Predicting Cognitive Strategies and Eye Movements in Hierarchical Visual Search

Anthony J. Hornof (hornof@cs.uoregon.edu)
Tim Halverson (thalvers@cs.uoregon.edu)
Department of Computer and Information Science
1202 University of Oregon
Eugene, OR 97403-1202 USA

Abstract

This article advances computational cognitive modeling of visual search, and the synergistic relationship between cognitive modeling and eye tracking. The paper presents cognitive models of the perceptual, cognitive, and motor processing involved in the visual search of a hierarchical layout. Two types of visual layouts are searched: *unlabeled layouts* in which words are arranged in groups but with no hierarchical organization, and *labeled layouts* in which each group is given a heading that guides the search. The two types of layouts motivate fundamentally different search strategies. The models are *post hoc* explanatory models of the search time data and *a priori* predictive models of the eye movement data. The models are evaluated here based on the eye movement data. The research demonstrates a methodology and provides guidance for predictive cognitive modeling of visual search.

Introduction

Cognitive modeling is useful to the field of human-computer interaction because it reveals patterns of human performance at a level of detail not otherwise available to analysts and designers (as in Gray, John & Atwood, 1993). The ultimate promise for cognitive modeling in human-computer interaction is that it provides the science base needed for predictive analysis tools and methodologies (Card, Moran & Newell, 1983). This article reveals patterns of human performance in visual search, and contributes to predictive analysis of visual search.

We recognize that cognitive modeling occurs in two distinct modes: (1) *explanatory* and (2) *predictive*. In the explanatory (or *exploratory*) mode, models are constructed to explain empirical data that have already been collected and analyzed. In the predictive mode, models are constructed to make *a priori* predictions of user performance; that is, predictions before human data has been collected and analyzed. Predictive models can reused in an exploratory mode when they are modified to provide a better fit with observed data. Note that in both modes the output from the model is referred to as a "prediction."

In this article, *post hoc* explanatory models of search time data are used to make *a priori* predictions of newly collected eye movement data. Based on what is learned here, the original models can now be updated and improved.

The work is presented in chronological order: The experiment was designed. Search times were observed. Models were built. Eye movements were observed. The models were evaluated based on this new data.

Eye tracking and cognitive modeling have much to offer each other, especially when eye tracking is used to identify the cognitive strategy used for a task (as in Salvucci & Anderson, 2001). This article further develops the synergy between eye tracking and cognitive modeling.

The Visual Search Experiment

The visual task studied here is finding a known target in a hierarchically-organized visual layout. Layout items are grouped, and sometimes the groups have useful headings. The task is somewhat analogous to looking for a piece of information on a web page or a product brochure, which is sometimes organized in a useful manner with groups and group headings, and sometimes arranged with no clear and useful organization. The task is specifically designed to reveal the core strategic components involved in a hierarchical search.

Experimental Procedure

Figure 1 shows a sample layout from the experiment. The layout has six groups of items, and each group is "labeled" with a heading of X*n*X, where *n* is a single numerical digit. In the figure, the groups are annotated with the letters A through F, though these letters did not appear in the experiment.

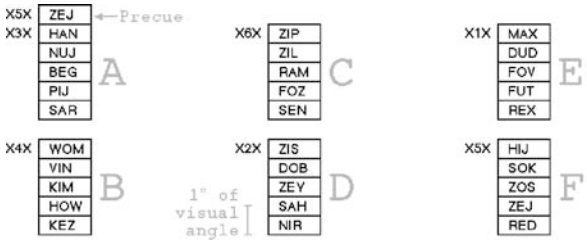


Figure 1. A "6-group labeled" layout. The precue, in the top left, would have disappeared when the layout appeared. The target is in group F. The gray text did not appear during the experiment.

Participants searched eight different screen layouts for a precued target object. Each layout contained one, two, four, or six groups. Each group contained five objects. One-group layouts used group A. Two-group layouts used groups A and B. Four-group layouts used groups A through D. The groups always appeared at the same physical locations on the screen. In each trial, the entire layout was displayed at the same moment, permitting any search order.

Layouts were either labeled or unlabeled. In unlabeled layouts, the X*n*X group labels did not appear. Each unique layout (such as "6-group labeled") was presented in a separate

block of trials.

Target and distractor items were three-letter words or pseudo-words, randomly selected for each trial. Group labels were randomly reordered for each trial. The target position was randomly selected for each trial.

Participants were precued with the target object and, for labeled layouts, the label of the group that would contain the target.

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.

Sixteen experienced computer users with no visual impairments completed the experiment. Search time was separated from mouse movement time by using a point-completion deadline (Hornof, 2001).

Eye Tracking Procedure

Eye movements were recorded using the LC Technologies Eyegaze System, a 60 Hz eye tracker that tracks eye movements using the pupil-center and corneal-reflection. A chinrest maintained an eye-to-screen distance of 56 cm, such that 1° of visual angle subtended 38.4 pixels. The precue always appeared at the same location, level with the participant's eyes.

A dispersion-based algorithm determined the center of fixations, using a minimum fixation duration of 100 ms and a deviation threshold of 0.5° of visual angle. Systematic error in the eye tracking data was reduced *post hoc* using "required fixation locations" (Hornof & Halverson, 2002).[1]

Observed Search Times

Figure 2 shows the search times observed when the experiment was run a second time, with eye tracking. There were no meaningful differences in the search time data between the two runs, though the models fit the data from the first run slightly better.

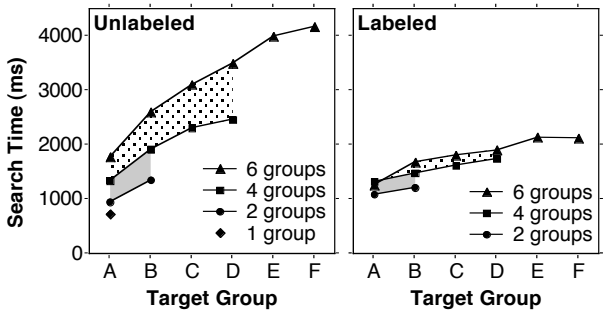


Figure 2. Mean search time for all sixteen participants as a function of the target group, for unlabeled (left) and labeled layouts (right). The shaded area shows the number-of-groups effect.

The three most salient trends in the search time data are: (a) Smaller layouts were faster. (b) Labeled layouts were faster. (c) Unlabeled layouts had a larger number-of-groups effect. The number-of-groups effect is measured in the distance between adjacent curves in a graph, and is shaded in the figure. The effect measures how much longer it takes to find an item in the same group as the layout gets bigger, and suggests an element of noise or randomness in the search process (Hornof, 2001).

[1] The experimental software and eye tracking data are downloadable at http://www.cs.uoregon.edu/~hornof.

Description of the Models

A number of computational cognitive models were built, using the EPIC cognitive architecture (Executive Process-Interactive Control, (Kieras & Meyer, 1997). EPIC captures human perceptual, cognitive, and motor processing constraints in a computational framework that is used to build simulations of human information processing and task execution. EPIC constrains the models that can be built, and the predictions that can be made, based on fundamental human processing capabilities and limitations.

As is required to use the architecture, we encoded into EPIC (a) a reproduction of the task environment, (b) the visual-perceptual features associated with each of the screen objects and (c) the cognitive strategies that guide the visual search. These components were specified based on task analysis, human performance capabilities, previous visual search models, and parsimony. The models are discussed in more detail in Hornof (2002).

Visual-Perceptual Features

Simple visual-perceptual features were used. It should be relatively straightforward to derive most of the features (location, relative position to other objects, size, and text) directly from the interface using an automated screen parser such as VisMap (St.Amant & Riedl, 2001). All feature values were set *a priori,* and the same values were used across all models. This approach emphasizes strategy over object features. Other approaches to modeling visual search emphasize object features (such as Anderson, Matessa & Lebiere, 1998; Fleetwood & Byrne, 2001).

The two visual features that we encoded with specific task knowledge in mind include object-type, which represents whether a screen object is a precue item, layout item, or group label, and next-group, which determines the global search order.

Cognitive Strategies

About eight different strategies were written to examine how people searched unlabeled and labeled layouts. Each strategy was encoded into EPIC, which executed the strategy and generated predictions that were compared to the observed data. Two strategies that provide a good fit with the search time data are described here.

Noisy-Systematic Search

The noisy-systematic search strategy for unlabeled layouts assumes that people attempt to make a "maximally-efficient foveal sweep" (Hornof & Kieras, 1997), in which the eyes move to capture everything in the high resolution foveal vision, which is roughly 2° of visual angle in diameter, with as few fixations as possible.

Noise is introduced into the strategy by having it sometimes overestimate how far the eyes can move and still

foveate everything with successive fixations. If the target is missed, another sweep will be required, substantially increasing the search time for that trial.

To vary the noise in the strategy, it was run with eighty-four different fixation distributions. In the model used here, the first fixation is on the first or second item in group A. Subsequent fixations are made to a randomly chosen item 3 to 7 items "down."

The "down" direction assumes people searched down the first column, down the second, down the third, back to the first. This order attempts to maximize the foveal coverage with as few eye movements as possible, and corresponds to the slope in the search time data. This search order is encoded into the next-group feature.

Mostly-Systematic Two-Tiered Search

The mostly-systematic two-tiered search strategy for labeled layouts assumes that people search the group labels until they find the target group, and then confine their search within that group. The strategy was based on task analysis and the significantly faster search times for labeled layouts. It is "mostly" systematic because it searches the labels in next-group order 75% of the time, and in random order 25% of the time.

Predicted Search Times

Figure 3 shows the search time predictions. The models predict unlabeled layout search time with an average absolute error (AAE) of 8%, and labeled layout search time with an AAE of 6%.

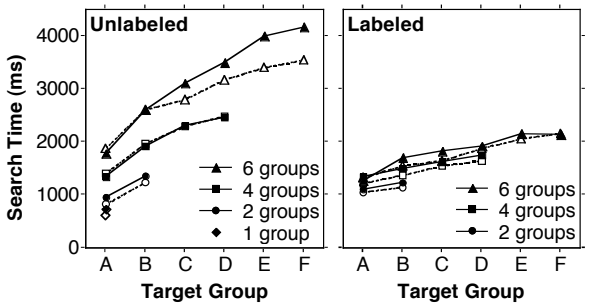


Figure 3. Search times observed (solid lines) and predicted (dashed lines) by the noisy-systematic model for unlabeled layouts (left) and the mostly-systematic two-tiered search model for labeled layouts (right).

The models demonstrate that different layouts will motivate different search strategies. Predictive visual layout analysis tools will need to incorporate different cognitive strategies for different layouts. The two strategies presented here could be used in such a tool.

Predicted and Observed Eye Movements

The *a priori* predicted and the observed eye movements will now be compared. Figure 5 shows the predicted and observed eye movements from one trial with an unlabeled layout, and from one trial with a labeled layout. The figure gives an idea of the similarities and differences between (a) the predicted and the observed and (b) unlabeled search and labeled search. Table 1 summarizes comparisons between the predicted and observed eye movements which will be elaborated in this section, starting with patterns that persisted across all layouts, not just unlabeled and labeled.

Table 1. A summary of the predicted and observed eye movements. Pluses indicate correct predictions.

Eye Movements	Predicted	Observed
Across All Layouts		
Fixations per trial (+)	7.9	7.4
Fixation duration (+)	228 ms	264 ms
Number of scan paths	One	Many
Anticipatory fixations (+)	Yes	Yes
Respond to layout onset (+)	Yes	Yes
Ignore white space (+)	Yes	Yes
Ignore shape (+)	Yes	Yes
Overshoot the target	Yes	Rarely
For Unlabeled Layouts		
Fixations per group	1.1	2.1
Groups revisited per trial	4.4	0.7
Items examined per fixation (+)	2.6	2.4
For Labeled Layouts		
Use group labels (+)	Yes	Yes
Groups revisited per trial	1.2	0.29

All Layouts

Fixations Per Trial. As can be seen in Figure 4, the models and the participants made a similar number of fixations per trial. The model overestimates the number of fixations per trial for unlabeled layouts, with an AAE of 18.0%. The model predicts an additional 1.1 fixations per trial, perhaps due to overshooting the target. If 1.1 fixations are removed from each trial, the AAE drops to 5.4%. The model accurately predicts the number of fixations per trial for labeled layouts, with an AAE of 5.1%.

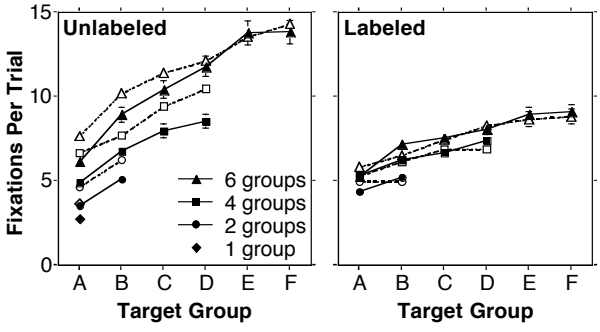


Figure 4. The average number of fixations predicted (dashed lines) and observed (solid lines) for each trial, as a function of the target group.

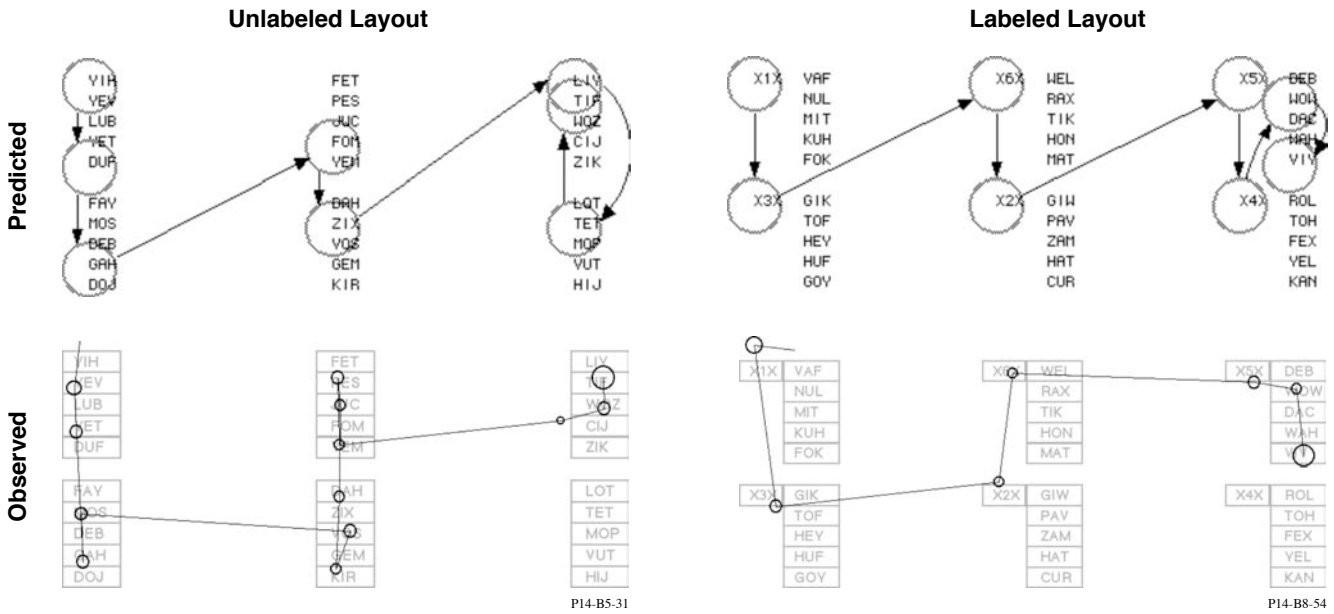
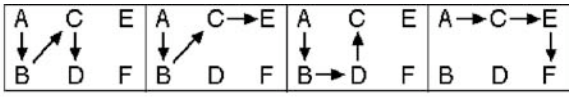


Figure 5. Fixations predicted (top) and observed (bottom) from one trial with an unlabeled layout (left) and one trial with a labeled layout (right). In the predicted, the circles represents the foveal region. In the observed, the diameters of the circles represent the fixation duration. The unlabeled layout fixations are predicted by the noisy-systematic strategy. The labeled layout fixations are predicted by the two-tiered systematic strategy.

Fixation Duration. When searching for the target, the models average one fixation every 228 ms (*SD* = 64). The average fixation duration observed when participants searched the layouts was 264 ms (*SD* = 146). The average fixation duration was a little longer for unlabeled layouts (283 ms, *SD*=152) than for labeled layouts (238 ms, *SD*=134).

Scan Paths. Figure 6 shows the most common orders in which the models and the participants searched the layouts. The models search the groups in the same order for most trials. Participants searched in many different orders. They started in group A but then followed numerous different paths.



Predicted				
Unlabeled:	**70%**	**12%**	0%	1%
Labeled:	**75%**	0%	0%	0%
Observed				
Unlabeled:	**17%**	5%	**30%**	**12%**
Labeled:	6%	7%	**19%**	**12%**

Figure 6. The predicted and observed order in which groups were searched when starting on a six group-layout. The percentages indicate how often each path was taken. Paths over 10% are in bold.

Anticipatory Fixations. The models predict anticipatory fixations, which are eye movements from the precue to the layout before the layout appears. Hornof and Kieras (1999) demonstrate that people make such eye movements. Participants exhibited anticipatory fixations. In 48% of the observed trials, a fixation started within 100 ms (before or after) the onset of the layout, before an eye movement could be prepared in response to the stimuli. The destination of these fixations is more regular for unlabeled layouts, typically to the second or third item in the layout.

Respond to Layout Onset. The models predict that an eye movement will occur in response to the layout onset. This fixation starts, on average, 287 ms after the layout appeared (*SD* = 28). Participants appeared to respond similarly, starting a fixation an average of 235 ms (*SD* = 117) after onset of the layout, which is roughly the time required to respond to a visual stimuli.

Ignore White Space. In the models, all fixations are to screen objects. No fixations land on the white space between the objects. Similarly, participants rarely fixated the white space. Ninety-nine percent of all fixations were within 1° of visual angle of a screen object.

Ignore Shape. The models move the eyes based on the physical structure of the layout and do not prefer items shaped like the target. These predictions build on menu models that explain search time data without considering the shape of menu items (Hornof & Kieras, 1997; Hornof & Kieras, 1999). Other menu models rely on the shape of menu items when shifting attention (Anderson et al., 1998; Byrne, 2001). Participants were minimally influenced by the shape of items. Items that had one or two letters-in-position in common with the target were only 5.7% more likely to receive a fixation than items with no letters-in-position in

common with the target.

Target Overshoot. The models tend to overshoot the target: They foveate the target, continue searching with one eye movement, and then move the eyes back to the target. The overshoot results from timing characteristics of the EPIC architecture and is not specific to these models. Participants rarely overshot the target. They did so on 6.9% of the trials.

Unlabeled Layouts

Fixations Per Group. While searching through an unlabeled layout, the model averages 1.1 fixations per group (*SD* = 0.4) up until the final group, and then 1.3 fixations in the target group (*SD* = 0.7). Participants tended to stay in a group longer, averaging 2.1 fixations per group (*SD* = 0.9) up until the target group was reached, and then 2.4 fixations within the target group (*SD* = 1.0).

Groups Revisits. When searching an unlabeled layout, the model revisits an already-examined group an average of 4.4 times per trial (*SD* = 6.5). Thirty-nine percent of the time the model moved to a group, it was a revisit. Participants required fewer revisits. Participants averaged 0.7 revisits per trial (*SD* = 1.9), and typically found the target on the first visit. Only 16% of the time that a participant moved to a group was it a revisit.

Items Examined Per Fixation. When searching the unlabeled layouts, the noisy-systematic model examines two or three items with each fixation because (a) two or three items land in EPIC's fovea simultaneously and (b) in EPIC, the text of all foveated items moves to visual working memory in parallel. One of the most interesting confirmations of the model provided by the eye tracking data is that participants also examined two or three items with each fixation. This is derived from the fact that participants averaged 2.1 fixations per group, there are five items per group (5 ÷ 2.1 = 2.4), and participants typically found the target on the first visit to a group.

Labeled Layouts

Use the Group Labels. For labeled layouts, the model searches the group labels until the target group is found, and then searches within the target group. This two-tiered search is the primary difference between the strategies for unlabeled layouts and for labeled layouts. The eye tracking data clearly demonstrate that participants used a two-tiered search strategy for labeled layouts. The strategy is evident when examining the eye movement data superimposed on the stimuli (as in Figure 5) and in that, up until the target group was reached, 64% of all groups were visited with a single fixation, and 80% of all fixations were recorded within 1° of visual angle of a group label.

Group Revisits. When searching labeled layouts, the models averaged 1.2 revisits per trial (*SD* = 1.5), roughly one extra revisit per trial. Participants used many fewer, usually finding the target group with a single pass of the group labels, averaging only 0.29 revisits per trial (*SD* = 0.7). Perhaps the model predicted more revisits because the *target overshoot* also occurred while searching group labels; after the target group label was found, the eyes typically continued to the next group, and then returned directly to search in the target group.

Discussion

The eye movement data confirm many aspects of the cognitive strategies and the visual-perceptual and oculomotor processing built into the models. The models accurately predict that a useful visual hierarchy motivates a two-tiered search, that multiple items are examined with a single fixation, and that the search strategy for this task ignores shape. The models accurately predicts initial fixations, and the timing and numerosity of fixations.

The eye movement data also reveal aspects of the models that can be improved. These *a priori* predictive models of eye movements can be reused in an explanatory mode, and rebuilt based on the following lessons learned for predictive cognitive modeling of visual search.

Lesson #1: Noise enters the process at several different levels. The models introduce one major element of noise--randomly skipped over and missing items while searching, which lead to revisits. This behavior contributes to accurate predictions of fixations-per-trial and search times, but poor predictions of fixations-per-group and revisits. There were more sources of noise in the human data. It was common for participants to make one, two or three fixations per group, whereas the models typically made just one. Additional fixations drove up the search time. Additional noise increased the number-of-groups effect. It remains to be seen what sources of noise will need to be included in predictive models.

Lesson #2: Search strategies are partially precompiled and partially filled in during execution. It is very interesting to see that participants consistently used the group labels in labeled layouts--a precompiled *global* strategic decision made before starting the search--and yet took many different paths through a layout, even from trial to trial--revealing a least-commitment, flexible, *local* strategic decision made during the search. The global search order imposed by the next-group feature in the models is wrong, and should perhaps be replaced by heuristics such as in the any-nearest production used in some menu models to move the eyes to any object near the current fixation (Byrne, 2001). However, even in the flexible planning of the search path, a high-level control maintained some order, avoiding paths that would lead to a long jump between the first and third columns.

Lesson #3: Cognitive architectures need a tight coupling between visual-perceptual and oculomotor processing. EPIC may need a faster interaction between visual-perception and oculomotor processing so that the architecture does not overshoot the target when running the strategies discussed here. This is a good result. The modeling has informed the development of the architecture.

Conclusion

This article presents computational cognitive models that predict the eye movements that people will make when searching a hierarchical visual layout. The predictions were evaluated with observed eye movements. All told, the models and the observed data provide a very interesting

explanation of how people conduct a hierarchical visual search, many ideas for how to improve these and future predictive models of visual search, and suggestions for improving cognitive architectures.

This research contributes to the synergistic relationship between cognitive modeling and eye tracking: Eye tracking data are best-understood in the context of models that simulate visual perception and oculomotor processing, and models of these processes can be improved with detailed analysis of eye tracking data.

Acknowledgments

The authors would like to thank Ronald Chong and Robert Wray for their feedback on this paper. This work was supported by the Office of Naval Research through Grant N00014-02-10440 to the University of Oregon, Anthony Hornof, principal investigator.

References

Anderson, J. R., Matessa, M., & Lebiere, C. (1998). The visual interface. In J. R. Anderson & C. Lebiere (Eds.), *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum, 143-168.

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., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fleetwood, M. D., & Byrne, M. D. (2001). Modeling Icon Search in ACT-R/PM. *Fourth International Conference on Cognitive Modeling*.

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

Hornof, A. (2002). *Cognitive strategies for the visual search of hierarchical computer displays*. Department of CIS Technical Report 02-04, University of Oregon, www.cs.uoregon.edu/~hornof/downloads/Hierarchical.pdf.

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., & 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., & Kieras, D. E. (1997). Cognitive modeling reveals menu search is both random and systematic. *Proceedings of ACM CHI 97: Conference on Human Factors in Computing Systems,* New York: ACM, 107-114.

Hornof, A. J., & Kieras, D. E. (1999). Cognitive modeling demonstrates how people use anticipated location knowledge of menu items. *Proceedings of ACM CHI 99: Conference on Human Factors in Computing Systems,* New York: ACM, 410-417.

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.

Salvucci, D. D., & Anderson, J. R. (2001). Automated eye-movement protocol analysis. *Human-Computer Interaction*, 16, 39-86.

St. Amant, R., & Riedl, M. O. (2001). A perception/action substrate for cognitive modeling and HCI. *International Journal of Human-Computer Studies*, 55, 15-39.