banker
professor
psychiatrist
artist
actor
secretary
manager

royalty
king
queen
prince
princess
duke
duchess
jester
lord
knight
lady

tools

surgical instrument
scalpel
needle
tweezer
laser
clamp
forceps
gauze
stethoscope
syringe
suture

garden tools
hoe
shovel
rake
spade
trowel
weeder
lawnmower
rototiller
pots
sprinkler

kitchen utensil
fork
knife
spoon
spatula
ladle
plate
bowl
tongs
cup
whisk

hiking equipment
backpack
rope
tent
compass
canteen
pick
matches
map
trailmix
raincoat

APPENDIX C

PRODUCTION RULES FOR THE FINAL MODEL

The following are the EPIC production rules, non-default parameters, initial memory contents, and well known (i.e. named) locations for the final model discussed in Chapter IV.

```
(Define Parameters

   // The eccentricity fluctuation factor affects
   // saccade destination selection.
   (Eye Eccentricity_fluctuation_factor Normal When_used 1 0.35)

   // Reduce ocular movement preparation cost to zero
   (Ocular Feature_time Uniform When_used 0 0.0)

   // Reduce manual movement preparation cost to zero
   (Manual Feature_time Uniform When_used 0 0.0)


   (Visual_perceptual_processor Text_recoding_failure_rate
      Normal When_used 0.09 0.0)

   (Visual_perceptual_processor Recoding_time Text 100)

   (Eye Availability Shape Zone 49 7.5)
   (Eye Availability Text Zone 49 1.0)

   (Visual_perceptual_store Property_decay_time Normal When_used 50 0)
)

(Define Initial_memory_contents
   (Goal Do Visual_Search Task)
   (Step Pretrial Tag Precues)
   (Tag Cursor Cursor)
)

(Define Named_location Away_from_precue 0.0 0.0)
```

```
///////////////////////////////////////////////////////////////////////
// PRE-SEARCH STRATEGY PRs
// The following rules for response to the precue.
///////////////////////////////////////////////////////////////////////

// Tag all precue objects for examination
(Tag-all-precue-objects-for-examination
IF
(
   (Goal Do Visual_Search Task)
   (Step Pretrial Tag Precues)
   (Visual ?Object Detection Onset)
   (Not (Tag ?Object Cursor))
)
THEN
(
   (Delete (Step Pretrial Tag Precues))
   (Add (Step Pretrial Look_at Any_Precue))
   (Add (Tag ?Object Precue_to_be_examined))
))


// Look at any random, unvisited precue object.
(Look-at-any-precue-object
IF
(
   (Goal Do Visual_Search Task)
   (Step Pretrial Look_at Any_Precue)
   (Tag ?Object Precue_to_be_examined)
   (Visual ?Object Eccentricity ?ecc)
   (Motor Ocular Modality Free)
)
THEN
(
   (Delete (Step Pretrial Look_at Any_Precue))
   (Add (Step Pretrial Memorize Precue))

   (Send_to_motor Ocular Perform Move ?Object)
))


// If looking at the precue, memorize it unless the text is unknown,
// which can happen because of text recoding failures.
(Memorize-and-point-to-precue
IF
(
   (Goal Do Visual_Search Task)
   (Step Pretrial Memorize Precue)

   (Tag ?Object Precue_to_be_examined)
   (Visual ?Object Text ?Text)
```

```
   (Not (Visual ?Object Left_of ?Anything))
   (Not (Visual ?Object Right_of ?Anything))

   // If the text is unknown, this object
   // will be revisited later.
   (Not (Visual ?Object Text Unknown))
)
THEN
(
   (Delete (Step Pretrial Memorize Precue))
   (Add (Step Pretrial Look_at Any_Precue))
   (Delete (Tag ?Object Precue_to_be_examined))

   (Add (Tag Target_Text ?Text))

   (Add (Tag ?Object Precue_Object))

   (Send_to_motor Manual Perform Point ?Object)
))


// If looking at the precue label, memorize it
(Memorize-target-group-label-in-precue
IF
(
   (Goal Do Visual_Search Task)
   (Step Pretrial Memorize Precue)

   (Tag ?Object Precue_to_be_examined)
   (Visual ?Object Text ?Text)
   (Visual ?Object Left_of ?Anything)

   // If the text is unknown, this object
   // will be revisited later.
   (Not (Visual ?Object Text Unknown))
)
THEN
(
   (Delete (Step Pretrial Memorize Precue))
   (Add (Step Pretrial Look_at Any_Precue))
   (Delete (Tag ?Object Precue_to_be_examined))

   (Add (Tag Target_Group_Label_Text ?Text))

   (Add (Tag ?Object Precue_Label_Object))
))


// If looking at a precue object and the text is unknown, look at
// another object.
// Note: Since an objects text will remain unknown until the model
```

```
// looks away from the object, this rule moves the eyes to an arbitrary
// named location so that the precue text will eventually be known.
// This does not affect the model's predictions as this occurs before
// search begins.
(Skip-Unknown-Precue-Object
IF
(
   (Goal Do Visual_Search Task)
   (Step Pretrial Memorize Precue)

   (Motor Ocular Modality Free)

   (Tag ?Object Precue_to_be_examined)
   (Visual ?Object Text Unknown)
)
THEN
(
   (Delete (Step Pretrial Memorize Precue))
   (Add (Step Pretrial Look_at Any_Precue))

   (Send_to_motor Ocular Perform Move Away_from_precue)
))


// The precue stage is over. Proceed with the timed portion of the
// trial. If the eyes are not on the precue, move them back to the
// precue.
(All-precue-objects-are-examined-so-proceed-with-trial
IF
(
   (Goal Do Visual_Search Task)
   (Step Pretrial Look_at Any_Precue)
   (NOT (Tag ??? Precue_to_be_examined))

   (Tag ?Object Precue_Object)

   (Motor Ocular Processor Free)
)
THEN
(
   (Send_to_motor Ocular Perform Move ?Object)

   (Delete (Step Pretrial Look_at Any_Precue))
   (Add (Step Pretrial Click Precue))
))


// Start trial.  Have the model wait long enough for the ocular motor
// processor to be free, to make sure the next rule is not delayed due
// to pre-trial activity.  The subject can wait here anyway as they
// memorize the precue.
```

```
// This does not affect the search time.
(Punch-Mouse-Button-To-Show-Layout
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Pretrial Click Precue)

      (Motor Ocular Processor Free)
      (Motor Manual Processor Free)
   )
   THEN
   (
      (Delete (Step Pretrial Click Precue))
      (Add (Step Move))

      // A location to move the eyes not been selected
      (Add (Tag nothing Selected))
      (Add (Tag nothing Current_Group))

      (Send_to_motor Manual Perform Punch B1 Right Index)
   )
 )


///////////////////////////////////////////////////////////////////////
// VISUAL SEARCH STRATEGY 'DAEMON' RULES
// These are PRs common to all strategies that occur at any time,
// largely for the maintenance of 'tags'
///////////////////////////////////////////////////////////////////////


// Decide to remove the group identified tags from all groups, except
// the currently fixated group.
// Note: This PR will fire every other cycle when there is only one
// group present. However, in that case, this PR  only triggers a rule
// that removes the step added in this PR.   See the PR 'Remove-Reset-
// Group-Step'
(Decide-to-Reset-Group-Identified-Tags
   IF
   (
      (Goal Do Visual_Search Task)

      // If not already resetting the tags
      (Not (Step Reset Identified_tags))

      // and not if there is only one group in the layout
      (Not (Tag Do_Not Retag))

      // If just one "unidentified" group remains
      (Visual ?Unidentified_Group Object_Type Group)
      (Not (Tag ?Unidentified_Group Object_Identified))
```

```
      (If_only_one)
   )
   THEN
   (
      (Add (Step Reset Identified_tags))
   )
)

// Remove all group object_identified tags
(Reset-Group-Identified-Tags
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Reset Identified_tags)

      (Visual ?Object Object_Type Group)
      (Tag ?Object Object_Identified)
   )
   THEN
   (
      (Delete (Tag ?Object Object_Identified))
   )
)


// Remove the step to reset group identified tags
// Note: This step must be separate from the PR that removes the tags
// since there may be only one group in the layout, in which case the
// PR that resets the tags will not fire.
(Remove-Reset-Group-Step
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Reset Identified_tags)
   )
   THEN
   (
      (Delete (Step Reset Identified_tags))
   )
)

// As the model moves to other groups, there is no need to remember
// which objects and labels have been identified in previous groups.
(Reset-Object-Identified-Tags
   IF
   (
      (Goal Do Visual_Search Task)

      // Find groups that have been identified
      (Tag ?Group Object_Identified)
      (Visual ?Group Object_Type Group)
```

```
      // Find identified objects in the identified groups
      (Visual ?Object In_Group ?Group)
      (Tag ?Object Object_Identified)
   )
   THEN
   (
      (Delete (Tag ?Object Object_Identified))
   )
)

// If the text of an object is perceived, mark it as identified if the
// object is in the currently selected group. This rule will only apply
// to group labels and menu item objects, as they are the only objects
// that have both text and are in a group.
(Mark-Objects-as-Identified
   IF
   (
      (Goal Do Visual_Search Task)

      // But not during the pretrial stages
      (Not (Step Pretrial ??? ???))

      // Find the selected group
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object In_Group ?Selected_Group)

      // Tag objects in the selected group
      (Visual ?Object Text ???)
      (Visual ?Object In_Group ?Selected_Group)

      // It is not necessary that the object has
      // not been identified before,
      // but it makes the model easier to debug
      // if this rule only fires for those
      // that have not been identified before.
      (Not (Tag ?Object Object_Identified))
   )
   THEN
   (
      (Add (Tag ?Object Object_Identified))
   )
)

// When a new group is visited, mark the previously visited group as
// "identified"
(Mark-Group-as-Identified
   IF
   (
      (Goal Do Visual_Search Task)
```

```
      // Get the group selected for fixation
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object In_Group ?Selected_Group)

      // Get the "current group"
      (Tag ?Current_Group Current_Group)

      // If the current group and fixated group are different
      (Different ?Current_Group ?Selected_Group)
   )
   THEN
   (
      (Add (Tag ?Current_Group Object_Identified))

      (Delete (Tag ?Current_Group Current_Group))
      (Add (Tag ?Selected_Group Current_Group))
   )
)


///////////////////////////////////////////////////////////////////////
// NOMINATE PRODUCTION RULES
// This is the first step in the of nominate-then-move search strategy.
// Each strategy is represented by a set of PRs.
///////////////////////////////////////////////////////////////////////

// LABELED LAYOUT NOMINATIONS

// Nominate the labels of all groups that have not been identified
// (i.e. fixated), if currently fixating a label, the target group has
// not been found
(Nominate-Labeled-Any-Direction-Unidentified-Labels
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Nominate)

      // If not currently reseting group identified tags
      (Not (Step Reset Identified_tags))

      // Once the eyes have stopped moving
      (Motor Ocular Modality Free)

      // Only if the target group has not been found
      (Not (Tag ??? Target_Group))

      // Nominate all group labels of unvisited groups, except
      // for a group label selected for fixation
      (Visual ?Nominate_Label Object_Type Group_Label)
      (Visual ?Nominate_Label In_Group ?Group)
      (Not (Tag ?Group Object_Identified))
```

```
      (Not (Tag ?Nominate_Label Selected))
   )
   THEN
   (
      (Add (Tag ?Nominate_Label Nominee Label))
   )
)

// Nominate group labels that have not been identified, if currently
// fixating a label and there was an recoding error with the target
// group label.
// Note: The only time other group labels are *not* nominated is when
// the target group label has been correctly identified.
(Nominate-Labeled-Any-Direction-Target-Group-Label-Unknown
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Nominate)

      // If not currently reseting group identified tags
      (Not (Step Reset Identified_tags))

      // Once the eyes have stopped moving
      (Motor Ocular Modality Free)

      // If the group label text was 'unknown'
      (Tag Target_Group_Label Unknown)

      // Nominate all group labels of unvisited groups, except for a
      // group label selected for fixation
      (Visual ?Nominate_Label Object_Type Group_Label)
      (Visual ?Nominate_Label In_Group ?Group)
      (Not (Tag ?Group Object_Identified))
      (Not (Tag ?Nominate_Label Selected))
   )
   THEN
   (
      (Add (Tag ?Nominate_Label Nominee Label))
   )
)

// Nominate all unidentified non-label objects in the currently fixated
// group, if this is a labeled layout, in anticipation of the fixated
// group being the target group.
(Nominate-Labeled-All-Current-Group-Objects-Unidentified
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Nominate)

      // If not currently reseting group identified tags
```

```
      (Not (Step Reset Identified_tags))

      // Once the eyes have stopped moving
      (Motor Ocular Modality Free)

      // Only if this is a labeled layout
      (Visual ??? Object_Type Group_Label)

      // Nominate all unidentified objects in the selected group
      (Visual ?Nominate_Word Object_Type Object)
      (Not (Tag ?Nominate_Word Object_Identified))
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object In_Group ?Selected_Group)
      (Visual ?Nominate_Word In_Group ?Selected_Group)
   )
   THEN
   (
      (Add (Tag ?Nominate_Word Nominee Object))
   )
)


// UNLABELED LAYOUT NOMINATIONS

// Nominate all unidentified menu objects in unidentified groups
(Nominate-Unlabeled-All-Unidentified
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Nominate)

      // If not currently reseting group identified tags
      (Not (Step Reset Identified_tags))

      // Once the eyes have stopped moving
      (Motor Ocular Modality Free)

      // Only if the target text has not been found
      (Not (Tag Target_Object ?Target_Object))

      // Only if this is an unlabeled layout
      (Not (Visual ??? Object_Type Group_Label))

      // Nominate unfixated, unidentified objects in unvisited groups
      (Visual ?Nominate_Word In_Group ?Nominate_Group)
      (Not (Tag ?Nominate_Word Object_Identified))
      (Not (Tag ?Nominate_Group Object_Identified))
      (Not (Visual ?Nominate_Word Text ???))

      // And, the word is not "too close" to where the model was just
      // looking
```

```
      (Greater_than ?Ecc 1.0)
   )
   THEN
   (
      (Add (Tag ?Nominate_Word Nominee Object))
   )
)



// Control PR for the labeled strategy. All rules in the labeled
// strategy fire after the eyes stop moving from the previous move
// step.
(Nominate-Control
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Nominate)

      // If not currently reseting group identified tags
      (Not (Step Reset Identified_tags))

      // Once the eyes have stopped moving
      (Motor Ocular Modality Free)
   )
   THEN
   (
      (Delete (Step Nominate))
      (Add (Step Move))
   )
)


//////////////////////////////////////////////////////////////////////
// MOVE PRODUCTION RULES
// This is the second step in the "search iteration" of nominate-move.
// The move rule to fire is chosen based on what has been nominated.
// If nothing has been nominated, the default move rule applies.
// Note: When the trial starts, move is the first step, not nominate.
//////////////////////////////////////////////////////////////////////

// Select the nearest nominated label if the target label has not been
// found and the text of the currently fixated label was identified
// properly.
(Move-to-Labeled-Strategy-Label-Nominee-Label-Known
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // When the eyes are free
```

```
      (Motor Ocular Processor Free)

      // Only if the target group label is not being fixated
      (Tag Target_Group_Label_Text ?T)
      (Tag ?Selected_Object Selected)
      (Not (Visual ?Selected_Object Text ?T))

      // After the selected text is seen and known
      (Visual ?Selected_Object Text ???)
      (Not (Visual ?Selected_Object Text Unknown))

      // Select the nearest nominated label
      (Tag ?Selected_Word Nominee Label)
      (Not (Visual ?Selected_Word Text ???))
      (Visual ?Selected_Word Eccentricity ?ecc)
      (Least ?ecc)
   )
   THEN
   (
      (Send_to_motor Ocular Perform Move ?Selected_Word)

      (Add (Tag ?Selected_Word Selected))
      (Delete (Tag ?Selected_Object Selected))

      (Delete (Step Move))
      (Add (Step Nominate))
   )
)


// The model "believes" the target group label may have just been
// found. That is, when the label is unknown, the currently fixated
// group is always searched.
(Move-to-Labeled-Strategy-Object-Nominee-Label-Unknown
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // When the eyes are free
      (Motor Ocular Processor Free)

      // After the selected text is seen and unknown
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text Unknown)

      // Select the nearest group object
      (Tag ?Selected_Word Nominee Object)
      (Not (Visual ?Selected_Word Text ???))
      (Visual ?Selected_Word Eccentricity ?ecc)
      (Least ?ecc)
```

```
   )
   THEN
   (
      (Send_to_motor Ocular Perform Move ?Selected_Word)

      (Add (Tag ?Selected_Word Selected))
      (Delete (Tag ?Selected_Object Selected))

      (Delete (Step Move))
      (Add (Step Nominate))
   )
)

// The model "believes" the target group label may have just been
// found. This rule also adds a tag to indicate that the model may be
// incorrect in it's “belief” that the target label has been found.
(Move-Target-Group-Possibly-Found
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // When the eyes are free
      (Motor Ocular Processor Free)

      // After the selected text is seen and unknown
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text Unknown)

      // Only if the model "believes" the unknown label is the target
      // group
      (Tag Unknown Is_Target)

      // Get the "target" group
      (Visual ?Selected_Object In_Group ?Target_Group)

      // The model does not already think the target group has been
      // found
      (Not (Tag ??? Target_Group))
   )
   THEN
   (
      (Add (Tag ?Target_Group Target_Group))

      // Add a tag representing that the model may have some
      // reservation that the correct group label was fixated.
      (Add (Tag Target_Group_Label Unknown))
   )
)

// The target group label was just found
```

```
(Move-to-Nearest-Object-Nominee-Target-Group-Found
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // When the eyes are free
      (Motor Ocular Processor Free)

      // A label is being fixated
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Object_Type Group_Label)

      // If the selected text is the target group label and was not
      // identified in a previous round
      (Not (Tag ??? Target_Group))
      (Tag Target_Group_Label_Text ?T)
      (Visual ?Selected_Object Text ?T)

      // Get the target group
      (Visual ?Selected_Object In_Group ?Target_Group)

      // Select the nearest group object
      (Tag ?Selected_Word Nominee Object)
      (Not (Visual ?Selected_Word Text ???))
      (Visual ?Selected_Word Eccentricity ?ecc)
      (Least ?ecc)
   )
   THEN
   (
      (Send_to_motor Ocular Perform Move ?Selected_Word)

      (Add (Tag ?Selected_Word Selected))
      (Delete (Tag ?Selected_Object Selected))

      (Add (Tag ?Target_Group Target_Group))

      (Delete (Step Move))
      (Add (Step Nominate))
   )
)

// If there are menu object nominees, select the nearest.
(Move-to-Nearest-Object-Nominee
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // When the eyes are free
      (Motor Ocular Processor Free)
```

```
      // Only if the target text has not been found
      (Not (Tag Target_Object ???))
      (Tag Target_Text ?Target_Text)
      (Not (Visual ?Target_Object Text ?Target_Text)
         (Visual ?Target_Object Object_Type Object))

      // After the selected text is seen
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text ???)

      // And there are no label nominations
      (Not (Tag ??? Nominee Label))

      // Select the nearest nominated menu object whose text has not
      // been identified
      (Tag ?Selected_Word Nominee Object)
      (Not (Visual ?Selected_Word Text ???))
      (Visual ?Selected_Word Eccentricity ?Ecc)
      (Least ?Ecc)

      // Get the old selected word
      (Tag ?Old Selected)
   )
   THEN
   (
      (Send_to_motor Ocular Perform Move ?Selected_Word)

      (Add (Tag ?Selected_Word Selected))
      (Delete (Tag ?Old Selected))

      (Delete (Step Move))
      (Add (Step Nominate))
   )
)

// Default movement
// This rule fires only if there are no nominations. If there are no
// nominations, then there are no objects to move the eyes to and the
// target
// has not been found yet. So, restart search by removing all object
// identified tags.
// Note: This will most likely only occur when there is more than one
// group. If there is more than one group, the object identified tags
// for menu objects are "reset" when moving to another group.
(Move-Default
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)
```

```
      // When the eyes are free
      (Motor Ocular Processor Free)

      // Only if the target text has not been found
      (Not (Tag Target_Object ???))
      (Tag Target_Text ?Target_Text)
      (Not (Visual ?Target_Object Text ?Target_Text)
         (Visual ?Target_Object Object_Type Object))

      // After the selected text is seen
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text ???)

      // And there are no nominations
      (Not (Tag ??? Nominee ???))

      (Tag ?object Object_Identified)
   )
   THEN
   (
      (Delete (Tag ?object Object_Identified))

      (Delete (Step Move))
      (Add (Step Nominate))
   )
)

// First movement
// This rule is identical to the default movement rule except that the
// condition for identification of the saccade destination text has
// been replaced by the condition for the selected of "nothing." This
// is because, when the eyes first move, there is no previous saccade
// and therefore no text in the selected saccade destination. This rule
// cannot be consolidated with the default move rule, because with the
// exception of the first eye movement, the text of the saccade
// destination must be identified.
(Move-First
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // When the eyes are free
      (Motor Ocular Processor Free)

      // This is the first eye movement
      (Tag nothing Selected)

      // Select from all objects
      (Visual ?Selectee_Word In_Group ???)
```

```
      (Visual ?Selectee_Word Eccentricity ?Ecc)

      // Except the closest objects
      (Greater_than ?Ecc 1.0)

      // Select the nearest such word
      (Least ?Ecc)

      // Get the old selected word
      (Tag ?Old Selected)
   )
   THEN
   (
      (Send_to_motor Ocular Perform Move ?Selectee_Word)

      (Add (Tag ?Selectee_Word Selected))
      (Delete (Tag ?Old Selected))

      (Delete (Step Move))
      (Add (Step Nominate))
   )
)

// Clean up nominations from the last nomination step
(Move-Clean-Up-Nominations
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // After the selected text is seen
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text ???)

      // For all nominations
      (Tag ?Nominee Nominee ?Nomination_Type)
   )
   THEN
   (
      (Delete (Tag ?Nominee Nominee ?Nomination_Type))
   )
)

// Clean "Unknown is target" tag
(Move-Clean-Up-Unknown_Is_Target
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // After the selected text is seen
```

```
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text ???)

      (Tag Unknown Is_Target)
   )
   THEN
   (
      (Delete (Tag Unknown Is_Target))
   )
)

// Clean up the Target_Group_Label Unknown tag
(Move-Clean-Up-Target_Group_Label-Unknown
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      // After the selected text is seen
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text ???)

      (Tag Target_Group_Label Unknown)
      (Tag ?group Target_Group)
   )
   THEN
   (
      (Delete (Tag Target_Group_Label Unknown))
      (Delete (Tag ?group Target_Group))
   )
)

//////////////////////////////////////////////////////////////////////
// END OF ALL STRATEGIES
// There are a couple of steps at the end that are common to all
// strategies.
//////////////////////////////////////////////////////////////////////

// The target has been found in an unlabeled layout. Move the eyes and
// cursor to the target.
(Move-Gaze-and-Cursor-to-Target-Unlabeled
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      (Motor Ocular Processor Free)
      (Motor Manual Processor Free)

      // The layout is unlabeled
      (Not (Visual ??? Object_Type Group_Label))
```

```
      // After the selected text is seen
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text ???)

      (Tag Target_Text ?Target_Text)
      (Visual ?Target_Object Text ?Target_Text)
      (Visual ?Target_Object Object_Type Object)
   )
   THEN
   (
      // Clean up potential remaining steps
      (Delete (Step Move))
      (Delete (Step Reset Identified_tags))

      (Add (Step Punch Mouse Button))

      (Add (Tag Target_Object ?Target_Object))

      (Send_to_motor Ocular Perform Move ?Target_Object)
      (Send_to_motor Manual Perform Point ?Target_Object)
   )
)

// The target has been found in a labeled layout. Move the eyes and
// cursor to the target.
(Move-Gaze-and-Cursor-to-Target-Labeled
   IF
   (
      (Goal Do Visual_Search Task)
      (Step Move)

      (Motor Ocular Processor Free)
      (Motor Manual Processor Free)

      // Only select from non-label objects' text if the target group
      // has been found
      (Tag ??? Target_Group)

      // After the selected text is seen
      (Tag ?Selected_Object Selected)
      (Visual ?Selected_Object Text ???)

      (Tag Target_Text ?Target_Text)
      (Visual ?Target_Object Text ?Target_Text)
      (Visual ?Target_Object Object_Type Object)
   )
   THEN
   (
      // Clean up any remaining steps
      (Delete (Step Move))
```

```
      (Delete (Step Reset Identified_tags))

      (Add (Step Punch Mouse Button))

      (Add (Tag Target_Object ?Target_Object))

      (Send_to_motor Ocular Perform Move ?Target_Object)
      (Send_to_motor Manual Perform Point ?Target_Object)
   )
)

// Click on the target
(Punch-Mouse-Button-On-Target
   IF
   (
    (Goal Do Visual_Search Task)
    (Step Punch Mouse Button)
    (Motor Manual Processor Free)
    )
   THEN
   (
    (Send_to_motor Manual Perform Punch B1 Right Index)
    (Delete (Step Punch Mouse Button))
    (Add (Step Pretrial Tag Precues))
    (Add (Step CLEANUP))
   )
)


// Clean up whatever needs to be cleaned up after the response
(Cleanup-All-Tags-Except-Cursor
IF
(
   (Goal Do Visual_Search Task)
   (Step CLEANUP)
   (Tag ?X ?Y)
   (NOT (Tag ?X Cursor))
)
THEN
(
   (Delete (Step CLEANUP))
   (Delete (Tag ?X ?Y))
))
```

REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

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.

Baddeley, A. D. (1986). *Working memory*. Oxford: Claredon.

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.

Brawn, P., & Snowden, R. J. (1999). Can one pay attention to a particular color? *Perception & Psychophysics, 61*(5), 860-873.

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,* ,

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 ACM CHI 1982 Conference on Human Factors in Computing Systems,* Gaithersburg, MD, 190-196.

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

Christ, R. E. (1975). Review and analysis of color coding research for visual displays. *Human Factors, 17*(6), 542-570.

Findlay, J. M. (1997). Saccade target selection duration visual search. *Vision Research, 37*(5), 617-631.

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

Halverson, T., & Hornof, A. J. (2004b). Link colors guide a search. *Proceedings of the Conference on Human Factors in Computing Systems,* Vienna, Austria, 1367-1370.

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. (2006). Towards a flexible, reusable model for predicting eye movements during visual search of text. *Proceedings of the Annual Meeting of the Cognitive Science Society,* Vancouver, B.C., Canada,

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

Halverson, T., & Hornof, A. J. (2008). The effects of semantic grouping on visual search. *Proceedings of the Conference on Human Factors in Computing Systems,* Florence, Italy, 3471-3476.

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

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., Centgraf, P., & Prevas, K. (2004). Predictive human performance modeling made easy. *Proceedings of the International Conference on Cognitive Modeling*, Pittsburgh, PA, 455-462.

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.

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.

Kootstra, 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,* Vancouver, B.C., Canada, 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.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(20), 279-281.

Mackworth, N. H. (1976). Stimulus density limits the useful field of view. In R. A. Monty, & J. W. Senders (Eds.), *Eye Movements and Psychological Processes.* (pp. 307-21). Hillsdale, NJ: Lawrence Erlbaum.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics, 2*(11), 547-552.

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

Mullet, K., & Sano, D. (1995). *Designing visual interfaces: Communication oriented techniques*. Englewood Cliffs, NJ: Prentice Hall PTR.

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, MA: Harvard University Press.

Nielsen, J. (2007). Top ten mistakes in web design. *Jakob Nielsen's Alertbox*. Retrieved February 5, 2008, from http://www.useit.com/alertbox/9605.html

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.

Poisson, M. E., & Wilkinson, F. (1992). Distractor ratio and grouping processes in visual conjunction search. *Perception, 21*(1), 21-38.

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

Salvucci, D. D. (2001b). 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,* Palm Beach Gardens, FL, 71-78.

Shen, J., Reingold, E. M., & Pomplun, M. (2000). Distractor ratio influences patterns of eye movements during visual search. *Perception, 29*(2), 241-250.

Shen, J., Reingold, E. M., & Pomplun, M. (2003). Guidance of eye movements during conjunctive visual search: The distractor-ratio effect. *Canadian Journal of Experimental Psychology, 57*(2), 76-96.

Shih, S. I., & Sperling, G. (1996). Is there feature-based attentional selection in visual search? *Journal of Experimental Psychology: Human Perception & Performance, 22*(3), 758-779.

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.

Spool, J. M., Scanlong, T., Schroeder, W., Snyder, C., & DeAngelo, T. (1999). *Web site usability: A designer's guide*. San Francisco, CA: Morgan Kaufman Publishers.

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 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.

Treisman, A. (1998). The perception of features and objects. In R. D. Wright (Ed.), *Visual Attention.* (pp. 26-54). New York: Oxford University Press.

U.S. Department of Health and Human Services (2006). Current research-based guidelines on web design and usability issues. *Research-Based Web Design & Usability Guidelines*. Retrieved February 5, 2008, from http://www.usability.gov/pdfs/guidelines.html

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505.

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine usable dictionary, version 2. *Behavior Research Methods, Instruments, and Computers, 20*, 6-11.

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: 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.

Yoon, C., Feinberg, F., Hu, P., Gutchess, A. H., Hedden, T., Chen, H. M., et al. (2004). Category norms as a function of culture and age: Comparisons of item responses to 105 categories by american and chinese adults. *Psychology and Aging, 19*(3), 379-393.

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? *Perception, 18*(2), 191-200.