Visual Availability and Fixation Memory in Modeling Visual Search using the EPIC Architecture

David Kieras (kieras@eecs.umich.edu)
Electrical Engineering & Computer Science Department, University of Michigan
2260 Hayward Street, Ann Arbor, Michigan 48109 USA

Sandra P. Marshall (smarshall@sciences.sdsu.edu)
Department of Psychology, San Diego State University
6475 Alvarado Road, Suite 132, San Diego CA 92120 USA

Abstract

A set of eye movement data from a visual search task using realistically complex and numerous stimuli was modeled with the EPIC architecture, which provides direct support for oculomotor constraints and visual availability constraints due to retinal non-homogeneity. The results show how the quantitative details of visual search can be explained within an architectural framework, and have useful practical, methodological, and theoretical implications.

Introduction

Many everyday and work activities involve visual search, the process of visually scanning or inspecting the environment to locate an object of interest that will then be the target of further activity. For example, one might search the kitchen to locate a package of desired coffee beans which will then be grasped. This work concerns computer interaction tasks, in which a particular icon coded by color, shape, and other attributes is searched for on the screen and is then clicked on using a mouse. In this domain, although the visual characteristics of the searched-for objects are artificially simple compared to most everyday objects, the task is a natural one in the sense that such activities are very common in the use of modern technology; they are perhaps the major performance bottleneck in important systems such as military workstations with which humans are expected to monitor and comprehend displays containing hundreds of moving icons in time-stressed and potentially lethal situations, and do so over extended periods of time. Hence analyzing such tasks in order to improve how well they can be done presents an opportunity to study real and important tasks that nonetheless involve relatively simple aspects of vision.

A currently developing approach to the design of systems that support such complex and high-performance tasks is to simulate human performance with a candidate design, using one of the computational cognitive architectures such as EPIC (Kieras & Meyer, 1997; see also Kieras, 2003, and Byrne, 2003). Any such effort will have to include reasonably accurate representations of the relevant aspects of human vision. In particular, the spatial and temporal extent of the tasks makes eye movements mandatory, but as argued forcefully by Findlay and Gilchrist (2003) in their presentation of *active vision*, mainstream cognitive psychology has under-represented this most salient feature of human vision as it works in natural environments and tasks, and thus has not developed the necessary theoretical components. These include the spatial non-homogeneity of the retina, which gives central and peripheral vision different roles, and the oculomotor mechanisms that move the eyes. The key cognitive activity in visual search is to use peripheral vision information to decide which object should be chosen to be the next target of central vision, and then to position the eyes accordingly, and repeat as needed.

Thus retinal non-homogeneity leads to differences in what properties of objects are currently visible as a function of their eccentricity and other properties, and are thus *available* to cognition for use in guiding the visual search. In general, the visual availability of a property of an object is a function not just of the eccentricity of the object, but also its size. An example is Anstis (1974) who showed that a letter can be recognized even in peripheral vision if it is large enough. A variety of other studies in the literature document similar effects for other properties, such as color (e.g. Gordon & Abramov, 1977). A third factor affecting availability is *crowding*; an object is less available if it is closely surrounded by other objects. Results by Bouma (1978) and others (e.g. Toet & Levi, 1994) show that the effects of crowding increase with eccentricity. However, these three factors combine in a complex way that appears to depend on the specific visual properties involved, and is not at all fully documented.

The EPIC architecture, which was developed to model humans in high-performance tasks, was perhaps the first computational cognitive architecture to explicitly represent visual availability and the time course of programming and executing eye movements, making it a natural framework for realizing models based on active vision concepts. The purpose of this paper is to present some recent results in which EPIC was used to model a complex realistic search task. The model, while not yet fully refined, demonstrates some key features of visual search.

The EPIC Cognitive Architecture

The EPIC architecture for human cognition and performance provides a general framework for simulating a human interacting with an environment to accomplish a task. Due to lack of space, the reader is referred to Kieras & Meyer (1997), Meyer & Kieras (1997), or Kieras (2004) for a more complete description of EPIC. Figure 1 provides an
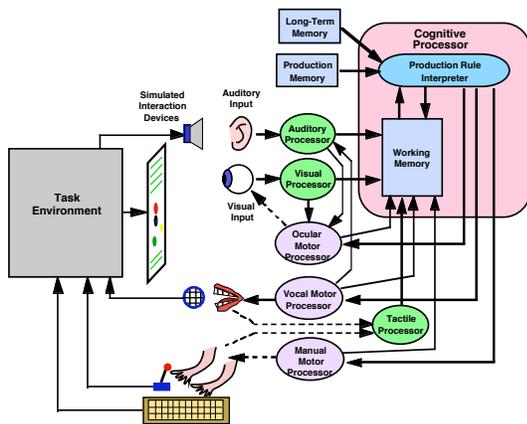
Figure 1. The overall structure of the EPIC architecture. Perceptual-motor peripherals surround a cognitive processor.

overview of the architecture, showing perceptual and motor processor peripherals surrounding a cognitive processor; all of the processors run in parallel with each other. To model human performance of a task, the cognitive processor is programmed with production rules that implement a strategy for performing the task. When the simulation is run, the architecture generates the specific sequence of perceptual, cognitive, and motor events required to perform the task, within the constraints determined by the architecture and the task environment.

EPIC's visual system follows the usual breakdown into a sensory store and a perceptual store, but there is also an eye processor that explicitly represents visual availability by determining which visual properties of objects are available in the sensory store as a function of the current position of the eye. The recognized objects and their relationships in the perceptual store are then available to the cognitive processor to match the conditions of production rules, whose actions can command the oculomotor processor to move the eyes.

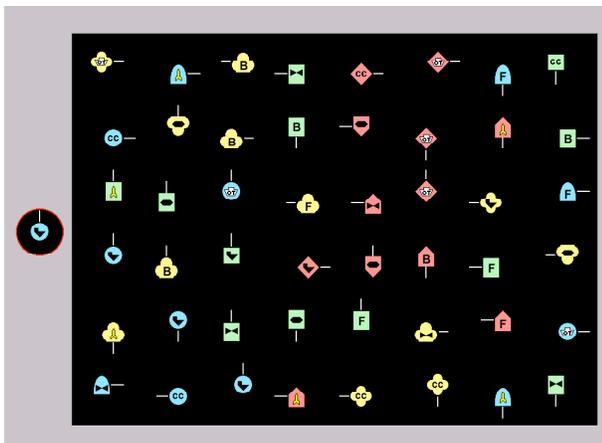The eye processor applies availability functions to

Figure 2. A sample display with 0% decluttering. The probe is at the left. One of the matching targets is immediately to the right; the other is third from the left in the bottom row.

determine what visual properties of an object are available given its size and eccentricity. Unfortunately, the research literature does not include studies from which one can construct availability functions that include a range of object sizes and eccentricities for realistic and useful visual properties, even for the highly artificial but practically significant stimuli such as would be found on computer displays.

The modeling work reported in this paper is part of an effort to arrive at useful visual availability parameters by constructing models for several visual search tasks and fitting them to the data, in hopes that the resulting parameter sets will be useful and generalizable. The data used in this work is a subset of the data described in St. John, Marshall, Knust, & Binning (2006).

The Data

The task. The experimental task required subjects to search for icons in a display that are identical to a probe icon. A sample display appears in Figure 2. A trial begins with a blank display and the appearance of the probe icon in the area at the left-hand edge of the display. After 2 sec, the to-be-searched icons all simultaneously appear to the right of the probe icon. In Figure 2, there are 48 icons to be searched, two of which match the probe icon. The subject's task is to click on each of the two matching icons. Once the second is clicked on, the display is blanked, and after an inter-trial-interval of approximately 2 sec, the next trial starts with the appearance of the probe icon.

Stimuli. The icons themselves are based on a new standard symbology for military displays called MIL-STD-2525B (Department of Defense, 1999), which is a military standard icon set for designating the kinds of objects that would appear on a military radar or tactical display with redundant coding for their militarily important properties. Each icon has a color, a shape, an internal symbol, and a "direction leader". Because these are actual military symbols, the properties are not at all orthogonal, but rather represent redundant codings. A full description can be found in St. John, Marshall, Knust, & Binning (2006).

Color represents the *Origin* of the object: red is hostile, blue is friendly, yellow is unknown, and green is neutral. Shape is redundant with the origin, but a certain feature of the shape also connotes the object's category: aircraft, ship, or submarine. Surface icons are shown as full shapes, air icons are truncated at the bottom, and subsurface icons are truncated at the top. Each specific shape appears in combination with one of eight correlated internal symbols that represents the kind of object, the *Platform*, such as cruiser, helicopter, aircraft carrier. For example, the bow-tie symbol represents a helicopter, and it appears in a shape that is missing its bottom, which means the object is a flying vehicle. The leader is a line segment commonly used on such displays to show the direction and speed of a moving object. In these icons, it is always the same length and appears in one of only four orientations to indicate *Direction* (N, S, E, W). Thus, three of the basic visual properties—

color, overall shape, and internal symbol—are somewhat confounded; a particular internal symbol can appear with any color, but only with a certain shape for each color. Color and overall shape are highly correlated. Blue (friendly) icons are curvilinear, red (hostile) icons are based on a diamond shape, yellow (unknown) icons are clover-leafed, and green (neutral) icons are square.

This data set thus has a serious disadvantage in that key stimulus properties are not orthogonal, but a considerable advantage in that the stimuli are undoubtedly of realistic complexity and subtlety. The choice of color as the code for the most important Origin property was based on long-standing results in human factors on the effectiveness of color coding in visual search; less salient properties are used for coding information that would be useful once an icon was fixated.

Stimuli were presented on a 17" color monitor with a resolution of 1024 x 768. The stimulus display occupied an 12.5" by 8.75" area. There were 48 possible positions for the searched-for icons on the display, each occupying a 0.5" by 0.5" area (excluding directional leader), with the icon positioned within the area in a randomized manner to avoid forming a strict grid configuration. The probe position was counted as a 49th position in the analysis.

These data are a subset from a large study that compared different forms of display "decluttering" that should improve visual search performance. Some of the icons were replaced with symbols that would be less distracting, but still informative. The subset of the data used here represented a baseline condition in which the removed symbols were replaced with an uninformative grey dot. Because the declutter manipulation was intended to be a realistic one in which irrelevant icons would be removed, the icons remaining were relatively similar to the probe icon, having a specified number of features in common. There were always three distractors that differed in only one of the three features from the probe. Most of the remaining distractors differed in two features, and the remainder differed in all three. The four conditions were 75, 50, 25, and 0% of the icons being decluttered – removed -, leaving respectively 12, 24, 36, and 48 icons remaining to be searched, randomly distributed over the display. Figure 2 is thus a 0% decluttered display containing 48 icons to be searched.

Procedure. The eye-tracking system used to collect data was the EyeLink System (SR International), which consists of a lightweight headset with three cameras. Two cameras record left and right eye; the third camera monitors head movement. Observations of eye movement and pupil activity were sampled at 250 Hz, providing 15,000 observations per minute for each eye. To record data, each undergraduate student participant underwent a short calibration procedure lasting 3-5 minutes, received instructions, and then proceeded to work through the task.

During a trial, the position of the eye was recorded every 4 milliseconds, and classified in which of the 49 regions of the display the current position was in. A duration of 80 ms or more in the same region was classified as a fixation on the display location. The location, start time, and end time of each fixation in each trial was computed, and combined with coded identity of the icon (or gray dot) at that location. After eliminating trials whose data were missing or ambiguous, the 21 participants with nine trials per declutter condition produced between 153 and 183 trials in each declutter condition. The fixations were then classified by type of match to the probe (e.g. *Origin* match, *Platform* match, *Target* (complete) match) and the mean number and duration of these classified fixations formed the basic data used in this study. Also calculated were mean latencies of the two responses in each condition.

The Model

Given the general machinery of the EPIC architecture, building a model for this task is relatively simple; task strategies and parameter values were explored iteratively to find one that fits the data reasonably well. Due to lack of space, details of this process are omitted, the task strategy implemented by the model's production rules will be only verbally summarized, and specific parameter values mentioned only when especially important.

At the beginning of the trial, the model fixates the probe position and waits for the probe to appear. It stores the color (Origin), shape and text label (Platform), and leader orientation (Direction) of the probe in working memory and waits for the main display to appear. It then begins the visual search process. This is implemented with sets of production rules that execute in two threads (see Kieras, in press). The first nominates objects to look at in anticipation of the next eye movement, and then moves the eye to a chosen candidate object as soon as possible. In the meantime, another thread evaluates the current candidate as soon as all of its necessary visual properties are available. An object is chosen as a candidate only if it has not been checked yet, and in the following order of descending priority: Encoding failed on a previous fixation on the object; an object that matches all of the probe attributes; an object that matches the probe color; an object chosen at random. After the candidate is chosen, the eye is moved to it. When the properties of the candidate become fully available, the candidate-choosing process is suspended, the candidate object is marked as checked, and then if it matches the probe, a mouse point movement is launched, followed by a button punch movement. The candidate-choosing process is then resumed. The result is that the model quickly and efficiently moves the eye from one candidate object to another, taking advantage of whatever properties about the objects are currently available.

The overall shape, Platform, and Direction information was assumed to be available only if the icon was foveated, but the color is available over an area up to several degrees in radius, depending on the display density. A good fit was obtained with the availability of the icon color for the four levels of declutter set at 9, 8, 7.5, and 7 degrees radius respectively. Any single value that produced a good fit at

one declutter level produced gross misfits at the other levels. Thus color provided visual guidance depending on display density.

The model assumed a large and reliable memory for which objects have been previously fixated or checked. This was motivated by the results in older eye movement studies such as Barbur, Forsyth, & Wooding (1993) , who reported a very low rate of repeat fixations on objects during visual search, and similar newer results of Peterson, Kramer, Ranxiao, Irwin, & McCarley (2001). Peterson et al. explained these few repeat fixations as a result of occasional encoding failures; with some small probability, the properties of an fixated object would fail to be encoded; when this was noticed, which tended to be quite soon, an eye movement would be made back to the object. The above task strategy implements this suggestion in the model with an encoding failure probability for the most detailed property, Platform, of 0.1.

Results

Availability Effects

Figure 3 shows the observed and predicted response times and number of fixations for each declutter condition; recall that decreasing declutter means increasing number of icons to be searched. In all graphs, observed points are plotted with solid symbols and lines, predicted with open symbols and dashed lines; the vertical brackets are 95% confidence intervals. The number of fixations is predicted very closely, as is the time for the first response. Note that the fourfold increase in number of objects to be searched resulted in only about a twofold increase in the number of fixations required, and even less of an increase in the time required, especially for the first response. The second response is consistently overpredicted, suggesting that some correction is needed in the assumptions about how long the search for the second target must wait while the first response is being executed.

As shown by the color availability parameter values, the effect of a more cluttered display is to reduce somewhat the area over which the most salient stimulus property is available. Note that decreasing the density of a display increases the average eccentricity between objects, making an object less available, but also decreases the average crowding, making an object more available. The two effects of density might partially cancel each other out, resulting the observed relatively small effect.

An additional timing result unnecessary to show graphically is that the fixation durations on non-targets were accurately predicted to be about 250 ms and essentially constant over conditions; for targets, the duration was substantially longer, because the icon would be pointed to, but the duration was not as long as would be expected from typical mouse movement times.

For brevity, further results are mostly shown only for fixations on non-target objects, which are gray dots or icons that mismatch one or more features of the probe. Figure 4 shows observed and predicted proportion of fixations of
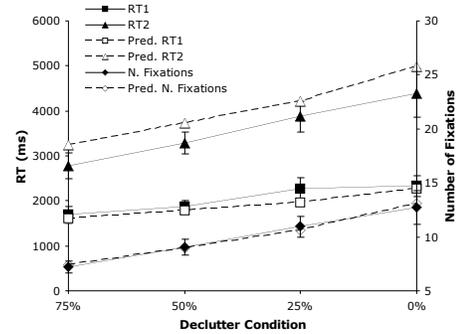

Figure 3. Response latency (RT, left axis) and number of fixations (right axis) for each decluttering condition.

different types on non-target icons that have one or more of the probe features for each declutter condition. The observed proportion of fixations on non-target icons of the same Origin increases with display density, and is quite high, at an average of 32% of the fixations; if fixations on Target icons were included, an average total of 66% of the fixations would be to objects that match on Origin.

However, the model overpredicts the proportion of non-target Origin fixations by about 0.14 on the average; the color-based visual guidance in the model was too dominant compared to the data. However, the proportion of fixations on non-target icons that match on one or more of the other target attributes (e.g. Platform) is correctly predicted to be low, about .12 on the average, and constant, showing that they do not guide the visual search significantly.

Figure 5 shows the proportion of fixations on objects that match none of the probe features. Although the model is fairly close for the gray dots, it clearly underpredicts the proportion of fixations on non-matching icons, corresponding to the too-frequent fixations on Origin matches.

Also shown in Figure 5 is the proportion of fixations on non-target objects that were previously fixated. The small proportion, 5%, of repeat fixations corresponds to the above-cited results, increases only slightly with the number of icons, and is fairly accurately predicted.
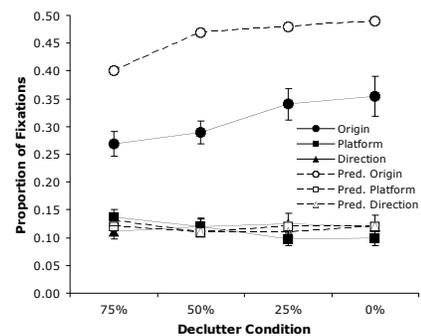

Figure 4. Observed and prediction proportions of fixations in each declutter condition for icons matching on at least one of the probe features.
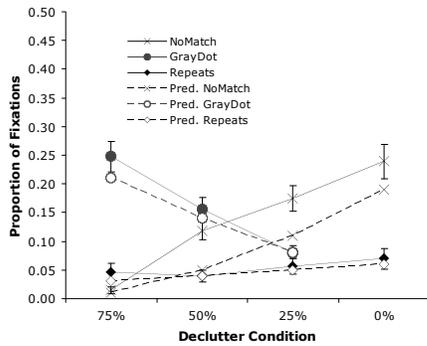
Figure 5. Observed and predicted proportions of fixations on objects matching none of the probe features, and on objects previously fixated (Repeats).
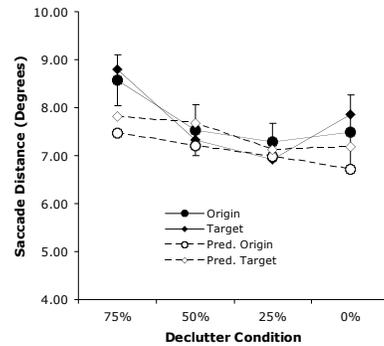

Figure 6. Observed and predicted saccade distances to arrive at Target icons or icons matching on Origin.

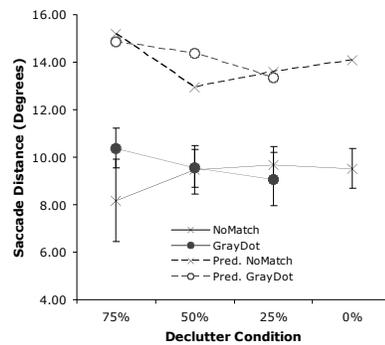

Figure 7. Observed and predicted saccade distances to arrive at objects not matching any probe features.

Figure 6 shows the saccade distance in degrees made to a fixation on icons that are a Target or match on Origin. Except for the least dense condition, the model makes saccades of about the same size as the data, reflecting how the availability of the color property in the model biases the next fixation to be nearby, where the color can be seen.

However, the saccade distance to non-matching objects is seriously overpredicted, as shown in Figure 7. In these cases, the model has picked the next object at random from the whole display, while in the data, such saccades are only somewhat longer on average than color-guided ones. This suggests either that the selection should be biased by distance, even though there is no limited-available property to guide it, or that objects out in the periphery can not be distinguished well enough to act as a saccade target; in other words, contrary to what the architecture assumes, perhaps "objectness" is also subject to availability.

Memory in visual search

A key component of this visual search model is a memory for which objects have been previously fixated. Wolfe (2003, Horowitz & Wolfe, 1998) has argued that visual search has little or no memory, but rather than being based on eye movements, this claim was based on an RT paradigm in which the display was modified during the search, making memory useless even if present (see von Muhlenen, Muller, & Muller, 2003 for more discussion of the strategic aspects of the task). The present data shows that repeat fixations occur at a very low rate, even for a large number of objects and fixations, as also observed by Peterson et al. (2001), and the model used their suggestion that the repeated fixations were due to stimulus encoding failures, rather than an unreliable memory for previous fixations. To investigate the need for a reliable fixation memory, the model was modified so that it had neither encoding failures nor a memory for which objects had been previously fixated. However, it could not locate the targets in a plausible amount of time or number of fixations unless color was available almost everywhere to provide guidance to the search – in effect assuming a homogenous retina, contrary to the active vision concept. Interestingly, this model

produced fairly accurate predictions of response times and number of fixations (see also St. John, et al., 2006), suggesting that these overall measures are not adequate to test different theories of visual search. But among other problems, the no-memory model predicted a 25% repeat fixation rate compared to the 5% observed in the data! This result disqualifies the no-memory model, meaning that a reliable memory for fixated objects is a critical component of a realistic model of visual search.

Conclusion

Practical conclusions. The data presented here are a useful case: although the stimulus properties are not ideally orthogonal, they are realistically complex, and the numerosity and density manipulations of the display are representative of the visual search problems in important practical tasks. Models based on a cognitive architecture that represents the visual availability and oculomotor mechanisms involved in visual search can account well for many important features of this data, meaning that the specific parameter values and architecture can be used in approximate models of human visual search performance for practical application in system design.

Methodological conclusions. The results suggest that visual search researchers should focus more on accounting for details of the eye movements – the fact that the no-

memory model could do reasonably well in predicting RT and overall number of fixations, and even the present model could do so quite well, even though some specific aspects of the fixations were incorrect, shows that these overall measures are not in fact sensitive enough to serve as a basis for testing theories of visual search.

Theoretical conclusions. The model predictions had enough quantitative detail that some aspects of the model can be clearly identified as incorrect in ways that have interesting implications for revisions to the architecture of the visual system. But the most important theoretical implication is that the memory for previous fixations is highly reliable and capacious; current cognitive theory has to satisfactorily explain this long-observed but unacknowledged memory system – it does not fit into any current cognitive architecture, at least not in any obvious way. In the present model, this memory was represented in a purely ad-hoc way; further work is needed to determine whether it could be represented as simply retention of the properties of those objects that constantly remain in the visual field. For example, suppose that the foveally available properties of a fixated icon persisted in visual memory for a relatively long time after the eye was moved away, and the next saccade target is chosen from those with unknown values of these properties. The effect would be to eliminate repeat fixations without the need of a special-purpose memory mechanism for which objects have been fixated.

Acknowledgments

References

Anstis, S.M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision research*, 14, 589-592.

Barbur, J.L., Forsyth, P.M., & Wooding. D.S. (1993). Eye movements and search performance. In D. Brogan, A. Gale, & K. Carr (Eds.) *Visual Search 2*. London: Taylor & Francis. 253-264.

Bouma, H. (1978). Visual search and reading: Eye Movements and the functional visual field: A tutorial review. In J. Requin (Ed.), *Attention and Performance VII*. Hillsdale, NJ: Lawrence Erlbaum Associates. 115-148.

Byrne, M. D. (2003). Cognitive architecture. In J. Jacko & A. Sears (Eds), *Human-Computer Interaction Handbook.* Mahwah, N.J.: Lawrence Erlbaum Associates. pp. 97-117.

Bouma, H. (1978). Visual search and reading: Eye Movements and the functional visual field: A tutorial review. In J. Requin (Ed.), *Attention and Performance VII*. Hillsdale, NJ: Lawrence Erlbaum Associates. 115-148.

Department of Defense (1999). Department of Defense interface standard: Common warfighting symbology: MIL-STD-2525B. *Reston, VA: DISA/JIEO/CFS.*

Findlay, J.M., & Gilchrist, I.D. (2003). *Active Vision*. Oxford: Oxford University Press.

Horowitz, T.S. and Wolfe, J.M. (1998). Visual search has no memory. *Nature,* 394, 575–57

Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. II. Hue and saturation. *Journal of the Optical Society of America*, **67**(2), 202-207.

Kieras, D.E. (2003). Model-based evaluation. In Jacko, J.A. & Sears, A. (Eds) *The human-computer interaction handbook*. Mahwah, New Jersey: pp. 1139-1151.

Kieras, D.E. (2004). EPIC Architecture Principles of Operation. Web publication available at ftp://www.eecs.umich.edu/people/kieras/EPIC/EPICPrinOp.pdf

Kieras, D. (in press). Control of cognition. In W. D. Gray (Ed.), *Integrated models of cognitive systems*. New York: Oxford University Press.

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.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, **104**, 3-65.

von Muhlenen, A. Muller, H.J., & Muller, D. (2003). Sit-and-wait strategies in dynamic visual search. *Psychological Science*, **14** (4), 309-314.

St. John, M., Marshall, S. P., Knust, S. R., & Binning, K. R. (2006). *Attention Management on a Geographical Display: Minimizing Distraction by Decluttering* (Technical Report No. 06-01). San Diego, CA: San Diego State University, Cognitive Ergonomics Research Facility

Peterson, M.S., Kramer, A.F., Ranxiao, F.W., Irwin, D.E., & McCarley, J.S. (2001). Visual search has memory. *Psychological Science*, **12**(4), 287-292).

Toet, A. & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, **32**, 1349-1357.

Wolfe, J.M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Science*, **7**(2), 70-76.