Attracting Attention

Participants and Paper Titles

Patrick Cavanagh (Harvard) - Pursuing moving objects with attention
Ken Nakayama (Harvard) - Visual search and its relation to depth cues
Jeremy Wolfe (MIT) -Three Aspects of the Parallel Guidance of Visual Attention
Steven Yantis (Johns Hopkins) - Attentional Priority in Vision

Chair
Anne Treisman (Berkeley)

Introduction

The early stages of visual processing appear to operate in parallel across the entire visual field. There are retinotopic arrays of neurons that appear to support the processing of basic features such as size, color, orientation, and so forth (VanEssan and Maunsell, 1983) but not of more elaborate properties (e.g. conjunctions of two basic features - Treisman and Gelade, 1980; Wolfe, Cave, and Franzel; 1989; Treisman and Sato, 1990). Parallel processing is not feasible for all visual tasks. The brain is simply not large enough (e.g. Tsotsos, 1990). The parallel stage seems to be restricted to a few simple operations on a few basic features (e.g. Wolfe et al, 1990). For more complex visual tasks, the ability to perform these tasks is restricted to a portion of the visual field. For many tasks, the relevant portion of field is not fixed but can be selected by attention. Thus, the large-scale architecture of the visual system seems to consist of a parallel front end leading to a processing bottleneck. After the bottleneck, the field of action of various limited capacity processes can be moved about under attentional control. Experimental evidence suggests that those movements can occur at a rate of 15-25 movements per second (e.g. Sagi and Julesz, 1986) though this rate may be much slower if attention becomes "engaged" at one locus (Mackeben and Nakayama, 1988).

For real-world visual tasks, it is not practical to allow visual attention to wander about the visual field at random. To use this architecture efficiently, the parallel stage should guide the deployment of attention. In this fashion, attention could focus limited capacity processes where they are needed and not waste time on sophisticated processing of the wrong part of the visual field.

While the logic of having a high capacity, parallel stage guide subsequent, limited capacity processes seems clear, it is not entirely clear how that guidance is provided. As noted above, the parallel stage has very limited capabilities. In this symposium we intend to discuss the nature of those limitations and the mechanics of attracting attention.

Pursuing moving objects with attention.

Patrick Cavanagh
Dept. of Psychology, Harvard U, Cambridge, MA 02138

When an observer fixates a point in space but attends to a moving object, there is a clear perception of motion. There are two possible sources for that perception. It may be due to the selection of the low-level motion signals generated by the object or it may be derived from the signals generated by the attention system in order to track the object. This second possibility is very much like the efferent copy model for the perception of target motion during pursuit eye movements. The target does not move on the retina during the tracking eye movements but it is nevertheless perceived to move in space. According to the efferent copy model, this motion percept is derived from the eye movement signals required to track the target.

To study whether the perceived motion of objects tracked without eye movements is derived from a similar but covert signal, tracking experiments were run using counterphase gratings. A counterphase grating is the sum of two identical gratings moving in opposite directions, left and right, for example. The stimulus contains no net motion but if it is tracked with eye movements in either direction, a moving grating is seen. To eliminate tracking eye movements, a radial, counterphase grating was constructed within a circular annulus. A radial grating rotating clockwise was added to a radial grating rotating counterclockwise, producing a counterphasing grating.

When an observer fixated the center of the radial grating, no net motion was evident. However, if the observer attended to individual spokes (while still fixating the center), a clear impression of motion was produced in the direction of the attentional tracking and this direction could then be changed at will. Observers most often reported attending to a pair of symmetrically opposed spokes in the grating and following their motion around the annulus. While attending to a given pair of spokes, the remaining areas of the annulus did not appear to rotate clearly in either direction. With practice, attention could be directed to other patterns of spokes such as a triple-pointed "Mercedes-Benz" symbol or a four-point cross. The attended regions in these patterns of spokes were not contiguous, but all the spokes in the attended pattern rotated in the same direction.

These results demonstrate that attention can track moving elements of an ambiguous display and disambiguate one direction of motion in a compound stimulus. The ability to attend to different groupings of elements (pairs, three-pointed and four-pointed stars) suggests a higher-order motion process that uses simple forms as place keepers to group separate elements into a unit that can be tracked over time. This process may be related to the place-keeping process that Pylyshyn and Storm studied with fields of independently moving luminance disks.

A second experiment showed that the motion perceived during attentional tracking is based on a higher-order motion analysis that differs greatly from low-level motion in its relative sensitivity to luminance and color. A clockwise grating defined by luminance was added to a counterclockwise grating defined by color. The contrast of the luminance grating was adjusted until neither direction of rotation dominated when fixating the center of the annulus. At this point, a small increase in the contrast of the luminance grating produced an impression of global rotation in the direction of the luminance grating(clockwise) whereas a small decrease produced an impression of global rotation in the direction of the color grating (counterclockwise). With the two directions of motion balanced and eyes fixated at the center, observers could easily track the counterclockwise motion of the color grating when asked to attend to individual spokes. However, they could not track the luminance spokes rotating in the opposite direction. In fact, very little contrast (about 10%) was required in the luminance grating to reach the point of motion balance. In the presence of the color patterns, the 10% contrast luminance patterns were difficult to see and it is reasonable to assume that the ease of tracking pattern elements depends on the visibility of the patterns. On the other hand, visibility is not such an important factor for the low-level motion system. The results here, as well as previous results, show that a luminance stimulus produces a strong low-level motion signal even when it is barely visible whereas a color stimulus produces only a weak low-level motion signal even though it is highly visible.

When the low-level motion signals from the color and luminance gratings were balanced in the stimulus, no net motion was seen in either direction without attention to individual elements. Selection of low-level signals can therefore be ruled out as the source of the perception of motion during tracking because, based on low-level signals, it should be as easy to track the motion of the luminance elements as that of the color elements. The results therefore suggest that the impressions of motion during tracking must be derived from the signals that displace the tracking window. Indeed, although the rotating color gratings appeared to slow down or even stop when the colors were equiluminous, the motion of the color elements re-emerged when they were being tracked.

Visual search and its relation to depth cues

Ken Nakayama
Dept. of Psychology, Harvard U, Cambridge, MA 02138

Some conjunctive searches can be done in parallel. That is, there is no increase in reaction time as the number of items is increased. Other conjunctive searches (using different visual dimensions), however, are more difficult and reaction time increases for increasing set sizes. Our working hypothesis is that feature grouping among some visual dimensions is stronger than in others. In particular, those features which reliably co-vary with the distance of surfaces in real world scenes seem most likely to support grouping. As such, there is the tendency for parallel conjunctive search to be most prominent for tasks employing the dimensions of binocular disparity, size, spatial frequency, and motion as opposed to color and orientation/form.

In a new series of experiments designed to evaluate the nature of perceptual grouping in visual search tasks, Mary Bravo and I have shown very different findings depending on whether the observer has knowledge of the target color. We find that when the color of a target is unchanged from trial to trial (the observer is searching for a known target), performance is unaffected by the number of distractors. When the color of a target is changed from trial to trial (the observer is searching for an unknown target), reaction times decrease with increasing numbers of distractors. The importance of distractor number in the search for an unknown target is even greater when target and distractors are randomly distributed across two disparity planes.

Three Aspects of the Parallel Guidance of Visual Attention

Jeremy M Wolfe
Dept. of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139

As noted in the Introduction, visual processing can be crudely divided into two parts: A "front-end" capable of processing a few simple features in parallel across the visual field and a "back-end" consisting of more powerful processes that can operate over only a restricted portion of the visual field. These processes can be directed to different parts of the field by attention. To move the limited capacity processes in a sensible and efficient manner, attention should be *guided* by information collected in parallel. The Guided Search model (Wolfe, Cave, and Franzel, 1989; Cave and Wolfe, 1990) is an effort to understand that guidance in the context of visual search. In a visual search task, a subject looks for a **target** item (or items) among a variable number of **distractor** items. Various measures of performance may be used (e.g. Klein and Farrell, 1989; Bergen and Julesz, 1983). In our experiments, we follow Treisman (1988) in measuring reaction time (RT) as a function of the total number of items presented. Some search tasks yield highly efficient searches, independent or nearly independent of set size. The clearest examples are searches for a target defined by a simple feature (color, size, etc) presented among homogeneous distractors (e.g. a red target among green distractors) (Treisman and Souther, 1985; Duncan and Humphreys, 1989).

When simple feature information is of no use, RTs increase linearly with set size. The increase is twice as great for target-absent trials as for target-present trials. These results are consistent with a self-terminating search governed by serial movements of attention at a rate of 40-60 msec/item. (e.g. Sagi and Julesz, 1986). A search for a "T" among "L"s is an example of such a search. If the Ts and Ls can rotate, they can be distinguished only by determining the relative position of the two line segments. This task appears to require serial deployment of attention.

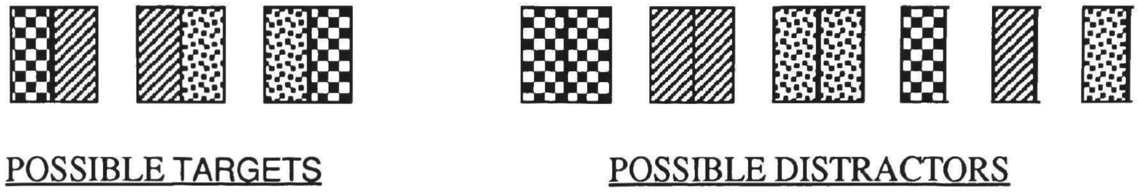
From the vantage point of Guided Search, the most interesting cases are those where simple feature information is present even if no single feature is adequate to differentiate target and distractors. An example is the search for a conjunction of two features. If the target is a red vertical item while the distractors are red horizontal and green vertical, no single feature can be used to discriminate targets from distractors. However, if attention can be guided toward "red" items and toward "vertical" items, the combination of these two sources of guidance could direct attention toward "red vertical" items. Guided search for a conjunction of two basic features can be much more efficient than a search where no guidance is possible. (Wolfe, Cave, and Franzel, 1989; Egeth, Virzi, and Garbart, 1984; Nakayama and Silverman, 1986; Treisman and Sato, 1990).

In the Guided Search model, attention is guided by the activation in the parallel stage. There are two components to parallel stage activation. Bottom-up activation is an accelerating, non-linear function of the mean of the differences between an item and other items in its neighborhood. Top-down activation is a function of the similarity between an item and the internal representation of the target. These activations are computed independently with each parallel feature module. Thus, in a search for a red vertical item, there would be relevant bottom-up activations from both color and orientation modules based on local variations in those features. Further, there would be independent top-down activation in the same modules based on the similarity of each item to "red" for the color module and "vertical" for the orientation module. All activations are summed and perturbed by noise.

Attention moves from location to location on the basis of total activation starting with the most active and continuing until the activation falls below some threshold. (See Cave and Wolfe, 1990, for details)

In this paper, we will discuss three aspects of the parallel guidance of attention:

1) Bottom-up activation can be based on local contrast, independent of stimulus identity. Consider the following task (Each pattern stands for a different color).



It is easy to find these targets because they contain a high chromatic contrast edge. It is not necessary for target colors to be unique or known. A simple high degree of local variation attracts attention.

2) Top-down activation is based on perceptual "categories". Top-down activation is based on the similarity between an item and an internal representation of target properties. Evidence from searches for a target orientation among two or more distractor orientations suggests that this internal representation is categorical (and not, for example, geometrical). Attention can be directed toward "steep" (or "shallow") items or toward "left" (or "right") tilted items but not toward targets specified by angle of orientation or even by the degree of steepness (i.e. it is not possible to attend to the "steepest" line).

3) Visual search can be performed as a sequence of guided operations. As discussed above, search for a conjunction of two features can be more efficient than an unguided serial search. For example, in a search for a black vertical item among black horizontal and red vertical items attention can be directed to the intersection of the set of all black items and the set of all vertical items. However, the following task poses a problem:

Now the target can be either white or black and either vertical or horizontal. Guiding attention to the intersection of the set of items that are black or white and the set of items that are vertical or horizontal is obviously useless. Instead, subjects appear to execute two sequential searches, first guiding attention to one target type and then, if no target is found, guiding attention to the other type. This yields RT x set size functions with shallow slopes but higher intercepts. That is, the double searches take somewhat longer but remain efficient.

ATTENTIONAL PRIORITY IN VISION

Steven Yantis
Department of Psychology
Johns Hopkins University

Coherent thought and action concerning visual object configurations require a control mechanism, visual attention, to select task-relevant subparts of the input image for rapid processing. This requirement arises because visual cognition is limited in capacity: there is often more sensory input available than the visual system can handle efficiently at one time. In this paper, I describe some new data that constrain the possible mechanisms for attentional capture, and I propose a model for attentional prioritization that is a hybrid of several recent architectures for visual selection.

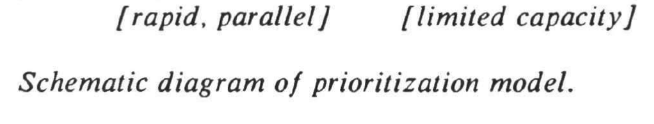
Experiments conducted in our lab reveal that spatiotemporal discontinuities (e.g., abrupt onset or movement) capture attention. In the experiments, subjects search a multielement array for a prespecified target element. When the target is presented with an abrupt onset and nontargets are presented without abrupt onset, then response times are rapid and do not depend on the number of nontargets in the display. When the target is among the objects without abrupt onsets, response times increase monotonically with the number of objects in the display. This suggests that the onset element is processed with high priority (e.g., if decisions are made serially, it is almost always processed first).

Attentional capture by abrupt onset is not absolute, of course. We have found that focussing attention on a spatial location that is likely to contain a target prevents capture by the appearance of an onset elsewhere. This is adaptively sensible, and suggests that there can be multiple criteria for selection, some that are more or less hard-wired and bottom-up (e.g., onsets) and others that are strategic and top-down (e.g., positional expectancies).

Our most recent data concern the mechanism by which attentional priorities are established and maintained when there are several high-priority and several low-priority objects in the image. We have analyzed response times when there are multiple onset and no-onset elements present in a single trial and found that the system queues or tags as many as four onset elements for attentional priority. These elements are processed earlier or receive more computational resources than lower-priority elements do. Preliminary data also suggest that a similar mechanism may operate in other stimulus domains when priority is specified by top-down selection criteria.

A model for attentional priority (see diagram) that can incorporate these results as well as other facts about visual selection incorporates a two-stage architecture, in which the first stage (Prioritization) rapidly and in parallel prioritizes objects in the image based on the extent to which various selection criteria are satisfied (including both bottom-up criteria like onset and top-down criteria like expected position). The importance of any given selection criterion is specified by a priority schedule, and priorities are assumed to be dynamically updated. The second stage (Identification) allocates limited computational resources based on the priorities assigned at the first stage and identifies objects via weighted stochastic sampling. To the extent the selection criteria in combination with the priority schedule effectively guide resources to task-relevant objects, search will be optimal.

A consensus is emerging that two-stage architectures of this type may account for a wide range of visual phenomena. Further empirical work in psychophysics, neuropsychology, and neurophysiology can place additional constraints on the detailed structure of these models.



Schematic diagram of prioritization model.

References

Bergen J. R. and Julesz B. (1983) Rapid discrimination of visual patterns. IEEE Trans on Systems, Man, and Cybernetics, SMC-13 857-863

Cave K. R. and Wolfe J. M. (1990) Modeling the role of parallel processing in visual search. Cognitive Psychology, 22

Duncan J. and Humphreys G. W. (1989) Visual search and stimulus similarity. Psychological Review, 96 433-458

Egeth H. E., Virzi R. A.and Garbart H. (1984) Searching for conjunctively defined targets. J. Exp. Psychol.:Human Perception and Performance, 10 32-39

Klein R. and Farrell M. (1989) Search performance without eye movements. Perception and Psychophysics, 46 476-482

Mackeben M. and Nakayama K. (1988) Fixation release facilitates rapid attentional shifts. Invest. Ophthalmol. Vis. Sci. (suppl), 29

Nakayama K. and Silverman G. H. (1986) Serial and parallel processing of visual feature conjunctions. Nature, 320 264-265

Sagi D. and Julesz B. (1986) Fast noninertial shifts of attention. Spatial Vision, 1 141-149

Treisman A. (1988) Features and objects: the fourteenth Bartlett memorial lecture. The Quarterly J of Experimental Psychology, 40A(2) 201-237

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

Treisman A. and Sato S. (1990) Conjunction search revisited. J. Exp. Psychol.:Human Perception and Performance, 16 ??

Treisman A. and Souther J. (1985) Search asymmetry: A diagnostic for preattentive processing of separable features. J. Exp. Psychol. - General, 114 285-310

Tsotsos J. K. (1989) Analyzing vision at the complexity level. Brain and Behavioral Sciences, ms ms

VanEssen D. C. and Maunsell J. H. R. (1983) Hierarchical organization and functional streams in the visual cortex. Trends in Neuroscience, 6 370-375

Wolfe J. M., Cave K. R.and Franzel S. L. (1989) Guided Search: An alternative to the Feature Integration model for visual search. J. Exp. Psychol. - Human Perception and Perf., 15 419-433

Wolfe J. M., Yu K. P., Stewart M. I., Shorter A. D., Friedman S. R.and Cave K. R. (1990) Limitations on the parallel guidance of visual search: Color X Color and Orientation X Orientation conjunctions. J. Exp. Psychol. - Human Perception and Performance, in press