Vision: **Modular analysis – or not?**
David Burr

It has commonly been assumed that the many separate areas of the visual system perform modular analyses, each restricted to a single attribute of the image. A recent paper advocates a radically different approach, where all areas in the hierarchy analyse all attributes of the image to extract perceptually relevant decisions.

Address: Isituto di Neurofisiologia, Via S. Zeno 51, Pisa and Department of Psychology, University of Florence, Italy.
E-mail: dave@in.pi.cnr.it

More than a half of our cerebral cortex is devoted to seeing. Visual analysis starts at the retina, proceeds through the thalamus to the striate visual cortex (V1) in the occipital lobe, and then on to other areas of extra-striate cortex. Over the past few years, anatomists have identified over 30 separate visual cortical areas [1], all richly interconnected within a complex hierarchical and parallel organisation. Obvious questions for vision research are "why do we have so many areas" and "what do each of these areas do". The classic approach [2,3] has been to assume that V1 separates information about different attributes of the image — position, colour, motion, depth and so on — and dispatches it to various areas of extra-striate cortex for specialised modular analysis. According to this view, the results of the independent analyses are combined at a later stage (see Figure 1). This idea draws heavily on David Marr's [4] 'principle of modular design', and has received wide support from a range of different disciplines, including physiology, psychology, neurology and, more recently, imaging studies.

In a recent paper, however, Peter Lennie [5] challenges this traditional view, and offers a thought-provoking alternative. Lennie suggests that the various different attributes — or dimensions — of an image are not parcelled out to separate areas, but that their analysis remains intimately coupled at all stages of analysis. Each area recovers perceptually relevant information to produce a perceptual decision about the image (not just about a single attribute). The next area in the hierarchy receives this decision as its input, but not the evidence on which it was based (which is effectively discarded after use). Furthermore, 'consciousness' has direct access to the perceptual decisions made at all levels, bypassing the hierarchy. This model, together with the traditional approach, is illustrated in Figure 1.

Lennie is quick to note that "the argument cannot yet be developed rigorously"; but over the 47 pages of his paper

[5] he has made a commendable effort to do so, discussing in detail the potential advantages and disadvantages, and scouring the literature for relevant evidence. Starting with electrophysiological single-cell recordings from the various areas (probably the most solid data available), he points out that these show no clear specialisation in their response properties — with the notable exception of area V5/MT (see below) — nor indeed is there good evidence for differences in attribute specialisation within individual areas. So, if the separate areas are not there to analyse different visual attributes, what might they be doing?

Consider first the primary visual cortex, V1, the largest visual area. Given that V1 covers 20% of the visual cortex and comprises 40% of all visual cortical cells, it does seem unlikely that this investment of machinery simply multiplexes the thalamic input out to other areas. Lennie also

Figure 1
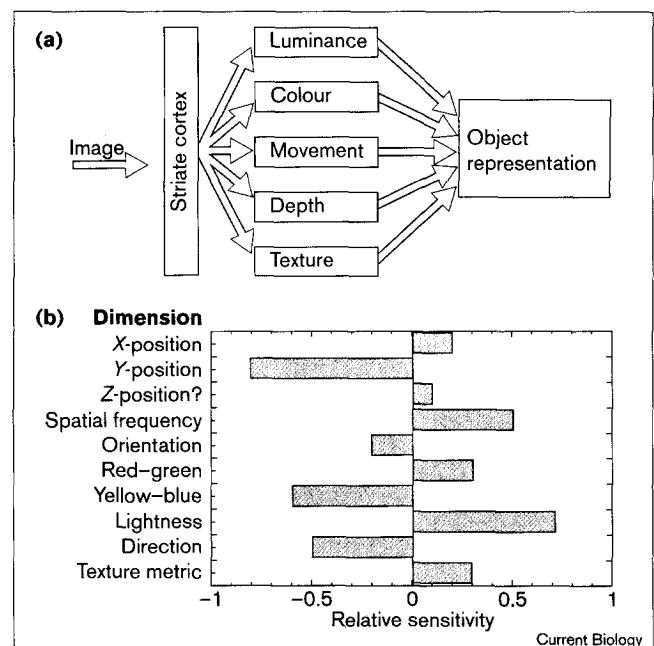
Two views of information transmission through the visual cortex. (a) Modular analysis [2,3]: the striate cortex, V1, separates information about the different dimensions of an image and dispatches it to different specialized extrastriate areas, or modules, that perform a parallel analysis of each dimension. At a later stage, the results of the analyses are combined. (b) Lennie's approach [5]: V1 preserves, in a closely-coupled relationship, all the information about different dimensions of the image. Each neuron is selective to a small region of a multi-dimensional space, so its properties can be represented by a vector in that space (shown diagrammatically here for a single cell).

questions the popular idea that V1 forms an economical representation of the image, given that it contains 200 times as many cells as its thalamic input. Pointing out that a major task for vision is to classify objects quickly and reliably, he develops an interesting argument that V1 encodes visual information in a sparse, distributed representation that greatly facilitates the classification task. This representation is achieved by employing a large number of neurons at each spatial position, each simultaneously selective over many visual dimensions, thereby occupying a small region in the multidimensional space (see Figure 1). As well as facilitating classification (for reasons that are well explained, but too complex to summarise here), this highly redundant coding strategy turns out to have a range of other advantages, including robustness and minimisation of energy consumption.

What else might be achieved by the various visual areas? One task for V1 could be contrast discrimination, as its neurons have a much larger working range than those of the higher areas. Other tasks might include texture discrimination (by complex cells), and contrast normalisation. V2 might be involved with binocular fusion, and element grouping over relatively short ranges. Grouping, or second-stage integration, is continued by V4, which also detects higher-order contrasts (including, but not limited to, colour contrast) and analyses complex symmetries that might help to segregate figure from background. V3 is somewhat enigmatic, and must remain without a clear occupation for the moment. The only area for which Lennie [5] concurs with the popular dogma is V5 (MT), which is unquestionably dedicated to motion analysis [6] (as is its adjacent area V5a or MST). Even here, however, Lennie does not delegate all motion analysis to V5, but only that related to the large-field motion, or optic flow, that is produced by self-motion. The local motion of individual moving objects, such as a red Ferrari, remains very much glued to the other dimensions — form, size, colour and so on — in all visual areas.

I find myself in sympathy with this view, having long been perplexed by the idea of modular design, given that the analysis of one attribute often depends so heavily on the analysis of others. One clear example is spatio-temporal interpolation, well illustrated by observing a person walk behind a slatted fence. Although the person is visible only through the gaps in the fence, our visual system interpolates the image concealed by the slats, so we see a complete figure. This seemingly magical phenomenon lends itself to a very simple explanation if we assume the existence of neurons with spatio-temporally tuned (motion-selective) receptive fields that serve not only to calculate velocity, but also to provide information about spatial structure [7]. Such neurons will automatically convert the temporal information given by the arrival times of the different features at the gaps to spatial information, from

which form is readily extracted. If, on the other hand, velocity and form are analysed in completely different areas, the task would be far more daunting, requiring specific and biologically implausible hardware.

In a recent issue of *Current Biology*, Snowden [8] has provided another example of an interaction between two distinct attributes, motion and spatial position. After a period of adaptation to unidirectional motion, stationary scenes appear to move in the reverse direction; but Snowden has observed that the adaptation to motion also produces a change in apparent spatial position, difficult to understand if the two attributes are coded separately in distinct areas. Many other examples of this type of interaction abound, such as the high sensitivity with which we can detect the motion of patterns defined only by their colour (see [9]). Perhaps the most famous example of perceptual interdependence is the colour-contingent orientation after-effect of McCollough [10], where adaptation to red vertical and green horizontal stripes causes subsequently viewed white stripes to appear green if vertical and red if horizontal. This is certainly consistent with Lennie's sparse distributed coding model, where each neuron encodes both orientation and colour, and indeed may reflect a dynamic adaptation that maximises efficiency, along the lines suggested by Barlow [11].

An interesting aspect of Lennie's [5] account is that so much of the visual cortex is devoted to seemingly 'simple' tasks, such as contrast encoding, perceptual grouping and surface extraction, leaving little hardware for the 'difficult' perceptual tasks, such as object recognition. But here our intuitions may betray us. Surface extraction is by no means a trivial feat, given the inherent ambiguities in the retinal input, and object recognition itself may not be that difficult, particularly if the visual information has been preprocessed in such a way as to facilitate rapid classification and the requisite perceptual decision-making. Indeed, computer vision programs have proven more successful with complex recognition tasks, such as handwriting, than they have with surface extraction and perceptual grouping, where they are easily fooled.

Lennie's [5] ideas will not be popular with everybody. He himself notes that the best evidence for modular visual analysis comes from specific perceptual impairments — such as loss of colour vision, or achromatopsia — caused by focal cortical lesions. Lennie promotes alternative interpretations for these curious neurological conditions, invoking damage to higher-level functions such as 'naming' of the colours, rather than colour perception per se; but I suspect that few neuropsychologists will be convinced. It is nevertheless remarkable that these neurological conditions are so rare, while television sets, unquestionably modular in design, readily demonstrate post-traumatic achromatopsia. It is also relevant that cortically colour-blind patients can

use chromatic information in many ways, such as to see the motion of purely chromatic stimuli [9], showing that not all of their central chromatic pathways are inoperative.

Other support for modular design comes from imaging studies, those colourful papers that now dominate the scientific weeklies with ever-new evidence of morphological specialisation for various tasks and attributes. Lennie [5] has chosen to assign little weight to these new studies, and perhaps this caution is well justified. The results of any research often depend strongly on the theoretical motivation for the study, which heavily conditions the experimental design (such as the choice of appropriate control conditions for image subtraction). It is possibly that, if the experiments were guided by different theoretical constructs, a different picture would emerge. Indeed, the most recent functional magnetic resonance imaging studies [12] have seriously questioned the involvement of V4 in colour perception, proposing in its stead a completely new area, V8. But is colour perception the function of V8, or might this apparent specialisation be an epiphenomenon, resulting from other factors, such as the high foveal representation of this area [12]?

Only time will tell whether the ideas presented in Lennie's [5] thesis prove to be correct, or even useful. But I am certain that most readers will find this paper scholarly, refreshingly novel and intellectually challenging, well worth a few hours reading and contemplation.

References

1. Fellerman DJ, Van Essen DC: **Distributed hierarchical processing in the primate cerebral cortex.** *Cerebral Cortex* 1991, 1:1-47.
2. Zeki S: *A Vision of the Brain.* Oxford: Oxford University Press; 1993.
3. Livingstone MS Hubel DH: **Segregation of form, color, movement and depth: anatomy, physiology and perception.** *Science* 1988, 240:740-749.
4. Marr D: *Vision.* San Francisco: WH Freeman; 1982.
5. Lennie P: **Single units and visual cortical organization.** *Perception* 1998, 27:889-935.
6. Shadlen MN, Britten KH, Newsome WH, Movshon JA: **A computational analysis of the relationship between neuronal and behavioral responses to visual motion** *J Neurosci* 1996, 16:1486-1510.
7. Burr DC, Ross J: **Visual processing of motion.** *Trends Neurosci* 1986, 9:304-306.
8. Snowden R: **Shifts in perceived position following adaptation to visual motion.** *Curr Biol* 1998 8:1343-1345.
9. Cavanagh P, Hénaff M-A, Michel F, Landis T, Troscianko T, Intriligator J: **Complete sparing of high-contrast color input to motion perception in cortical color blindness.** *Nat Neurosci* 1998, 1:242-247.
10. McCollough C: **Color adaptation of edge detectors in the human vision system.** *Science* 1965, 149:1115-1116.
11. Barlow HB: **A theory about the functional role and synaptic mechanisms of visual after-effects.** In *Vision: Coding and Efficiency.* Edited by Blakemore C. Cambridge: Cambridge University Press; 1990:363-375.
12. Hadjikhani N, Liu AK, Dale AM, Cavanagh P, Tootell RBH: **Retinotopy and colour sensitivity in human visual cortical area V8.** *Nat Neurosci* 1998, 1:235-241.