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 research advances computational cognitive modeling of visual search, and the synergistic relationship between cognitive modeling and eye tracking. The research examines 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 based on the eye movement data. The research demonstrates a methodology and provides guidance for predictive cognitive modeling of visual search.

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 known item 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.

The task was presented to sixteen experienced computer users. 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.

Participants searched eight different screen layouts for a precued target object. Each layout contained one, two, four, or six groups. 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, group labels, and the target position were randomly selected for each trial.

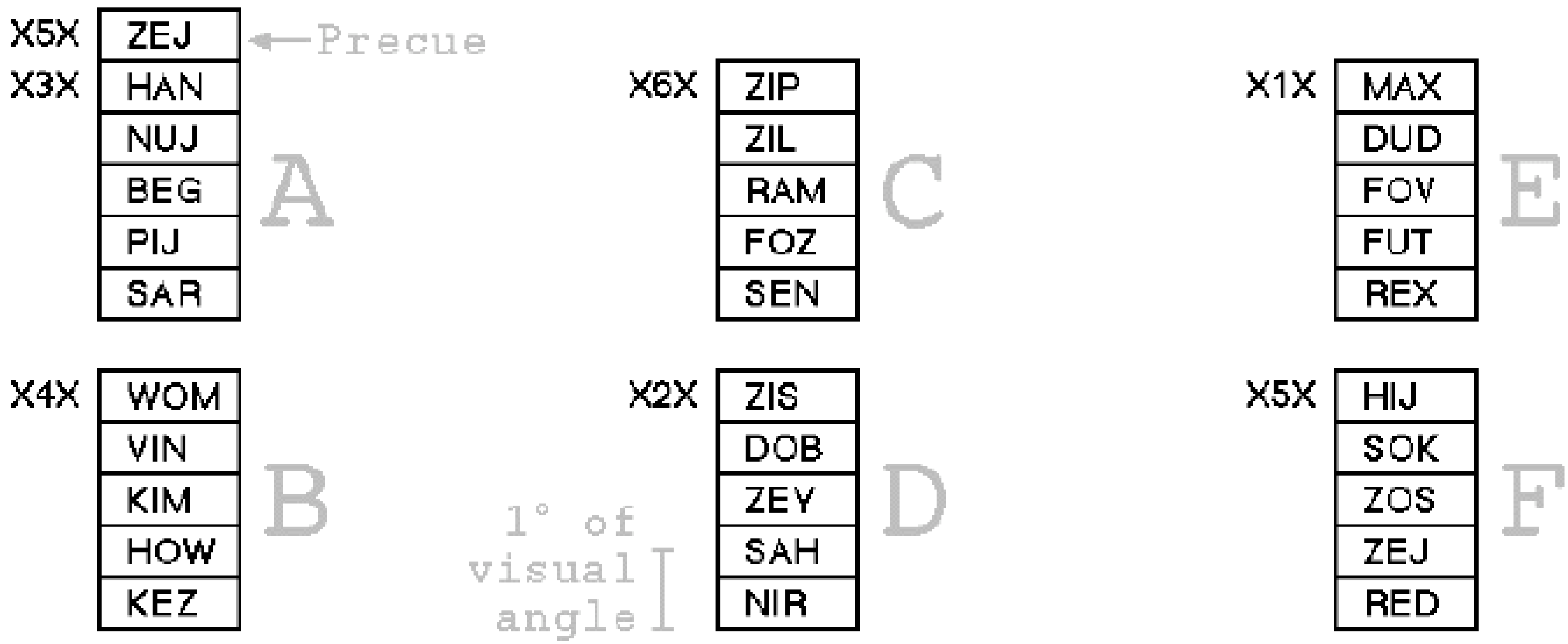

Figure 1. A “6-group labeled” layout. The precue, in the top left, would have disappeared when the layout appeared. The target is in group F. The gray text did not appear during the experiment.

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.

As is required to use the architecture, we encoded into EPIC the cognitive strategies that guide the visual search. The following two strategies provided a particularly good fit with the observed data.

The noisy-systematic search strategy for unlabeled layouts assumes that people attempt to make a “maximally-efficient foveal sweep”, 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.

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.

Predicted and Observed Eye Movements

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.

The *a priori* predicted and the observed eye movements were compared. Figures 2 and 3 show the predicted and observed eye movements from one trial with an unlabeled layout, and from one trial with a labeled layout. The figures give 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. These data, as well as other aspects of this research, are elaborated in Hornof & Halversion (2003).

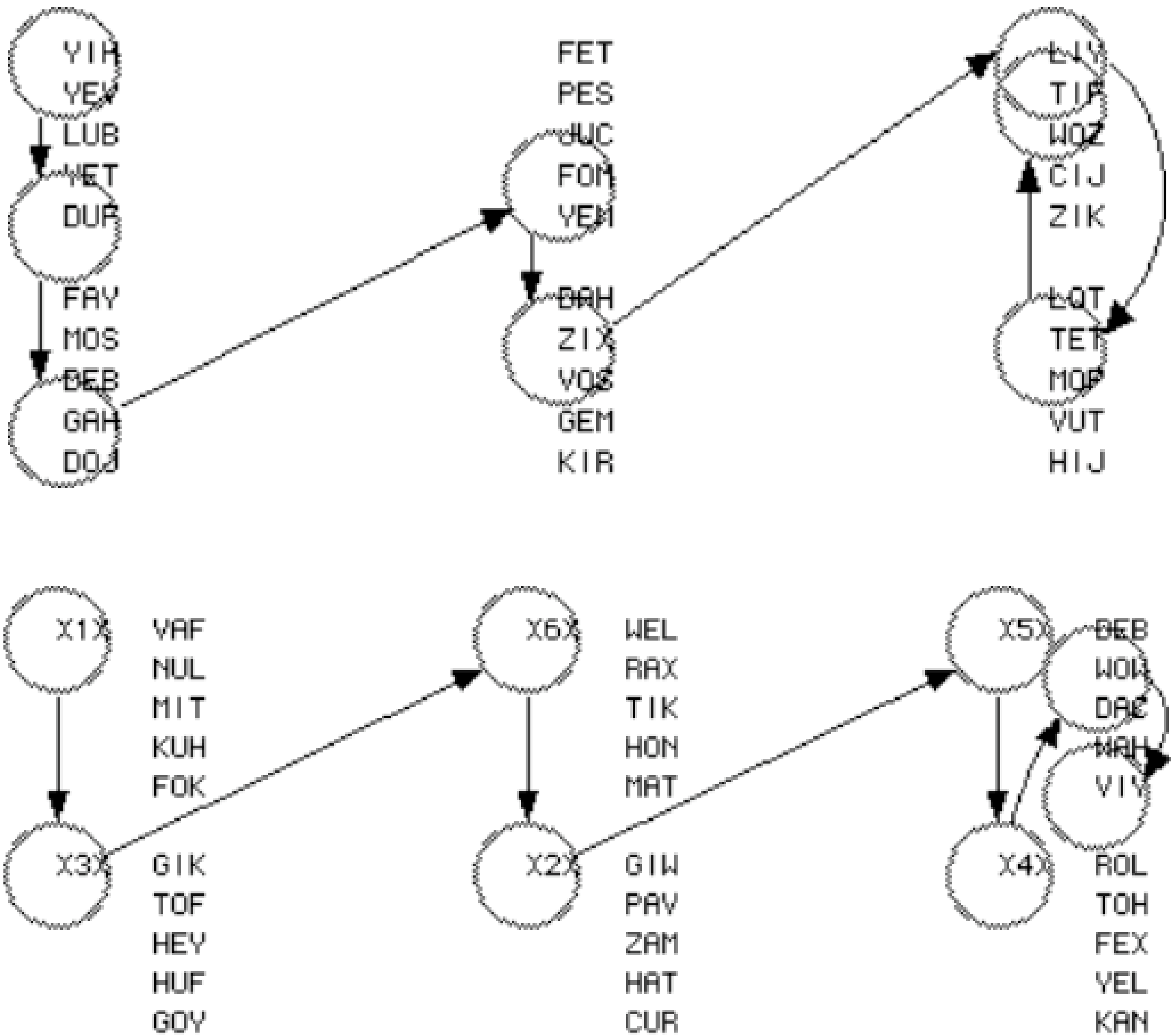

Figure 2. Fixations predicted from one trial with an unlabeled layout (top) and one trial with a labeled layout (bottim). The circles represents the foveal region. The unlabeled layout fixations are predicted by the noisy-systematic strategy. The labeled layout fixations are predicted by the two-tiered systematic strategy.

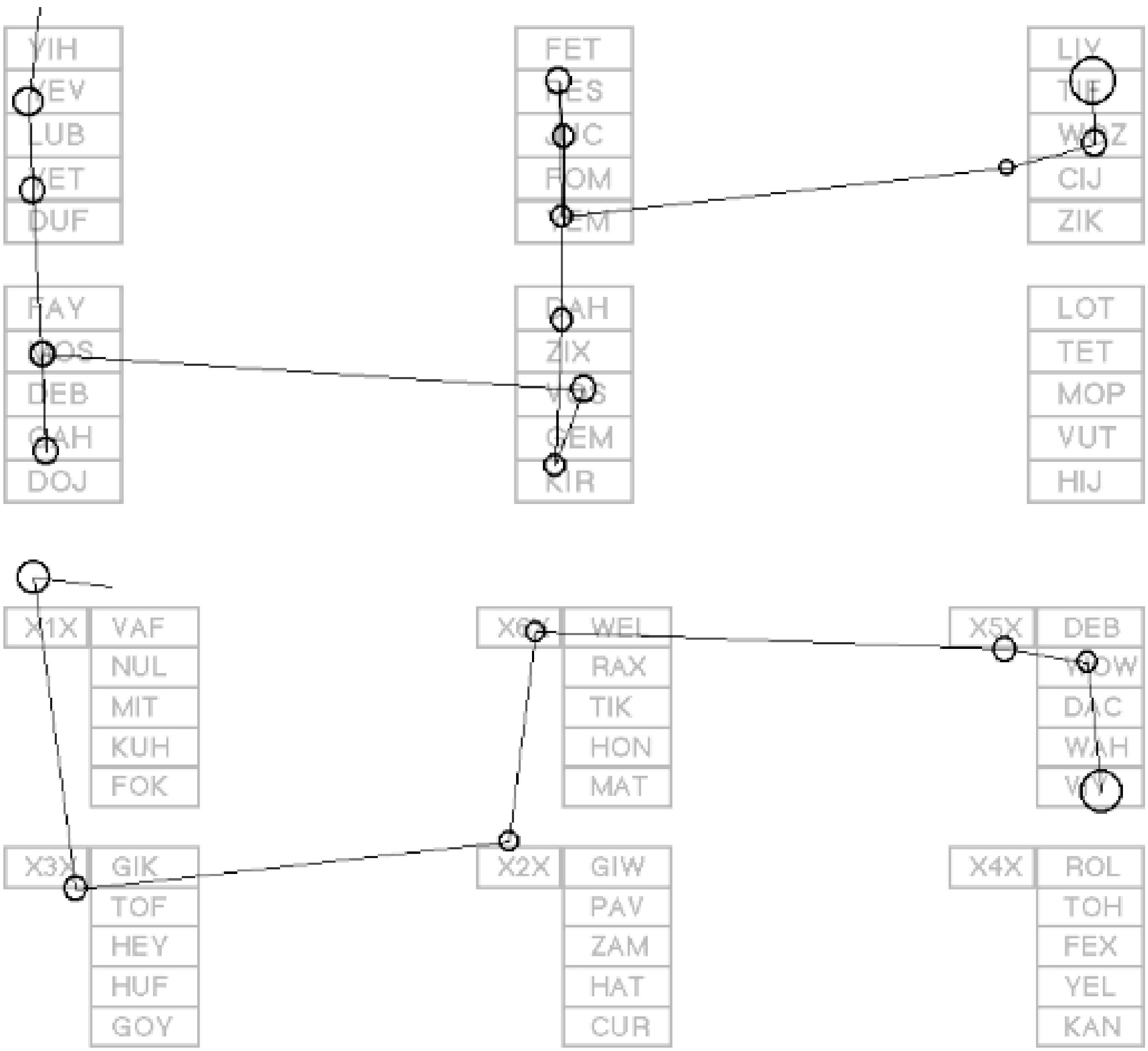

Figure 3. Fixations observed from one trial with an unlabeled layout (top) and one trial with a labeled layout (bottom). The diameters of the circles represent the fixation duration.

Table 1. A summary of the predicted and observed eye movements.
Plus signs 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 targe	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

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.

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

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.

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, New York: ACM. Downloadable at http://www.cs.uoregon.edu/~hornof/downloads/CHI03.pdf

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.