Adding Resolution to an Old Problem: Eye Movements as a Measure of Visual Search

Gregory J. Zelinsky[1] **Rajesh P.N. Rao**[2] **Mary M. Hayhoe**[1] and **Dana H. Ballard**[2]
[1]Center for Visual Science and [2]Computer Science Department
University of Rochester
Rochester, NY 14627-0270
gzelinsk@casper.beckman.uiuc.edu, rao@cs.rochester.edu
mary@cvs.rochester.edu, dana@cs.rochester.edu

The process by which we locate a target in a cluttered visual scene has been the topic of fierce debate for the past several decades. While the obvious reason for this debate is that many competing models do an excellent job of describing the basic search phenomenon (an increase in manual reaction time [RT] with the number of objects in the display), a more distal cause may be linked to a limitation of the dependent measures commonly used to study search. For example, one popular class of models liken search behavior to the movements of a fixed-diameter beam or spotlight (Treisman *et al*, 1980, Treisman, 1988). The more items in a display, the more times this spotlight must move, resulting in longer search times. Note however that this very specific theory about the spatial evolution of search over time is being inferred from a RT result without any direct measure of this spotlight or its movements. This failure to directly measure the theoretical construct in question opens the door for other models of search to offer their own explanations for why RTs increase with set size. In this sense a RT measure simply lacks the resolution needed to unambiguously describe the spatiotemporal changes occurring during search.

In response to this resolution problem, three experiments used the sequence of saccades and fixations accompanying search to more directly describe the search process. Stimuli for the first experiment were color images of either 1, 3, or 5 realistic objects arranged on an appropriate surface (*i.e.*, toys on a crib, tools on a workbench, etc.). Each of these objects subtended about 2.5 degrees of visual angle, and appeared initially at 7 degrees eccentricity from a fixation cross. The entire scene subtended 16 degrees. The six subjects participating in this experiment had to indicate the presence or absence of a designated target object in 360 of these scenes by making a speeded keypress response.

Based solely on the RT data, it would be possible to frame a fairly strong argument for a serial search process underlying this task. Increasing the number of objects in the search display yielded a concomitant increase in RT, and the rate of this increase for the target-absent trials was about 1.5 times the slope observed in the target-present data. However, when the pattern of eye movements are considered, evidence for a different search process emerges. Rather than a sequence of saccades directed to individual items in the display, initial saccade landing positions clustered near the center of the images, even though no objects ever appeared at this location. This center-of-gravity averaging behavior (Coren & Hoenig, 1972, Findley, 1982) was most pronounced in the 3 and 5 item displays where the mean distance between the endpoints of these saccades and the target was approximately 5 degrees. In the single item trials, this measure of initial oculomotor error decreased significantly to 3.1 degrees. Targeting accuracy improved markedly by the second saccades, mainly due to gaze shifting to the side of the display containing the target. Second saccade endpoint error at a set size of 1 was 0.8 degrees and increased linearly to 2.7 degrees at the largest set size. The slope of this error function flattened by the third, and typically last, fixations.

Instead of the serial movements of a spotlight, we interpret this eye position data as evidence for a global-to-local search strategy. According to this interpretation, the inaccurate first saccades reflect an initial distribution of the search process over a broad region of the scene, with each subsequent saccade describing a more restricted region until only the target is selected. Recent work in our lab has shown how a simple color and spatial filtering computation also unfolding over time can implement such a search dynamic and parsimoniously account for the eye data results (Rao *et al.*, 1995). Early in the search process there may be many parts of a realistic scene that can be easily confused with the target, thereby forcing the search computation to be distributed over much of the image. Programming an eye movement at this stage in the search process would allow each of these spurious candidates to contribute (weighted by target similarity) to the spatial computation, giving rise to the averaging tendency observed in the initial saccades. Over time, as the decision process converges and less likely candidates drop out of the computation, the search region narrows and gaze moves steadily towards the target.

To test this interpretation, two additional experiments varied the signal-to-noise characteristics of these search displays while again monitoring eye movements. We predicted that a lower signal-to-noise ratio would cause more points in the scene to match the target, yielding increased averaging behavior and a more difficult search task. Likewise, removing background noise from the scene should reduce the number of matching points and attenuate oculomotor averaging. To create stimuli with a lower signal-to-noise ratio, we simply removed chromatic information from the original color scenes. To make stimuli with a more favorable signal-to-noise ratio, the complex backgrounds from the original scenes were replaced with uniform colored fields. As predicted, these manipulations had opposing effects on oculomotor averaging. Initial saccades in the single item grayscale trials landed almost a degree further from the target than in the corresponding color scenes, whereas making the background uniform resulted in a 1.5 degree increase in first saccade accuracy relative to the original single item data. Initial saccade endpoint errors for the 3 and 5 item displays remained essentially unchanged across all three experiments, demonstrating the robustness of this averaging tendency.

To conclude, the greater spatio-temporal resolution of an oculomotor measure allows the search process to be viewed more directly and in finer detail than what is possible from a manual RT. Note also that this direct assessment of search behavior makes it unnecessary to speculate about an ill-defined attentional process. The only assumption being made here is that eye movements are aligned with search during free viewing (Gould, 1973, Jacobs, 1986, Zelinsky & Sheinberg, in press). Based on this oculomotor evidence, search in these realistic tasks can be better described as a global-to-local behavior rather than a serial process. Furthermore, the extent to which search is globally distributed over the display and the timecourse over which this process becomes more local appears to be influenced by the signal-to-noise characteristics of the scene.

References

Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, A. (1988). Features and Objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology, 40A*, 201-237.

Coren, S. & Hoenig, P. (1972). Effect of non-target stimuli upon length of voluntary saccades. *Perceptual and Motor Skills, 34*, 499-508.

Findlay, J. (1982). Global visual processing for saccadic eye movements. *Vision Research*, 22, 1033-1045.

Rao, R., Zelinsky, G., Hayhoe, M. & Ballard, D. (1995). Modeling saccadic targeting in visual search. *Advances in Neural Information Processing*. MIT Press.

Gould, J. (1973). Eye movements during visual search and memory search. *Journal of Experimental Psychology, 98*, 184-195.

Jacobs, A. (1986). Eye-movement control in visual search: How direct is visual span control? *Perception & Psychophysics, 39*, 47-58.

Zelinsky, G. & Sheinberg, D. (in press). Eye movements during parallel/serial visual search tasks. *Journal of Experimental Psychology: Human Perception and Performance.*