Towards a Flexible, Reusable Model for Predicting Eye Movements During Visual Search of Text

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

Abstract

Visual search is an integral component in many human activities. The eye movements produced during such activities can provide valuable information about people's cognitive processes. This research investigates, with detailed eye movement data analysis and computational cognitive modeling, the perceptual, strategic, and oculomotor processes people use to visually search. A cognitive model is evolved in a principled manner based on eye movement data, past modeling efforts, and recent psychological literature. In the model, re-usable, parsimonious, local strategies interact with perceptual-motor constraints to predict the bulk of the eye movement data, including aspects of the data that appear to require task-specific global strategies in addition to fixation-to-fixation local strategies. The analysts evolve a base level model with a random strategy into a robust and reusable model with a flexible strategy that could work with a wide range of visual stimuli.

Keywords: cognitive modeling; visual search; EPIC; eye movements

Introduction

The visual search strategies people employ have a substantial effect on the time it takes people to find a target in a visual layout. A fair amount of research has been done on visual search strategies people use. For example, Shen, Reingold, and Pomplun (2003) found that people tend to shift their visual search strategy very quickly based on which visual feature is most informative for a given layout. Burke, et al. (2005) found that people ignore the most salient objects that do not relate to the task, flashing banner advertisements, in simulated web pages.

One way to better understand what visual search strategies people use, and why they use them, is through computational cognitive modeling. The models instantiate the theory, make testable numeric predictions, and facilitate identification of unanswered questions. Several computational models of visual search have been proposed (e.g. Pomplun, Reingold, & Shen, 2003; Wolfe, 1994). For the most part, these computational models of visual search account for one or two of the perceptual, strategic, or oculomotor processes involved in visual search, but not all three. Ideally, a model of visual search would explain some aspect of each process involved in visual search.

This research proposes a flexible and reusable computational cognitive model of text search that builds directly on a number of previous studies of structured, menu-like visual layouts. The purpose of this modeling effort is to further clarify and build a framework for understanding (scientifically) and predicting (scientifically and for design purposes) how people integrate perceptual, strategic and motor processes in visual search. This paper describes the evolution of a visual search model from a constrained, random search strategy into a robust and flexible model of menu search that accounts for a wide variety of eye movement data. We believe the resulting model, while developed using data from one task, has been evolved by the analysts with sufficiently few task-specific requisites. That is, the model is flexible and reusable.

Building on Previous Visual Search Models

This research builds directly on previous research of menu search. Hornof (2004) studied the visual search of layouts with and without a useful visual hierarchy. The task relevant to the current research is the visual search of layouts without a visual hierarchy. Figure 1 shows a sample layout from the experiment.

Sixteen participants searched four different screen layouts for a precued target object. 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 used group A. Two-group layouts used groups A and B. Four-group layouts used groups A through D.

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.

Hornof (2004) presented models that predicted and

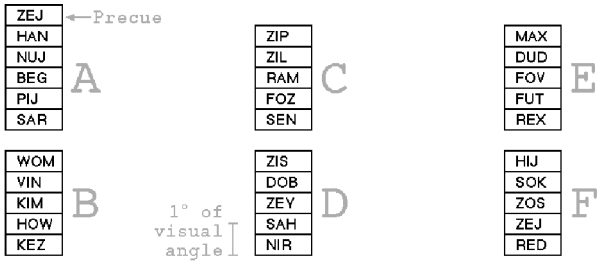


Figure 1. A 6-group layout. The precue, in the top left, would disappear when the layout appeared. The gray text did not appear during the experiment.

explained the search time data collected from the visual hierarchy task. Hornof and Halverson (2003) replicated the study to collect eye movement data to verify the eye movement strategies predicted by the models. 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. This selection strategy resulted in a very plausible explanation for how people did the task. The model accounted for the reaction time and a fair number of eye movement measures, especially considering that the model was built without the eye movement data to guide its development.

However, the model's strategy is perhaps somewhat tuned to aspects of this one visual task and layout. Aspects 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 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 need for a more flexible model that would predict the eye movements with greater fidelity and would do so in a more general, task-independent manner. The data collected by Hornof and Halverson (2003) are used in the current research.

The EPIC Cognitive Architecture

A series of computational cognitive models described in this study were built using the EPIC (Executive Process Interactive Control) cognitive architecture (Kieras & Meyer, 1997). EPIC captures human perceptual, cognitive, and motor processing constraints in a computational framework that is used to build cognitive models. Into EPIC, we encoded (a) a reproduction of the task environment, (b) the visual-perceptual features associated with each of the screen objects, such as the text feature, and (c) the cognitive strategies that guide the visual search, encoded as production rules. These components were added based on task analysis, human performance capabilities, previous visual search model, and parsimony.

After these components are encoded into the architecture, EPIC executes the task, simulates the perceptual-motor processing and interactions, and generates search time and eye movement predictions. EPIC simulates ocular-motor processing, including the fast ballistic eye movements known as saccades, as well as the fixations during which the eyes are stationary and information is perceived.

Evolving the Cognitive Strategy

This paper presents the several steps in the principled evolution of a model of visual search. The motivation for creating the model is the need for a computational model that is flexible enough to predict performance on a variety of menu-like visual layouts, and that can explicitly account for a wider range of eye movement measures than previous models.

The principled approach adopted here for building the model was to make gradual improvements based on "low-level" eye movement data (for example, fixation duration and saccade distances). At each step in the evolution of the model, a sub-strategy was added or a perceptual parameter was changed to increase the fidelity of the model. Basic visual search research or previous computational modeling motivated each change. It should be noted that each strategy or perceptual parameter change was considered "fixed" for later iterations of the model.

This model-building procedure resulted in gradual improvements, which we believe results in a model that meets our goal of a flexible, reusable model that accounts for how people search a visual layout. The following sections discuss four substantial steps made in the evolution of our cognitive model, starting with the motivation and explanation of the baseline model.

Step 1: Start with the baseline model

This modeling endeavor started largely as an attempt to integrate two pre-existing visual search models—the best-fitting model for the (unlabeled) visual hierarchy layouts from Hornof (2004) and the best-fitting final "mixed density" models from Halverson and Hornof (2004). In an effort to integrate the two, we started by finding the common elements between the best-fitting models for each of the visual search tasks. Interestingly, in the process of stripping down each of the models to find the common elements of both models so that they could be merged, we ended up with pretty much the same purely random model promoted by Hornof (2004), and the same purely random search strategy used in Halverson and Hornof (2004). They were integrated and used to start the exploratory modeling discussed here.

The new purely-random baseline model started with a strategy in which saccade destinations were selected at random from among potential targets. Beyond that, the model imposed a minimal number of constraints, primarily imposed by the EPIC cognitive architecture and task analysis, including:

(a) Search proceeded without replacement. In other words, objects were not selected as a saccade destination after their text had been identified. Analysis of our eye movement data suggested that people rarely fixated an object more than once. A model with no memory for fixated locations or objects would predict way too many fixations.

(b) Saccades were initiated after the fixated objects are identified. This was a feature of the "mixed density" model from Halverson and Hornof (2004). In EPIC, the visual properties of objects are available at varying eccentricities. For the *text* property, the default availability radius is one degree of visual angle from center of fixation. Once an object enters the availability region of a property, that property enters working memory after an amount of time determined by two parameters: (i) transduction time (50 ms for text), the time it takes from the information to reach sensory memory, and (ii) recoding time (100 ms for text), the time it takes to recognize the property. Given the strategy used in these

models, these constraints directly affect when the next saccade is initiated.

(c) EPIC's oculomotor feature preparation time parameter was changed to zero. Recent progress with the EPIC cognitive architecture has found that oculomotor preparation time may not be necessary or may occur in parallel with saccade destination decisions (Kieras & Meyer, 2005). Movement feature preparation time was previously determined based on shared features (e.g. direction and extent) with the previous motor movement. Initiation and execution times are still required.

These three constraints persisted throughout all models discussed in this paper.

Combining these constraints with the baseline random strategy, the resulting model predicted only one eye movement metric quite well, namely mean fixation duration. Figure 2 shows the predicted and observed fixation durations by layout size. The model predicts the mean fixation duration with an average absolute error (AAE) of 7.8%. In that our goal is an AAE of less than 10%, this is an acceptable error.

The model did a poor job of predicting other eye movement data, including saccade distance, fixations-per-group, fixations-per-trial, and scanpaths. Many of these shortcomings result because the model does not accurately predict trends in saccade destinations. Though a purely-random search strategy is good first approximation for predicting mean search times, a more refined strategy is needed for a robust, reusable, general purpose model of visual search.

Step 2: Refine the saccade destinations

As discussed previously, the two models whose integration initially motivated this research used either task-specific or purely random strategies. Step 2 pursued a more flexible strategy. To this end, Step 2 worked to improve the prediction of saccade destinations.

Two metrics were used to determine saccade destinations: mean saccade distance and mean fixations per group. Saccade distance measures the distance between contiguous fixations. Fixations-per-group measures the number of contiguous fixations within one group in the layout.

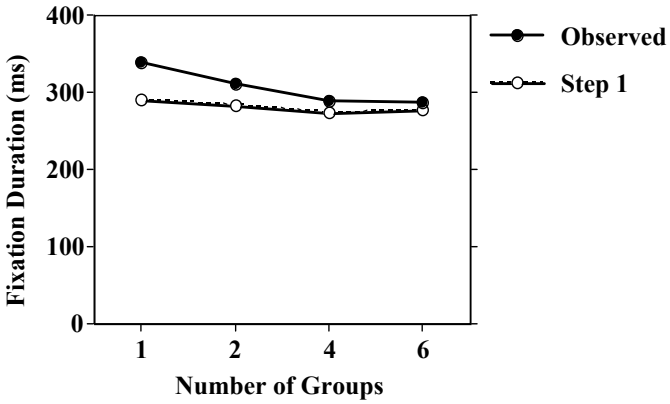


Figure 2. Fixation duration observed (solid line) and predicted by the Step 1 model (dashed line). AAE = 7.8% Error bars are too small to be visible (standard errors < 15)

Direct visual inspection of hundreds of individual eye movements made by participants revealed two clear patterns not accounted for by the Step 1 random model. First, once participants had finished dwelling on a group, they tended not to revisit that group until the remainder of the layout had been searched (this was true 94% of the time). Second, participants were more likely to saccade to nearby objects rather than to distant objects. Step 2 introduces two modifications that account for these behaviors.

To maintain forward progress in the search, a sub-strategy was added to prohibit group revisits until all groups had been searched. If two contiguous fixations land on two different groups, then objects in the first of the two groups are no longer potential saccade destinations until the entire layout had been searched. This sub-strategy uses layout-specific information, that objects are organized into groups, but we suspect that most visual layouts will have some sort of natural grouping that can be similarly used.

People do not search randomly. When searching, they are more likely to saccade to objects that are relatively nearby rather than objects across the layout. In visual search, saccade destinations are based on proximity to the center of fixation (e.g., Motter & Belky, 1998). Other models of visual search prefer nearby objects as saccade destinations (e.g., Barbur, Forsyth, & Wooding, 1990).

The Step 1 model was modified so that saccade destinations were selected based on proximity to the center of fixation. Objects in EPIC have a property, *eccentricity*, which reflects the object's distance (in degrees of visual angle) from the center of fixation. The random saccade destination selection strategy was changed to select the potential target with the least eccentricity. To account for variability in the human saccade distances, noise was also added to the eccentricity to vary saccade distances, while at the same time preferring nearby objects.

Saccade destinations are thus selected as follows: (a) After each saccade, the eccentricity property is updated based on the new eye position. (b) The eccentricity property is scaled by the eccentricity fluctuation factor, which has a mean of one and a standard deviation of 0.3. This scaling factor is individually sampled for each object after each saccade. (c) Objects whose text has not been identified and that were in unvisited groups are marked as potential candidates for the saccade destination. (d) The candidate object with the lowest eccentricity property, after the scaling factor is applied, is selected as the next saccade destination.

The standard deviation of the fluctuation factor was determined by varying the fluctuation factor (by increments no smaller than 0.01) to find the best fit of both the mean saccade distance and mean fixations per group. We recommend this parameter setting for future modeling.

Figures 3 and 4 show the Step 1 and 2 model predictions for mean saccade distance and mean fixations per group. As can be seen, the Step 2 model predicts the data much better. The two modifications made to the model dramatically decreased the error in the predicted eye movement data.

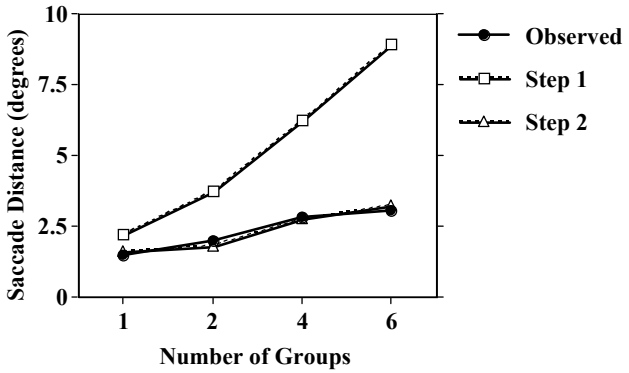


Figure 3. Saccade distance observed (circle), predicted by the Step 1 model (squares), and predicted by the Step 2 model (triangles).
Step 1 AAE = 112%, Step 2 AAE = 5.8%
Error bars are too small to be visible (standard errors < .2)

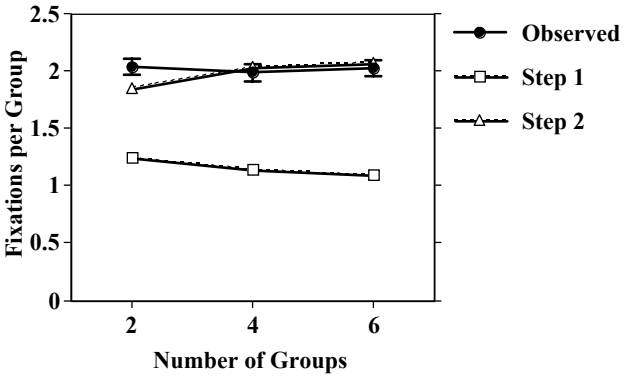


Figure 4. Fixations per group observed (circles), predicted by the Step 1 model (squares), and predicted by the Step 2 model (triangles). Step 1 AAE = 42%, Step 2 AAE = 4.6%
Error bars indicate ±1 standard error.

The improvements made to the model in this step have improved the fidelity of the model, while making the model more reusable and flexible. The model requires only one directly-extractable, task independent object feature—location. However, the model still requires improvement. As will be seen, the model still finds the target too quickly, even though the model correctly predicts how long people dwell in each group.

Step 3: Account for whole-task performance

Our goal is to produce a model that accounts for multiple eye movement measures. This includes accounting for eye movements at multiple scales. The previous iteration of the model accounted for the number of fixations per group, which can be viewed as accounting for a sub-task (searching each group) of the whole task (searching the entire layout). We next investigated means of improving the model at the "whole-task" level.

Again, a qualitative analysis of the participants' eye movement behavior suggests what might be needed in the model. It was observed that the participants sometimes looked at or near the target but continued to search. This suggests that the participants may be failing to recognize the target occasionally. It should be noted that it is unlikely that the participants did not react to the target because they had forgotten the target, as the participants eventually found the target and completed the task successfully.

Previous modeling research suggests that people do occasionally fail to recognize fixated text. In Halverson and Hornof's (2004) "mixed density" model, a perceptual parameter was introduced to explain an increase in the likelihood of missing a target based as a function of the text density. The modeling suggested that even in sparse text, people fail to recognize the target with approximately a 10% probability.

The model was modified to include a *text recoding failure rate*. Text recoding failure rate has only recently been added to EPIC, and the default value is zero (i.e. no chance of failing to identify text). The parameter represents the probability that the text property of fixated visual objects will be unknown.

This perceptual parameter was used in the current work for two reasons. First, to explore ways to account for the observation that participants missed the target occasionally. Second, to potentially provide converging support for the validity of using this parameter. If the current exploratory modeling predicts observed eye movement data with a text recoding failure rate similar to that used in the previous modeling, this would not only support the use of the parameter here, but also suggest a recommended default value for the parameter for future modeling.

The text recoding failure rate was initially set 10%, the value used in the previous modeling effort for sparse text (Halverson & Hornof, 2004). This failure rate was varied by 1% increments until the model predicted the mean number of fixations per trial. A value of 9% provided the best fit for the number of fixations per trial, the eye movement measure used to evaluate "whole-task" level performance.

Figure 5 shows the observed and predicted number of fixations per trial. As can be seen in the figure, the Step 2 model under-predicts the total number of fixations required to

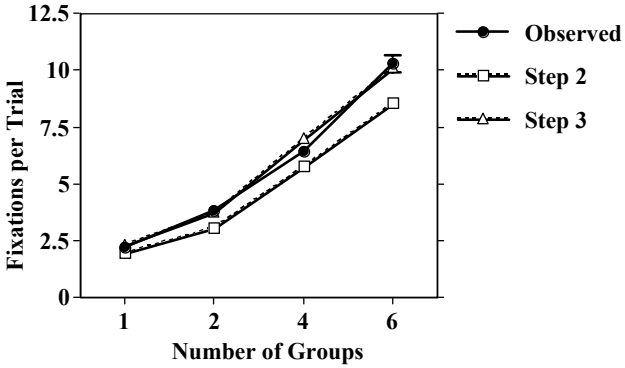


Figure 5. Fixations per trial observed (circles), predicted by the Step 2 model (squares), and predicted by the Step 3 model (triangles). Step 2 AAE = 14.3%, Step 3 AAE = 4.2%
Error bars indicate ±1 standard error. Some error bars are too small to be visible.

find the target by 14%. This is not a bad prediction, but an error of less than 5% would be ideal.

As shown in Figure 5, the Step 3 model predicts the number of fixations per trial with an error of 4.2%. This is a very good prediction. The decreased error and the similarity between the best fitting text recoding failure rate found here and the rate found in past research provides support for the use of this perceptual parameter here. This finding suggests that future modeling of menu-like search tasks should use a text recoding failure rate of around 9-10%.

Step 3A: Increase visual working memory decay

An interesting interaction between small layouts and EPIC's visual working memory gave rise to a surprising prediction. Occasionally the model would search a small layout without finding the target (due to text recoding failures introduced in Step 3) and stall. The model assumed that objects whose text properties (regardless of a text recoding failure) existed in the visual perceptual store were not candidate saccade destinations. EPIC's visual perceptual store retains the properties of objects for 500 ms after the eyes moves. Therefore, the text properties of all objects were known after the second fixation and there were no candidate destinations for a third saccade. A variety of solutions were pursued, but only one was consistent with recent literature and did not worsen eye movement predictions.

Woodman, Vogel, and Luck (2001) showed that when VWM is occupied, visual search remains efficient. When people are given a task that fills VWM with visual properties like shape, and then perform a second task searching for a shape, search rates are unaffected. One interpretation of these findings is that VWM decays quickly for goal-irrelevant information, like non-targets.

The model was modified by setting the perceptual store property retaining-time parameter to 50 ms. We would recommend this setting for future visual search models that include small layout conditions.

Step 4: Add a global strategy

Step 4 adds a global search component to improve the robustness of the model for predicting the frequency with which various scanpaths are followed.

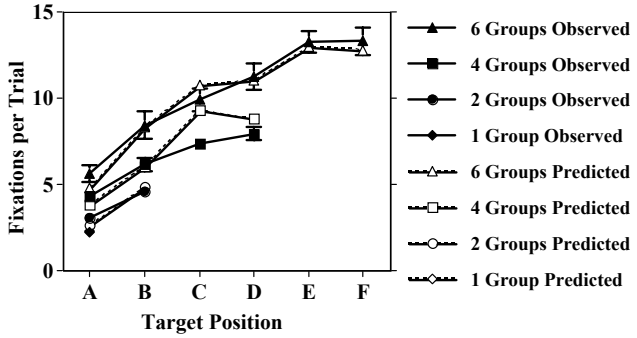


Figure 6. Fixations per trial observed (solid lines) and predicted by the Step 3 model (dashed lines). AAE = 8.1%. Error bars indicate ±1 standard error. Some error bars are too small to be visible.

Figure 6 shows the number of fixations per trial as a function of target group. (Figure 1 identifies the six groups as A through F.) There is a slight bump in the data when the target is located in group C. The purely local strategy for selecting nearby objects as saccade destinations motivates the model to reach group D before group C, which was not the case with people. Though the effect is slight (with an overall AAE of 8.1%), we believe this trend points to the need for some sort of global component to the strategy.

In local strategies, saccade destinations are determined based on what is encountered during the course of the search. In global strategies, saccade destinations are planned out in advance based on the task and stimuli.

A global component was added to the strategy such that the model could develop a global "preference" for scanning horizontally or vertically. A preferred scanning direction is established after the model, using the local strategy, starts searching horizontally or vertically. Once a direction is established, it is preferred unless no more groups exist in that direction.

Figure 7 shows that the global component slightly improved the model's prediction of fixations per trial. Most important, the bump in the data for group C is diminished.

Figure 8 shows the three most frequently observed scanpaths, as well as the predictions of the Step 3 and Step 4 models. People tended to start by going either down the first column or across the top. As shown in the predictions, the Step 3 model almost never goes across the top. However, the Step 4 model increased the frequency.

The improvements made by adding the global strategy are

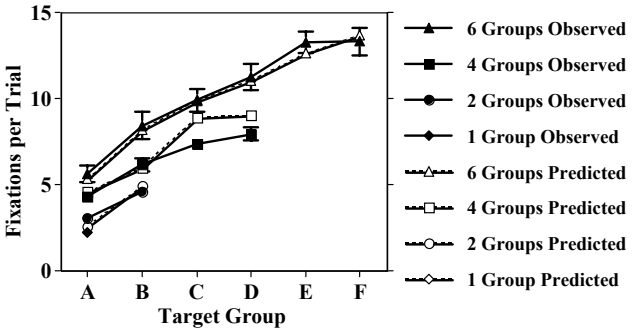
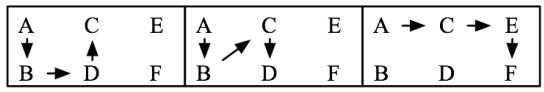


Figure 7. Fixations per trial observed (solid lines) and predicted by the Step 4 model (dashed lines). AAE = 6.5%. Error bars indicate ±1 standard error. Some error bars are too small to be visible.



Observed:	30%	18%	11%
Step 3:	30%	21%	1%
Step 4:	36%	27%	6%

Figure 8. The most commonly observed scanpaths in six-group layouts and how often each path was taken by the participants (observed) and the models (Step 3 and 4).

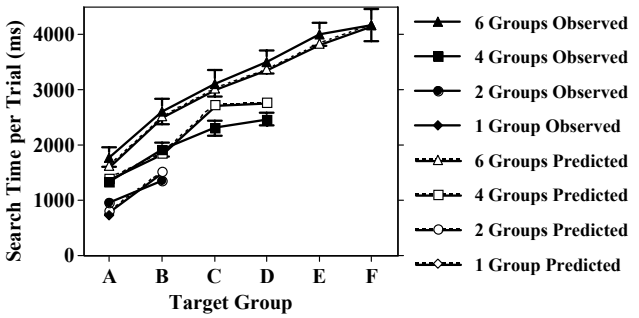


Figure 9. Search time per trial observed (solid lines) and predicted (dashed lines). AAE = 7.1%. Error bars indicate ±1 standard error. Some error bars are too small to be visible.

subtle, but they add to the fidelity of the model. However, the addition of the global strategy does not improve the quantitative fit of the model substantially and the addition may be considered overfitting as the additional production rules introduce additional free parameters. Therefore, the Step 3 model may be a better candidate on which to build successive flexible, reusable models of visual search that account for more factors.

Figure 9 shows the observed and predicted search times of the Step 4 model. The model predicts the observed search time quite well. This is a validation of the principled approach used to gradually improve the model using a variety of eye movement data. Moreover, it is gratifying to find that the model is able to make such accurate predictions without using the more brittle strategies of its predecessors.

Conclusion

A flexible and reusable model of visual search was developed that accounts for a wide variety of search data by (a) saccading to nearby objects when the fixated text is recognized, (b) positing a partial inspection of some objects and an occasional failure to identify others, (c) remembering more-or-less where but not what's been searched, and (d) accounting for people's tendencies to follow regular scanpaths with an element of a global strategy. The model explains the observed saccade distances, the number of fixations to each group in a layout, the total number of fixations in a trial, the number of fixations to find an object based on the object's location in the layout, the fixation duration, and to a slightly lesser extent the scanpaths that people used. The prediction of such a wide variety of measures bodes well for *a priori* prediction of visual search.

The model is flexible and reusable. The strategy is not tuned to the visual layout of the task. The only features required by the model are the location and identification (*text*) of the visual objects to be searched. If the visual layout is divided into clearly distinguishable groups, the model can utilize that information, but this division is not required. The model is currently limited to the visual search of textual layouts, but most aspects of the model are clearly generalizable to other stimuli.

The integration of recent, relevant psychological phenomena benefits the continued integrative development of computational models and advances in basic psychological research, and thus for Cognitive Science in general. Phenomena include general saccadic selection behavior (Motter & Belky, 1998), visual working memory (Woodman, Vogel, & Luck, 2001), and the integration of both local *and* global strategies. This work will continue with further integration of cognitive models of visual search from various cognitive architectures.

Acknowledgments

This research was supported by the Office of Naval Research and the National Science Foundation.

References

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*. London: Taylor & Francis.

Burke, M., Hornof, A. J., Nilsen, E., & Gorman, N. (2005). High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *Transactions on Computer-Human Interaction, 12*(4), 423-445.

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*, 124-129, Pittsburgh, Pennsylvania.

Hornof, A. J. (2004). Cognitive Strategies for the Visual Search of Hierarchical Computer Displays. *Human-Computer Interaction, 19*(3), 183-223.

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

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., & Meyer, D. E. (2005). Epic Progress. *Paper presented at the presented at the Office of Naval Research Cognitive Modeling Grantee Meeting*, Carnegie Mellon University, Pittsburg, PA.

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

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.

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.

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

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.