GUIDED SEARCH 3.0

A Model of Visual Search Catches Up With Jay Enoch 40 Years Later

JEREMY M WOLFE *Harvard Medical School and Brigham and Women's Hospital*
GREGORY GANCARZ *Boston University*

Abstract

Forty years ago, Enoch reported a central bias in eye movements made during visual search tasks. Here we discuss how our Guided Search 3.0 model incorporates that observation into a general account of visual search behavior.

Visual Search

Vision research has a cyclical component. Important issues fade into the background for a while, only to reappear when a new generation takes them up. In this paper, we present our most recent work on visual search. Some of the core ideas in this work are prefigured in papers written by Jay Enoch some 4 decades ago(1, 2). Enoch had subjects search for targets in aerial photographs and found that Ss were biased to search the center of the image. Understanding visual search in natural images of this sort has proven to be a difficult topic. Most of the progress in the last generation has been made with more easily analyzed but less realistic stimuli. In these laboratory tasks, the independent variable is the number of items (set size) and one dependent variable is the reaction time (RT) - the time to find the target or to determine that no target is present. Different tasks produce characteristically different results. In feature searches (Fig. 1a), the target is defined by one basic feature (e.g. orientation). RT is roughly independent of set size. Adding more horizontal lines does not slow search. In conjunction searches(1b), the target is defined by a combination of 2 (or more) basic features and RT increases about 5-10 ms for every additional distractor. Targets in serial searches (1c) are defined by the spatial relationship of features. Here each "T" and "L" is composed of a vertical and a horizontal line. Each additional "L" adds 20-30 ms to RT on a target present trial - twice that, 40-60 ms, on a target absent trial. Cases like 1c can be called "serial searches" because the data are consistent with a serial, self terminating search through the items at a rate of 20-25 HZ (3)(4).

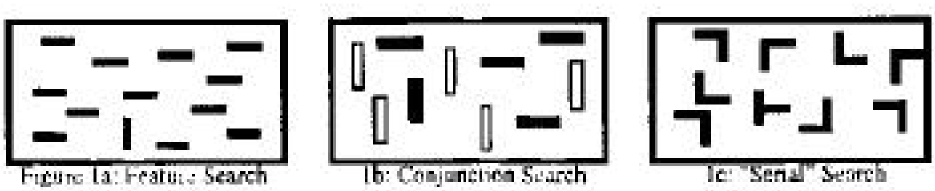

Figure 1a: Feature Search 1b: Conjunction Search 1c: "Serial" Search

Performance on these search tasks can be understood in Neisser's (5) framework. He distinguished between preattentive processes that could perform simple tasks across the entire visual field in parallel and attentive processes that perform more complex tasks only in a limited area at one time. Treisman applied this idea to visual search in her Feature Integration Theory (6). It held that preattentive feature processes could perform feature searches in parallel, while attention from item to item was required to do all other searches. Subsequent research showed that conjunction searches were too efficient to be explained as purely serial, attentive searches (7). Wolfe's Guided Search (GS) model accounted for this efficiency by proposing that preattentive feature processes could guide the deployment of attention in conjunction search. The basic idea is shown in Figure 2. No preattentive process can identify black vertical conjunctions but a color process can mark all black items and an orientation process can mark all vertical items. If the output of those two preattentive processes is combined into an attention-guiding *activation map*, attention will be most strongly guided to items that are both black and vertical - the target.

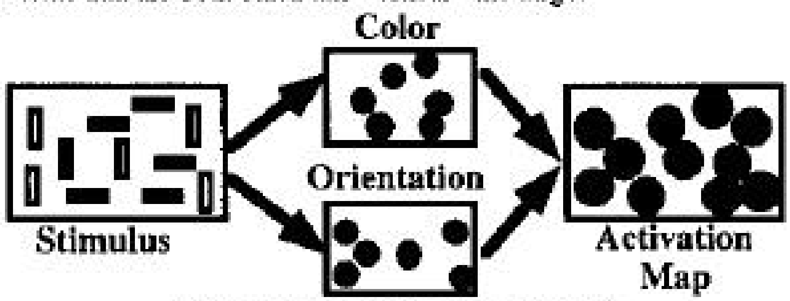

Figure 2: The basic idea of guided search.

Guided Search 3.0

GS (8) and its revision, GS2 (9) successfully modeled a wide range of laboratory search tasks. However, the eventual goal of a model of visual search must be to handle real world images like those that interested Enoch in the 50's. GS had ignored two factors central to Enoch's interests. First, the eyes move during real-world visual search. Most searches involves an interplay of covert attentional movements and overt eye movements. Second, visual processing is much more detailed at the fovea. Presumably, eye movements in search exist to direct the fovea toward regions of interest. Enoch found that viewers spent much more time looking at the center of an image. More recently, it has been shown that visual search targets are found faster and more accurately near fixation than in the periphery (10). GS3, the newest version of our model (Fig. 3), incorporates eye movements and eccentricity effects as part of an effort to make GS a more realistic model of human search behavior.

The simulation of GS3 starts with a Gray Scale Image as its input. Next the image is processed by an array of On and Off-Center Units, approximating retinal ganglion cell response. These provide input to Preattentive Feature Maps for brightness and orientation. These maps model the representation of the stimulus in V1 by a complex log transform (11). This realistic overrepresentation of the fovea creates a role for eye movements and a substrate for eccentricity effects in search.

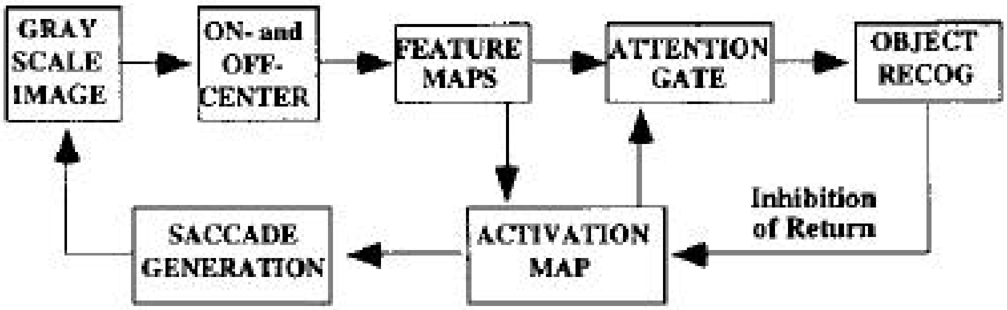

Figure 3: The architecture of Guided Search 3.0

In GS3, attention serves as a gate, allowing feature information from only one object at a time to reach higher processes like object recognition. This attentional gate is under the control of the activation map. Activity in the activation map is a weighted sum of activity in the preattentive feature maps. Thus, if the target is dark and vertical, information from "dark" and "vertical" feature maps is given a large weight. The activation map is a winner-take-all network that converges on a winner about every 50 ms. This allows featural information about new items to pass through to the identification stage at a rate of 20-25 Hz. If the selected item is not the target, feedback from the identification stage to the activation map inhibits that item, permitting a new item to win access to the identification stage. Since the eyes move, this inhibition of return needs to be in head-centered coordinates. Accordingly, the activation map is converted to these coordinates.

In order to move the eyes, a saccade map is created (12). It is the GS3 analog of the superior colliculus. Activity in this map is a blurred version of activity in the activation map. Every 200-250 ms, the eyes are moved to the point of highest activation in the saccade map. The spatial blurring of this map can cause short latency saccades to be directed to locations between nearby items if the eyes move before the activation map has picked a clear winner. This speed-accuracy trade-off is seen in human eye movement data (13). The GS3 architecture leads to a cooperative relationship between eye movements and attentional deployments. Because central portions of the visual field are overrepresented in the feature maps, activation tends to be highest for central items. Thus, during a fixation, attention is deployed to 4 or 5 central items. Assuming that these are not targets, their representations are inhibited in the activation map. As a result, when it is time for the eyes to move, they foveate new items. These are then examined by attention and this process repeats until the target is found or a target absent response is made.

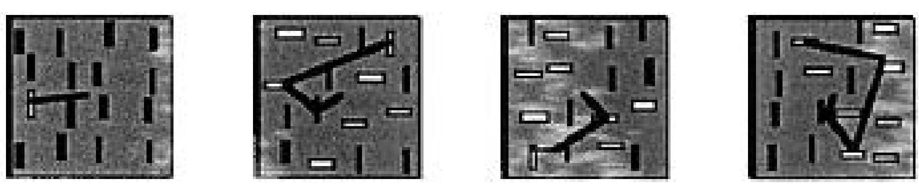

Figure 4: Typical GS3 scan paths for feature and conjunction search tasks.

The GS3 simulation produces data that mimic human data on a number of tasks. Searches for "white vertical" targets in gray scale images similar to those in Fig. 1 produce slopes of RT x set size functions that are comparable to human data. Moreover, the increase of RT with eccentricity is the same in the GS3 data as it is in human data. GS3 eye movements look similar to those recorded in human studies (14). Samples are shown in Fig. 4.

To summarize, in GS3, eye movements and covert attentional deployments work together to find a target in a representation with a realistic overrepresentation of the central visual field. Further refinement of the model will be required before GS3 can be asked to perform the photointerpretation tasks that Enoch's subjects did in the 50s. However, having moved away from real images and natural tasks for many years, we are heading back toward models of real world visual search tasks.

Acknowledgments: This work was supported by AFOSR 49620-93-1-0407. We thank Todd Horowitz and Sara Bennett for their comments on drafts of this paper.

References

1.Enoch JM. Natural tendencies in visual search of a complex display. Visual Search Techniques 1959;NAS/NRC Pub. 171:187-193.
2. Enoch JM. Effect of the size of a complex display upon visual search. J. Optical. Soc. Am. 1959;March:280-286.
3. Kwak H, Dagenbach D, Egeth H. Further evidence for a time-independent shift of the focus of attention. Perception and Psychophysics 1991;49(5):473-480.
4. Bennett SC, Wolfe JM. Serial search can proceed at 50 msec per item. Investigative Opthalmology and Visual Science. 1996;37(3):S298.
5. Neisser U. Cognitive Psychology. New York: Appleton, Century, Crofts, 1967.
6. Treisman A, Gelade G. A feature-integration theory of attention. Cognitive Psychology 1980;12:97-136.
7. Wolfe JM. Visual search. In: Pashler H, ed. Attention. London, UK: University College London Press, 1996:
8. Wolfe JM, Cave KR, Franzel SL. Guided Search: An alternative to the Feature Integration model for visual search. J. Exp. Psychol. - Human Perception and Perf. 1989;15:419-433.
9. Wolfe JM. Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin and Review 1994;1(2):202-238.
10. Carrasco M, Evert DL, Change I, Katz SM. The eccentricity effect: Target eccentricity affects performance on conjunction searches. Perception and Psychophysics 1995;57(8):1241-1261.
11. Schwartz EL. Spatial mapping in primate sensory projection: analytic structure and relevance to perception. Biological Cybernetics 1977;25:181-194.
12. Sheinberg DK, Zelinsky GJ. A cortico-collicular model of saccadic target selection. In: d'Ydewalle G, Van Rensbergen J, eds. Perception and Cognition. Amsterdam: Elsevier, 1993:333-348.
13. Findlay JM. Visual search: eye movements and peripheral vision. Optometry and Vision Science 1995;72:461-466.
14. Zelinsky GJ, Sheinber DL. Eye movements during parallel / serial visual search. J. Exp. Psychology: Human Perception and Performance 1996;in press.