

Density Guides Visual Search: Sparse Groups are First even when Slower

Katherine A. Tarling & Duncan P. Brumby

UCL Interaction Centre
University College London
London UK WC1E 6BT UK

How do people efficiently locate content in a display? We investigate the effect of text layout on how people decide which area of a display to search first. Using a visual search paradigm, participants were required to locate a known target within a two-column display, in which items were grouped into semantic clusters, and the physical distance between items varied. For ‘mixed’ trials, the distance between items in each column was varied. Results showed that participants preferred to search the sparser of the two columns first, even though they were faster at locating the target in the denser column. This finding suggests that participants were adopting an inefficient search strategy for locating the target item. Discussion focuses on the implications for models that assume people rationally adapt their search strategy to maximize the gain of task-relevant information over time.

INTRODUCTION

Many interactive computing systems require the user to search a visual display. A core question concerns how people control the visual search process to decide in what order to look at the various visual elements in the display. One area where this question has been studied extensively is investigating how people search web pages to locate content information (e.g., Brumby & Howes, 2008; Curtrell & Guan, 2007; Fu & Pirolli, 2007). When searching the web, users tend to focus their visual attention on regions of the display that are more likely to contain informative content (Buscher, Cutrell, & Morris, 2009) and avoid regions that contain irrelevant content, such as banner advertisements (Burke, Hornof, Nilsen, & Gorman, 2005). To locate content efficiently, users exploit visual design conventions, such as the organization of semantically related items into distinct spatial regions under headers (Halverson & Hornof, 2008; Hornof, 2001).

These examples show that people have well-honed strategies for controlling the visual search process to efficiently locate content in a display. The stronger theoretical claim that has been made is that people rationally adapt their search strategy to maximize the gain of task-relevant information over time (Brumby & Howes, 2008; Cox & Young, 2004; Fu & Pirolli, 2007; Pirolli & Card, 1999; Tseng & Howes, 2008). A core prediction of these rational models is that the optimal strategy for searching a given display is sensitive to changes in cost and benefit, and that people will adopt the most efficient strategy given the constraints on their cognitive and perceptual processing abilities.

In support of this rational perspective, it has been shown that very small changes to the layout of a display can lead people to adopt quite different search strategies. For instance, Everett and Byrne (2004) found that moving a visual icon by just 1.6 degrees of visual angle from its corresponding text label dissuaded people from looking at it because doing so would require an additional eye movement to encode the icon.

Evidently people employed a search strategy that skipped the icon based on an assumption that the benefit to be had from encoding the icon was not worth the time cost of looking at it.

Very little work has investigated how people decide where to look in a display that has regions that differ in how quickly and easily they can be searched. One exception is a study by Halverson and Hornof (2004) in which participants were required to locate a predefined target in a display. The density of items in the display was varied, resulting in regions where there were fewer items in a large font (sparse regions) and regions where there were more items in a smaller font (dense regions). Halverson and Hornof found that when searching a ‘mixed’ display, that contained both sparse and dense regions, participants showed a preference for looking at items in the sparse regions first.

To explain this preference for searching sparse regions of the display first, it is worth considering how participants in Halverson and Hornof’s (2004) study searched each region in isolation. The sparse layout contained fewer items, of greater font size, that were positioned farther apart from each other, than items in the dense layout. Consequently, when participants searched displays conforming to an all-sparse layout, they were faster at locating the target and made fewer fixations, each of shorter duration. Given that sparse regions could perhaps be searched faster, possibly because the larger text could be processed faster, it would be rational to direct attention to items in these regions first. This processing benefit for sparse regions does not always hold, however. Many studies in the visual search literature have shown that when item size is controlled for, search times tend to decrease, as opposed to increase, when items are brought closer together (Bertera & Rayner, 2000; Ojanpää, Näsänen, & Kojo, 2002).

In one particular study, Ojanpää, Näsänen, and Kojo (2002) had participants search a list of words for a given target. The distance between items in the list was varied, but unlike in Halverson and Hornof’s (2004) study, the font size and number of items that were presented was held constant

across conditions. Ojanpää, Näsänen, and Kojo (2002) found that when the distance between items was increased, search times increased. This increase in search time occurred because more fixations were required to evaluate each item in the display when they were spaced farther apart.

Furthermore, a number of studies have shown that when items are positioned close together, information from multiple items is available from a single fixation, meaning that not every item assessed is directly fixated (Brumby & Howes, 2008; Hornof, 2004; Tseng & Howes, 2008). Therefore items in a more densely packed display can be searched more efficiently, with fewer eye movements, than items that are placed farther apart from one another.

In this paper, we report an experiment that investigated how people search displays that contain distinct visual regions that can be searched with varying degrees of efficiency. Participants were required to look for a known target within a two-column display in which the vertical distance between items was varied. We were primarily interested in how people chose to search ‘mixed’ layout displays, in which the distance between items in each column was varied. Given the choice between columns, we expect a rational searcher to show a preference for evaluating the column with more closely spaced items because it can be more efficiently searched, requiring fewer eye movements. Such a finding would be in contrast to Halverson and Hornof’s (2004) earlier result that showed people have a preference for searching sparse areas first.

A secondary aim of the experiment was to investigate how semantic structure is used to guide the search process. It is known that people use semantic structure to rapidly focus their attention on the most relevant region of the display (e.g., Brumby & Howes, 2008; Halverson & Hornof, 2008). In Halverson and Hornof’s study, participants searched menus that contained spatially distinct groups of semantically related items. Participants were able to use this semantic structure to determine which group was most likely to contain the target. The boundary of each group was visually salient, however, and we were keen to see whether people were sensitive to semantic structure even when there were no visual cues to indicate the boundaries of each group. This study varied the number of items in each semantic group, on the assumption that people would be more efficient at searching menus that contained larger semantic groups. The study primarily focuses, though, on the effects of spacing.

METHOD

Participants

Twenty-two people (nine female) were recruited from the psychology research participation panel at University College London to take part in the study. Participants were between 22 and 44 years of age ($M=29$ years). All were native English speakers, with normal or corrected-to-normal vision. Participants were paid £7 for their time.

Materials

Stimuli were presented on a 17inch TFT monitor set at a resolution of 1024 by 768 pixels. Participants were seated approximately 60cm from the monitor. Eye movements were

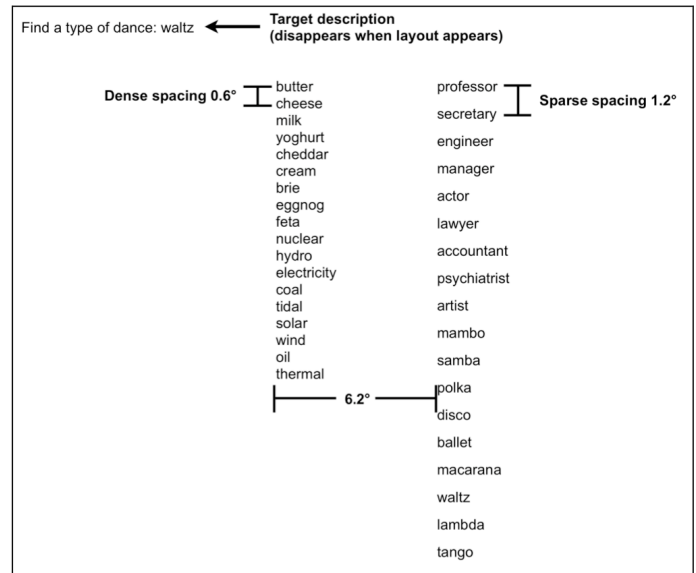


Figure 1. A mixed layout where items are arranged as four distinct semantic groups of nine items each. All angle measures represent approximate degrees of visual angle.

recorded using a Tobii 1750 eye tracker. The sampling rate of the eye tracker was 50 Hz, with gaze point accuracy of less than 0.5 degrees of visual angle. The experimental software ran on a Dell Optiplex machine with 1 GB of RAM running Microsoft Windows XP. An optical mouse was used, set at the ‘medium’ speed via the system control panel.

The visual search task required participants to locate a predefined target item in a menu containing 36 items. Figure 1 shows an example layout for a mixed display. Menu items were displayed in two separate vertical columns, with each column containing 18 items. All text was presented in an Arial font, size 12. The vertical distance between items in each column varied between conditions. Items in the dense column were 33 pixels apart, giving a separation of approximately 0.6° visual angle between items. Whereas items in a sparse column were 20 pixels apart, giving a separation of approximately 1.2° visual angle between items. Regardless of layout, the first item in each column was always positioned 90 pixels from the top of the screen, and there was a horizontal distance of 200 pixels between columns from left edge to left edge, giving a separation of approximately 6.2° visual angle.

Each menu contained sets of words that belonged to a natural category. The set of words used were taken from Halverson (2008, pp. 158-164), originally reported in Yoon et al (2004). Across the set of materials, there were 14 top-level categories (e.g., animal, building, entertainment). Underneath each of these top-level categories there were four mid-level categories (e.g., bird, farm animal, tropical fish, and wild animal, all types of animal), and for each of these there were 10 items (e.g., bluebird, canary, starling, and eagle, all types of bird). In total, there were 560 items available. No word appeared more than once in the database.

For each trial, 36 menu items were randomly chosen from the set of available items. The constraints on this sampling process were determined by the semantic group size condition of the participant. In the large group size condition, each menu

contained semantic groups made up of nine items from four different top-level categories. In the small group size condition, each menu contained groups of three items from 12 different top-level categories. The target item was randomly chosen from one of the available semantic groups and appeared in a random location within that group of items. The participant was given a specific target and category description before each layout appeared.

Design

The experiment used a 2 x 3 (semantic group size x display layout) mixed design. Semantic group size was manipulated as a between-subjects factor, such that participants searched menus where items were grouped into either large or small sets of semantically related items. There were three display layouts used: all dense, all sparse, and mixed. For the mixed layout, the distance between items in each column was varied, giving one sparse and one dense column (see Fig. 1). Column density was randomized between trials, meaning that the dense column could appear on either the right or left side of the display. The position of the target in the menu was also randomized, such that it was equally likely to appear in either column.

The main dependent variable of interest was search time and a number of eye movement measures that were used to infer visual search strategy. In particular, we were interested in whether participants would show a preference for searching the sparse column first during mixed layout trials.

Procedure

Participants were informed that they would be required to perform a series of simple searches. For each search trial, a predefined target and category description was shown in the top-left corner of the screen (e.g., “Find a type of dance: waltz”). After encoding the target, participants displayed the menu by clicking on the target word with the mouse. This made the target and the description disappear, and the menu appear. Participants were instructed to locate the target as quickly and as accurately as possible. The trial ended when the participant selected the target with the mouse. If an incorrect selection was made, then the participant was instructed to make another selection from the same menu after looking at the target again.

Participants completed six practice trials, before completing 40 trials for each of the layout conditions (all-sparse, all-dense, and mixed). Trials were grouped by condition. Mixed trials were always presented last, and the order of all-sparse and all-dense trials was counter-balanced across participants. After completing a block of trials participants were given a break of three minutes. The eye tracker was calibrated before the start of each block to ensure accurate gaze tracking. The entire experiment took approximately 20 minutes to complete, and participants were informed that they were free to leave as soon as they had completed all of the trials.

RESULTS

For each trial we consider data from when the menu first appeared to when the participant selected an item. Data from

Table 1. Dependent measures for each layout condition.

| <i>Measure</i> | All-dense <i>M (SD)</i> | Mixed <i>M (SD)</i> | All-sparse <i>M (SD)</i> | Sig. |
|--|----------------------------|------------------------|-----------------------------|-------------|
| Search time (ms) | 2,990 (510) | 3,200 (750) | 3,330 (650) | * |
| Number of fixations | 13.8 (5.0) | 14.6 (4.6) | 15.5 (5.2) | * |
| Fixation duration (ms) | 214 (98) | 199 (83) | 193 (75) | <i>n.s.</i> |
| Number of items visited at least once | 9.8 (2.6) | 10.4 (2.6) | 11.0 (2.8) | * |
| Saccade distance based on item spacing | 2.2 (0.5) | 2.3 (0.4) | 2.1 (0.4) | * |

* $p < .05$

one participant in the small semantic group size condition was excluded because their mean search time was greater than two standard deviations from the participant mean. We also excluded data from a randomly chosen participant from the large semantic group size condition to balance the design. Data from trials in which an incorrect item was selected on the first selection were also removed (26 trials out of 2640). For statistical analysis, a 2x3 mixed factorial ANOVA was used.

In terms of the manipulation of semantic group size, participants were slightly slower at searching the smaller three-item semantic groups ($M=3,350\text{ms}$, $SD=340\text{ms}$) than the larger nine-item semantic groups ($M=3,030\text{ms}$, $SD=690\text{ms}$). The effect of group size on search time was non-significant ($p=.23$), as was the interaction with layout condition ($p=.36$). Moreover, there was no effect of group size on any of the other dependent measures considered. We therefore focus the remainder of our analysis on the effect of layout density.

Table 1 shows data for the main dependent measures for each of the layout conditions. In terms of average search time per trial, there was a significant effect of layout condition, $F(2,36)=4.46$, $p<.05$, $MSE=.14$. Pairwise comparisons showed that participants were faster at locating the target in the all-dense layout than the all-sparse layout ($p<.05$). There was also a significant trend for participants to be faster when searching the all-dense layout compared to the mixed layout ($p=.09$), but the difference between the mixed and the all-sparse layout was non-significant ($p=.25$).

We next consider a number of eye movement measures to better understand why participants were more efficient at locating the target when items were positioned closer together. Eye gaze position was recorded at a rate of 50 Hz, and converted to fixations using Tobii's ClearView analysis software. Fixations were defined with a radius of 20 pixels (approx. 1.2° visual angle) and minimum duration of 40 ms.

Table 1 shows fixation data for each layout condition. There was a significant effect of layout condition on the mean number of fixations per trial, $F(2,36)=3.41$, $p<.05$, $MSE=4.85$. Pairwise comparisons showed that significantly fewer fixations were needed to locate the target in the all-dense layout than the all-sparse layout ($p<.05$). It can also be seen in Table 1 that fixations tended to be longer in the all-dense layout than the all-sparse layout, but the effect of layout condition on fixation duration was not significant ($p=.17$).

We next consider whether the reason why fewer fixations were required to locate the target in the all-dense layout was

because more items could be assessed within a single fixation. If this were the case, then we would expect consecutive fixations to be farther apart in the all-dense layout (i.e., for there to be an increase in mean saccade distance). Of course, we cannot directly consider saccade distance in terms of physical distance because this measure would be confounded by the manipulation of item spacing. Instead, we consider saccade distance in terms of item spacing (i.e., the number of items skipped over between consecutive fixations). To do this, we formed a *visit sequence* for each trial in which fixations were mapped to items in the menu and multiple contiguous fixations to the same item were aggregated (c.f., Brumby & Howes, 2008). We consider the median number of items between each consecutive visit in the sequence, where moving between columns is considered to be a single item traversal. For instance, if items 1, 2, 19, and 18 were visited in order, then the mean saccade distance would be 1. Using this method we calculate the mean saccade distance for each condition.

The final row of Table 1 shows the average saccade distance between consecutive visits for each layout condition. There was a significant main effect of layout condition on the number of items skipped over between consecutive visits, $F(2,36)=3.66, p<.05, MSE=.10$. Pairwise comparisons show that there was a trend for participants to skip over more items when searching the all-dense layout than when searching the all-sparse layout ($p=.08$). This suggests that participants were more efficient when searching the denser layout because multiple items could be evaluated within a single fixation, resulting in fewer fixations being needed to locate the target.

We next consider how participants chose to search the mixed layouts. Recall that for mixed layouts, one column of text was sparse while the other was dense. Column density was randomized between trials, and the target was equally likely to be present in either column. Because participants were faster at locating the target in denser layouts, a rational searcher ought to search the denser column first.

Figure 2 shows the proportion of visits to the right column over the first eight visits. It can be seen that participants had a strong preference for starting their search in the left column (i.e., only 26% of first item visits are to the right column, regardless of the layout of that column). With each consecutive visit there was a growing propensity to shift to the right column. Interestingly, there is a far greater likelihood of shifting to the right column when that column had a sparse layout than when it had a dense layout.

We support these observations by performing a 2x2x8 mixed ANOVA on these data. Across the first eight visits participants were more likely to visit the right column when that column had a sparse layout ($M=53\%, SD=12\%$) than when it had a dense layout ($M=46\%, SD=12\%$), $F(1,18)=6.35, p<.05, MSE=.10$. There was also a main effect of item visit, reflecting the trend that participants were more likely to shift their attention from the left column to the right over time, $F(7,126)=10.79, p<.001, MSE=.04$. There was no effect of semantic group size ($p=.49$), nor any interactions ($p>.05$).

GENERAL DISCUSSION

The participants in our study had a preference for searching the sparser of the two columns first, even though

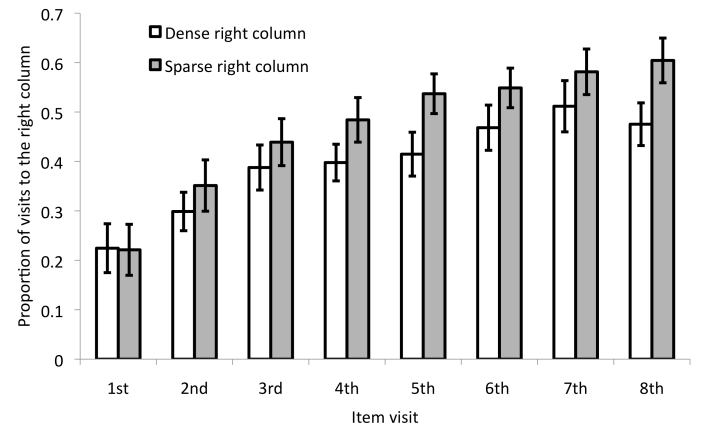


Figure 2. Proportion of visits to items in the right column for mixed trials. Error bars represent standard error of the mean.

they were faster at locating the target in the denser region of the display. Critically, this preference for searching the sparse region emerged after only a few fixations (see Figure 2). This finding is consistent with the results of Halverson and Hornof's (2004) study, which showed that people have a preference for searching sparse regions of a display first. In Halverson and Hornof's study, however, the sparse regions could be searched more efficiently. The results of the current study are novel because they show that people's preference for searching sparse regions remains even when there is no clear efficiency benefit to searching these areas first. In fact, the results of the current study show that participants were more efficient at locating the target in denser, and not sparser, regions of the display. The reason why participants were faster at searching layouts in which items were closer together is that multiple items could be evaluated within a single fixation, meaning that not every item that was assessed needed to be fixated.

The results of this study suggest that by choosing to search the sparser column first, participants were adopting an inefficient search strategy that impacted the average search time to locate the target. In particular, search times for the mixed layout condition were closer to the slower all-sparse layout condition than to the faster all-dense layout condition (see Table 1). This suggests that participants were adopting a relatively inefficient strategy for searching the mixed layouts. This finding seems at odds with various accounts that assume that people rationally adapt their search strategy to maximize the gain of task-relevant information over time (e.g., Brumby & Howes, 2008; Cox & Young, 2004; Fu & Pirolli, 2007; Pirolli & Card, 1999; Tseng & Howes, 2008).

But why did participants show a preference for searching the sparser region of the display first, given that it was inefficient to do so? It might be the case that this seemingly sub-optimal strategy reflects a lower-level adaptation to the demands of the environment. For instance, it might be the case that while faster, participants found locating the target in the dense region more effortful. Previous research has shown that eye movement fixations tend to be longer when items are more densely packed together (Bertera, & Rayner, 2000; Halverson & Hornof, 2004; Ojanpää, Näsänen, Kojo, 2002; Tseng & Howes, 2008). However, in the current study we

observed a small though not significant increase in the duration of fixations when participants searched denser layouts. This might suggest that participants were not lingering long enough to encode all of the information available in each fixation for the denser layout. If this were the case, then encoding errors should be more common in the all-dense layout and we might expect to see more re-fixations to the target item (i.e., because it has been seen in the periphery of vision but not fully encoded before the next fixation is executed).

One approach to evaluate the validity of this argument is to use a computational cognitive model. Halverson's (2008) model is an ideal candidate. The model is developed within the EPIC cognitive architecture (Kieras, & Meyer, 1997) and aims to provide an account of eye movements in visual search tasks such as the one used here. One possibility to explore with a model of this kind is whether encoding errors are more likely in dense regions of text. Kieras (2009) recently demonstrated how encoding errors play a critical role in explaining visual search data. Further analysis of the available data is required to address this question.

Finally, it is worth noting that a secondary aim of the study reported here was to investigate how semantic structure is used to guide search. Previous research has shown that people use semantic structure to rapidly focus their attention on the most relevant region of the display (e.g., Brumby & Howes, 2008; Halverson & Hornof, 2008). In the current study, we did not find an effect of semantic group size on performance, though participants were marginally faster when searching larger semantic groups. Paying attention to semantic grouping may not have conferred much value in the context of the current study, in that the boundaries between groups were not visually salient. Further work is needed to investigate whether visual boundaries are required to make semantic groupings useful in the kind of rapid, known item search tasks used here.

CONCLUSION

We have shown that people have a preference for searching sparse regions of a display first, even when there are regions of the display that might be more quickly searched first. This finding is at odds with various accounts that assume people rationally adapt their search strategy to maximize the gain of task-relevant information over time. Instead, we show that people have a propensity to search sparse regions of the display first. This finding supports the common design practice of placing important or leading content information (such as navigational elements, 'quick links' or 'refine search' options) in groups that are sparse relative to the other information on the layout (such as the main body of text or search results) because these regions are more likely to be looked at first.

ACKNOWLEDGEMENTS

The study reported here was undertaken as part of the first author's Master's dissertation work. We thank Tim Halverson, Anthony Hornof, and four anonymous reviewers for their valuable comments on this manuscript.

REFERENCES

- Bertera, J.H., & Rayner, K. (2000). Eye movements and the span of the effective visual stimulus in visual search. *Perception and Psychophysics*, 62, 576-585.
- Brumby, D.P., & Howes, A. (2008). Strategies for Guiding Interactive Search: An Empirical Investigation Into the Consequences of Label Relevance for Assessment and Selection. *Human-Computer Interaction*, 23, 1 – 46.
- Buscher, G., Cutrell, E., & Morris, M. R. (2009). What do you see when you're surfing? Using eye tracking to predict salient regions of web pages. In the *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09*. New York, NY: ACM Press.
- Burke, M., Hornof, A.J., Nilsen, E., & Gorman, N. (2005). High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Transactions on Computer-Human Interaction*, 12(4), 423-445
- Cox, A.L., & Young, R.M. (2004). A rational model of the effect of information scent on the exploration of menus. In *Proceedings of the 6th International Conference on Cognitive Modeling*, (pp. 82-87). Mahwah, NJ: LEA.
- Cutrell, E., & Guan, Z. (2007). What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07* (pp. 407-416). New York, NY: ACM Press.
- Everett, S.P., & Byrne, M.D. (2004). Unintended effects: Varying icon spacing changes users' visual search strategy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'04* (pp. 695-702). New York, NY: ACM Press.
- Fu, W.-T., Pirolli, P. (2007), SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human-Computer Interaction*, 22, 355-412.
- Halverson, T., & Hornof, A.J. (2004). Local density guides visual search: Sparse groups are first and faster. In the *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1860-1864). Santa Monica, CA: HFES.
- Halverson, T. (2008). *An "Active Vision" Computational Model of Visual Search for Human-Computer Interaction*. Doctoral Dissertation in Computer and Information Science. The University of Oregon, Eugene, OR.
- Halverson, T. & Hornof, A.J. (2008). The effects of semantic grouping on visual search. In the *Extended Abstracts of ACM CHI 2008: Conference on Human Factors in Computing Systems* (pp. 3471-3476). New York, NY: ACM Press.
- 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.
- Kieras, D. (2009). The persistent visual store as the locus of fixation memory in visual search tasks. In the *Proceedings of the 9th International Conference on Cognitive Modeling, ICCM2009*, Manchester, UK.
- Kieras, D. & Meyer, D.E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12, 391-438.
- Ojanpää, H., Näsänen, R., & Kojo, I. (2002). Eye movements in the visual search of word lists. *Vision Research*, 42, 1499-1512.
- Pirolli, P., & Card, S.K. (1999). Information foraging. *Psychological Review*, 106, 643-675.
- Tseng, Y.C. & Howes, A. (2008). The adaptation of visual search strategy to expected information gain. In the *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08*. New York, NY: ACM Press.
- Yoon, C., Feinberg, F., Hu, P., 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, 379-393.