

Spotlight on attention

Computer scientists have always emphasized that the brain represents the ultimate in parallel computer architecture, its highly interconnected neurons performing up to billions of operations simultaneously. This parallelism is often cited as the critical advantage of the brain with respect to our serial computers still based on a von Neumann architecture, performing single operations step-by-step. The construction of new parallel computers (such as the Connection Machine¹) with thousands of simple processors and the power to solve previously intractable problems in computational vision is therefore an especially exciting event for both computer and brain scientists. Ironically, just as the new technology of powerful parallel computers begins to close the gap between machine and brain, new psychophysical, anatomical and physiological findings^{2–6} suggest it might not be so wide: in certain simple but crucial tasks of early vision, the brain acts as a serial processor. These tasks are mediated by the 'spotlight of attention' which can scan the visual field independently of eye movements^{7–10}.

Treisman's psychophysical experiments^{11,12} demonstrate the nature of this spotlight: A subject rapidly recognizes the letter 'S' mixed into a random field of green 'Xs' and brown 'Ts'. The 'S' pops out at him, suggesting that its distinguishing feature (its shape) is tracked down independently of and in parallel with the mechanism that groups together letters of a common color. When color is the distinguishing feature, the same result occurs: a green 'X' pops out of a field of brown 'Xs'. This pre-attentive stage is fully consistent with the model of the brain as a parallel computer. But when a green 'T' in the same field is the target, the subject is slow to find it, and his slowness linearly increases with the number of objects in the field. The search for an object distinguished by 'conjunctive' features (color and shape) seems to be a serial, self-terminating scan of each spatial location.

The pop-out effect suggests the existence of separable features which satisfy Garner's three criteria¹³: they can be attended to selectively, processed independently and in parallel, and used as distinct tests for similarity.

Color, line orientation, line ends (terminators), and possibly crossings have been diagnosed as separable features in pop-outs and similar experiments on texture discrimination by Julesz^{14,15}. Direction and speed of motion and stereoscopic disparity, although not tested for explicitly in such experiments, are expected to be separable features, from physiological evidence: they are detected by tuned neurons. Barlow¹⁶ speculated that such features (which he called 'linking features') may be analysed locally in a topographic map of the visual field, but then may be 'sent' to non-topographic feature maps, where features of similar dimensions are grouped together. The question which follows naturally from this idea is one which Treisman and Julesz address in their ingenious experiments, and which resonates with artificial intelligence: How are separable features, having been teased apart in the primary analysis of an image, put back together to make coherent objects? Or, how are feature maps put in register to restore the topography of a scene?

Barrow and Tenenbaum¹⁷ suggested that local computations are carried out

in parallel and the results represented in distinct maps, 'intrinsic images', each of which separately encodes a parameter such as orientation, reflectance, or intensity. The images are in spatial register with the original image and with each other. Marr's 'primal sketch'¹⁸ condenses the results of local parallel computations into a single map, grouping parameters according to location in the original image. Both computational methods analyse the features of an object by parallel processing which resembles that of the pre-attentive stage in psychophysics, and in both methods spatial location serves as a passive link to reconstruct the scene.

Minsky¹⁹ answered the artificial intelligence question differently. He suggested that a fixed set of pattern-recognition techniques scan each image location separately and serially, computing parameters and linking them together in the same set of operations. Psychophysical experiments indicate that in human vision, the spotlight of attention acts in the same way, focusing on each location in series. When the spotlight is prevented from scanning normally, as when attention is overloaded, one may expect illusory conjunctions – combinations of features which do not actually exist in one object – to occur. Treisman²⁰ finds that these conjunctions appear to the subject when his attention is diverted to another task, or otherwise overloaded.

The plausibility of a spotlight of attention in the brain which operates as the above model suggests has been strengthened by a recent emphasis on functional localization in cortical anatomy²¹ and by new results in the physiology of attention, particularly selective visual attention.

These results raise but do not fully answer several independent questions:

how are feature maps constructed physiologically and how are they represented anatomically? if the maps are constructed from spatially parallel processing, are the links between them necessarily forged by serial processing? and what is the physiological mechanism underlying the spotlight?

At least 12 distinct visual areas have been identified in monkey cortex⁴, each of which contains topographic representations of part or all of the visual hemi-field and may be characterized by one or more distinct visual functions. Present evidence suggests, for example, that visual areas MT, MST and 7a are specialized for motion analysis: in MT the preponderance of cells are selective for form or color; in MST and area 7a, cells are also direction-selective but perform more sensitive tests on motion stimuli than MT cells²². In contrast, areas V4, VP, and IT seem specialized for color and form analysis. V1, with its variety of cells selective for wavelength, orientation, direction, speed or disparity, has been described as the segregator of visual functions. In the current picture, the functional pathways which analyse motion and color and form, emanate in parallel from V1 and terminate in the higher-order, specialized areas.

Although it is a vast oversimplification to equate different areas with different 'feature maps', the anatomical segregation of distinct visual functions does support the evidence for parallel processing of separable features. But are parallel functional pathways necessarily integrated by serial processing, and if so, at what level? The very facts which muddy the distinctions between hierarchical and parallel structures in cortical anatomy provide a hint: the connections between areas in a functional pathway are not in a strictly forwards direction, but are also backwards and lateral, suggesting both feedback and crosstalk within the pathway. Because the cell to cell connections have not been functionally characterized, it is possible that a pop-out mechanism and even a spotlight search is imbedded within the feedback and crosstalk.

The physiology underlying pop-out effects and spotlight activity is virtually unexplored although several models have been proposed. Crick⁶ suggested that a specific physiological mechanism underlies the spotlight: the bursting of a subset of active thalamic neurons, which creates a conjunction

of cortical neurons, which in turn signals a coherent set of features. Although the existence of several visual areas in the thalamus makes it unlikely that the spotlight activity is located primarily in the gating of input to V1 by the perigeniculate nucleus, as Crick originally proposed, the idea that feature conjunctions are mediated by transient neuronal assemblies is still feasible and attractive. Koch and Ullman²³ have proposed a simple mechanism, performed by an abstract network of neurons, which automatically yields both pop-out effects and serial scanning.

Recent physiological experiments have revealed the importance of attentive mechanisms in modulating neuronal response, but they have not specifically addressed the questions raised by the spotlight model. Although experiments have explored far beyond the effects of general unanaesthetized arousal, they have yet fallen short of providing direct evidence for (a) pop-out mechanisms or (b) a spotlight which scans the visual field and mediates operations such as conjunctions between features.

A typical cell in the superficial layers of the monkey superior colliculus shows an enhanced response just before the alert animal saccades to a target in its receptive field⁵. A similar enhancement also occurs in the frontal eye fields and in prestriate cortex. Most strikingly, attention-mediated modulation has been demonstrated in area 7 of the posterior parietal lobe^{24,25}, where neuronal response is enhanced not only when the animal saccades to a target in the tested receptive field, but also when the animal touches the target without making an eye movement, or, in general, whenever the animal attends to the target, regardless of how it attends. On the basis of these findings Mountcastle suggested that mechanisms in area 7 are responsible for 'directing visual attention' to selected stimuli.

Haenny, Maunsell and Schiller²⁶ recently demonstrated attentional mechanisms in V4. In an alert monkey trained to detect and signal an agreement between oriented tactile and visual stimuli, the responses of orientation-specific cells in V4 varied: some were specific for the visual cue independently of the tactile one; some were specific only for a single pair of matching visual and tactile stimuli; and some were specific for the tactile cue,

independently of the visual one. These results suggest that attention involves higher-level processing in which low-level information from different sensory modalities is combined and encoded in an abstract representation.

Although a number of visual areas, including V1, V2 and MT, do not seem to show enhanced responses associated with performing specific visual tasks, it is quite possible that this lack simply reflects experimental limitations. If these limitations were overcome, more direct experiments could be performed. A pop-out effect may be detected, for example, by recording from a cell stimulated by a preferred feature in its receptive field (for example a vertical bar), while randomly changing the field surrounding the bar, so that some times it is the odd man out among many horizontal bars, some times simply one among many vertical bars. One might also expect to find neurons which, in an inattentive animal, are responsive to a single preferred feature, but, in an attentive animal, are responsive only to certain conjunctions between features.

A similar experimental paradigm has been described by Braitman²⁷, although it does not strictly match these suggestions. Neurons in the inferotemporal cortex responded differently to the same physical stimulus, a colored checkerboard, depending on whether the monkey was made to attend to the color or the size of the squares. These results perhaps come the closest to demonstrating specific attentional effects on a neuronal level, and although further and more direct results will certainly be difficult to obtain, they are essential to address critical questions on the anatomy and physiology of attention.

The theoretical work on computational geometry initiated by Minsky and Papert²⁸ suggests a more complex and wide-ranging role for the spotlight of attention than in the conjunction of features. They demonstrated in their work on perceptrons that certain deceptively simple visual operations such as determining the connectedness of a contour could not be performed efficiently by parallel processing, but instead required serial processing. Their work implies that sequentiality in the brain is not the result of a capricious choice of evolution, but a requirement imposed by the intrinsic nature of visual computations (see also Ref. 29). The spotlight of attention may be essential not simply as the link

to join different feature maps (or intrinsic images), but as a processing focus to scan the image or its maps and perform certain abstract operations on each location. As Ullman suggested³⁰ the spotlight of attention in this role would underlie the computation of spatial relations, such as 'inside-outside', in addition to the simple conjunction of features or parameters.

The theoretical conclusion that serial processing is needed for some simple visual tasks follows from the fact that an enormous explosion of connectivity would result if information from the retina were sent to a single parallel network in the cortex. In a sense, each processing unit in the network would have to be connected to each point in the whole visual field to make a decision about the connectedness of a geometric figure, for example. There is, yet, an alternative and intermediate possibility to the strictly serial processing that the spotlight we have discussed may perform. In particular, we suggest that at each instant of time only a small part of the image may be 'routed' to a small processor specialized for the task at hand. Each processor itself may be highly parallel, but the routing to it is necessarily serial.

These parallel processors, perhaps similar to small perceptrons, may be realized in the brain as heavily interconnected cell assemblies. We suggest that one or several of the many small parallel machines would be directed by the spotlight to analyse a portion of the image – or its feature maps – which the spotlight illuminates. Interestingly,

how to route information in a parallel computer is emerging as the main theoretical and technological problem of the new computer architectures presently under construction. It is intriguing to speculate that the attentional spotlight may play a key role in solving exactly the same problem of how to route information in the brain.

Whether these ideas make any biological sense and what the biophysical basis of routing could be, are open questions that await new data. It is fascinating nonetheless that computational considerations and psychophysical and neuroanatomical data on attention now illuminate each other in new and intriguing ways.

Selected references

- 1 Hillis, D. (1981) *Artificial Intelligence Lab. Memo*, No. 646, MIT, Cambridge, MA
- 2 Treisman, A. (1983) in *Physical and biological processing of images*, (Braddick, O. J. and Sleigh, A. C., eds), Springer, Berlin
- 3 Julesz, B. (1984) *Trends NeuroSci.* 7, 41–48
- 4 Van Essen, D. C. and Maunsell, J. (1983) *Trends NeuroSci.* 6, 370–375
- 5 Wurtz, R. H., Goldberg, M. E. and Robinson, D. L. (1980) *Progress in Psychobiology and Physiological Psychology* 9, 43–83
- 6 Crick, F. (1984) *Proc. Natl Acad. Sci. USA*, 81, 4586–4590
- 7 Posner, M. I. (1980) *Q. J. Exp. Psychol.* 32, 3–25
- 8 Posner, M. I., Cohen, Y. and Rafal, R. D. (1982) *Philos. Trans. R. Soc. Lon. Ser. B* 298, 187–198
- 9 Posner, M. I., Snyder, C. R. R. and Davidson, B. J. (1980) *J. Exp. Psychol.* 109, 160–174
- 10 Tsai, Y. (1983) *J. Exp. Psychol.* 9, 523–530
- 11 Treisman, A. (1982) *J. Exp. Psychol.* 8, 194–214
- 12 Treisman, A. and Gelade, G. (1980) *Cog. Psychol.* 12, 97–136
- 13 Garner, W. R. (1974) *The processing of information and structure*, Lawrence Erlbaum, Potomac, MD
- 14 Bergen, J. R. and Julesz, B. (1983) *Nature (London)* 303, 696–698
- 15 Sagi, D. and Julesz, B. (1984) *Perception*, A23
- 16 Barlow, H. B. (1981) *Proc. R. Soc. Lon. Ser. B* 212, 1–35
- 17 Barrow, H. G. and Tenenbaum, J. M. (1978) in *Computer Vision Systems* (Hanson, A. and Riseman, E., eds), Academic Press, New York
- 18 Marr, D. (1976) *Philos. Trans. R. Soc. Lon. Ser. B* 275, 483–524
- 19 Minsky, M. (1961) *Proc. IRE* 49, 8–30
- 20 Treisman, A. and Schmidt, H. (1982) *Cog. Psychol.* 14, 107–141
- 21 Zeki, S. M. (1978) *Nature (London)* 274, 423–428
- 22 Newsome, W. T. and Wurtz, R. H. (1981) *Soc. Neurosci. Abstr.* 7, 732
- 23 Koch, C. and Ullman, S. (1984) *Artificial Intelligence Lab. Memo*, No. 770, MIT, Cambridge, MA
- 24 Bushnell, C., Goldberg, M. E. and Robinson, D. L. (1981) *J. Neurophysiol.* 46, 755–772
- 25 Mountcastle, V. B., Andersen, R. A. and Motter B. C. (1981) *J. Neurosci.* 1, 1218–1235
- 26 Haenny, P. E., Maunsell, J. H. R. and Schiller, P. H. (1984) *Perception* 13, A12
- 27 Braitman, D. J. (1984) *Brain Res.* 307, 17–28
- 28 Minsky, M. and Papert, S. (1969) *Perceptrons*, MIT Press, Cambridge, MA
- 29 Poggio, T. (1983) *Physical and Biological processing of Images* (Braddick, O. J. and Sleigh, A. C., eds), pp. 128–153, Springer, Berlin
- 30 Ullman, S. (1984) *Cognition*, 18, 97–159

A. HURLBERT and T. POGGIO

Center for Biological Information Processing and Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139, USA.