

---

# 11 Evolution of Visual Processing

G. Adrian Horridge

---

## Introduction

---

The distribution of eyes, their anatomy and the visual behaviour that they make possible, are well documented in comprehensive texts, of which several are referenced with an asterisk (\*) at the end of this article. Here I will endeavour to draw attention to a few general principles that apply to the simpler examples of natural visual processing so far as we know it in invertebrates, and to our primitive efforts to copy low-level natural vision into artificial systems. These principles apply to a very large number of scattered examples of eye types and to a wide variety of visual behaviour, but most of the data actually refer to insect vision, which is reasonably well known.

### The Anthropocentric View of Vision

The first hurdle to be overcome, and the most difficult, is the acceptance of the view of primitive vision as necessarily resembling human vision. Humans create a marvelous visual world which arrives in consciousness already endowed with meaning. We see trees, dogs and faces already given names and nouns by deep unconscious interaction with language and other memories. We overcome these preconceptions with the aid of comparative anatomy and comparative visual behaviour, which immediately reveals the limitation of the view that (say) insect vision is like all VISION which is like human VISION. I prefer to call the visual performance of animals like insects *semivision*, because we have no evidence that they have our kind of visual world, although they obviously see very well.

### No Comprehensive Theory

The evolution of eyes and visual processing encompasses so many topics that it is not easy to see the significant aspects, especially because numerous disciplines are involved. Our relatively isolated example of technology which mimics natural vision, the camera, gives a false impression that 'vision' sees 'pictures', but the camera

only *transmits* the picture. In Canberra in the 1970s we elaborated a general theory to explain how the dimensions of eyes and the angles between visual axes are related to the limits of lens resolution, as fixed by the diffraction of light, and the noise limits arising from the random arrival of photons. The theory referred to the sensitivity which depends on receptor size and lens diameter and to the spatial resolution which also depends on receptor size and spacing. Basically, it was a sampling theory (Kirschfeld, 1976; Snyder, 1975, 1979). However, the advent of the new architectures of computers with parallel distributed processing brings another framework, which can now bind together a new synthesis to span the gap downstream between the array of photoreceptors, the functions of single neurones and the coordinated processing of visual behaviour. Indeed, for almost a century there have been discussions about the emergent properties of nervous systems; in the case of the visual system we can now see clearly that emergent properties reveal themselves because in cooperative parallel processing the neurones are individually ambiguous and inadequate exactly because they have evolved to operate in groups. However, I anticipate my conclusion and must return to the evolution of vision.

## Importance of Broad Absorption Spectra

### Pigment Spots and Visual Pigments

The commonest visual pigments (called rhodopsins) are combinations of a protein (often called an *opsin*) with a carotenoid, which is lipophilic and has a broad absorption spectrum. Carotenoid molecules have a long hydrocarbon chain of alternating double and single carbon-carbon bonds which resonate in many ways and cause a broad absorption spectrum. The position of the peak, that is, the colour of the visual pigment, depends on the constitution of the protein, so that pigments can readily evolve to suit the visual world and the visual tasks in that world. Pigments are known with a peak in the range from 350 nm in the ultraviolet to about 620 nm in the mid-red. Most of

the broad-spectral sensitivity curves have a shape that agrees with the Dartnall (1953) nomogram. Visual pigments apparently evolved independently many times from related molecules already in the cell. Eyes with 3 photoreceptor colour types are common in several groups of animals; eyes with a single pigment are also widespread, and a few are known with 2 (cockroach), 4 (butterflies and some birds) or 5 (dragonflies).

The carotenoid part of the molecule is held by the cell membrane and the absorption of a photon opens channels in the membrane that cause a polarization or depolarization of the cell via an amplifying cascade. In many photoreceptors the single photon captures cause individual miniature potentials (called 'bumps'), best known in arthropods. An important consequence of this mechanism is that the magnitude of the response is independent of the wavelength of the photon. Photoreceptors therefore act as photon counters irrespective of the colour of the captured photon, and the discrimination of colour must depend on an accurate measurement of the ratios of photon numbers captured by photoreceptors with different spectral peaks. The spectral sensitivity of a photoreceptor is its field of view in the colour dimension: these fields are broad and they overlap a great deal where there are several spectral types. Colour discrimination is excellent because it depends on overlapping spectral sensitivity of different receptors which are excited in different ratios. This principle is important and we will return to colour discrimination as a model for processing in other dimensions.

## The Laws of Optics Govern Eye Evolution

### Absorbing Light Guides and Fused Rhabdomeres

*The rhabdomere* is the organelle composed of tightly packed microvilli in the photoreceptor cell. The microvilli contain the visual pigment molecules and, being composed of a high proportion of lipid membrane, they have a relatively high refractive index (of about 1.39 compared with 1.34 in the adjacent cytoplasm). The visual pigment molecules in a layer 1  $\mu\text{m}$  thick absorb only about 1% of the light, and therefore there has been strong evolutionary pressure to increase the depth of the absorbing layer. At the same time the cross-sectional area of the rhabdomere must remain small in order to optimize the resolution and match the cross-section of the absorbing organelle to the resolution of the lens (see Fig. 11.3). As a result, we find that many excellent eyes have the rhabdomere in the form of a long, thin rod which points towards the nodal point of the optical system. Examples occur in coelenterates, annelids, arthropods and molluscs and in the rods and cones of vertebrate eyes. The rod acts as a light guide which conducts the light falling on its end, absorbing the light as it

passes along the rod (Figs 11.1 and 11.2). Where these receptor rods occur we can infer that resolution and sensitivity are being optimized so that they do not act against each other, as they would if the rhabdomeres were merely increased in size, like the silver grains in a sensitive film.

Increasing sensitivity by increase in the length of the rod soon leads to the self-absorption effect. Absorption in the rod is about 1% per  $\mu\text{m}$  at the absorption peak, so that most of this peak light has been absorbed at 100 to 200  $\mu\text{m}$  along the rod. The off-peak photons are absorbed also, although less effectively, with the nett effect that the longer the receptor rod the broader the absorption curve. The ultimate result is that when all the light is absorbed the receptor is *black*. This achieves a maximum sensitivity but rules out the possibility of colour vision by receptors of different colours. This is presumably why the cones of vertebrates are shorter than the rods. The above considerations mean that the absorption must be below about 50% and the receptor rod must not exceed 30 to 50  $\mu\text{m}$  in length, depending on the exact numerical values of the coefficients (Snyder, 1975, p 197).

*The fused rhabdom* is an anatomical solution to the problem of increasing sensitivity and retaining colour vision. The rhabdomere is the contribution of a receptor cell to the whole rhabdom, which is now the light guide. The light is therefore absorbed by each rhabdomere as it is carried down the rhabdom as a unit (Fig. 11.2). If the rhabdomeres differ in their absorption peaks, having for example peaks in the ultraviolet, blue and green, each absorbs a different fraction of the total light and it is now possible to absorb all of the light and retain colour vision. Most of the larger insects have 2, 3 or 4 spectral types of rhabdomeres fused into rhabdoms more than 100  $\mu\text{m}$  long. Elongated *fused* rhabdoms are therefore a sign of colour vision.

The same principle, combining 100% absorption with complete discrimination, applies to the plane of polarization of polarized light. Crustacea commonly have interleaved layers of microvilli arranged so that microvilli of one set of receptor cells are at right angles to those of the other cells. Insects, spiders, molluscs and others commonly have rhabdomeres of adjacent cells pressed together side by side with microvilli of different orientations. This arrangement is not necessarily proof that discrimination of the polarization plane is important for animals that have it, because absorbing light in all polarization planes is also a way of increasing sensitivity in general.

### Matching Receptor Resolution and Lens Resolution

Following a seminal paper by Kirschfield (1976), and inspired by Allan Snyder, a long series of papers have

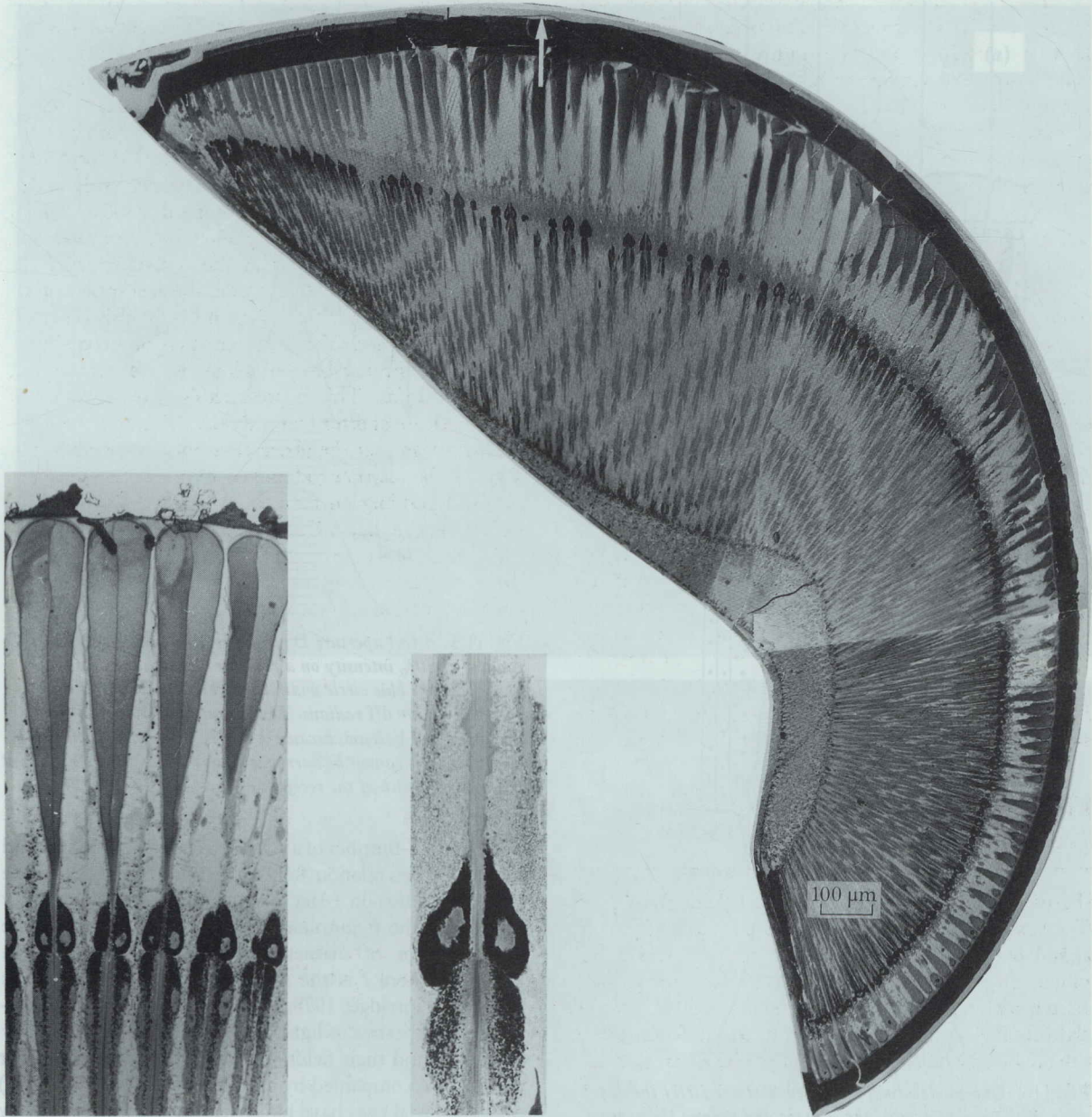


Fig. 11.1 *The anatomical basis of the regional differences in spatial resolution in the eye of the praying mantis. This is a horizontal section through the compound eye of Tenodera. The forward-looking part of the eye has smaller angles between visual axes (smaller  $\Delta\phi$ ), longer cones and therefore longer focal length (smaller  $\Delta\rho$ ) and larger facets than the side of the eye. The insets show the critical region where the cone meets the rhabdom, where the ray optics is matched to the light-guide optics.*

appeared from Canberra on the geometry of receptors, the dimensions of eyes and the diffraction limitation of light. Essentially, the optics and receptors must be matched to make the best use of the lens resolution. All eyes are sampling devices in angular coordinates. To resolve contrast with least blurring, the focus must be perfect, the optics of

the lens must be as good as diffraction allows, and then to detect as small a contrast as possible, the size of the receptors must match the optical resolution of the finest detail by the lens. These conditions of matching spatial resolution, least-motion resolution and lens resolution, are met by working at the limit set by the wavelength of light, and

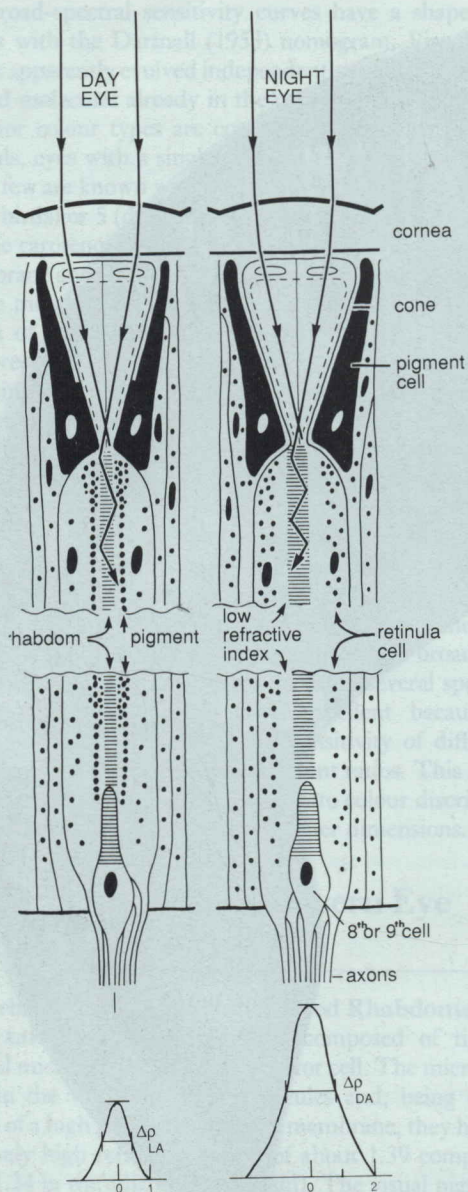


Fig. 11.2 One ommatidium in vertical section: (left) the day state, with pigment around the cone tip and around the narrow rhabdom; (right) the night state, with a clear space around the track of the light and a broader rhabdom. As shown by the angular sensitivity fields plotted at the foot of each figure, the day eye is less sensitive and has a narrower field at the 50% sensitivity contour. This diagram includes features of several of the groups of large diurnal insects.

the best compromise is found when the photoreceptors are only  $1$  to  $2 \times 10^{-6}$   $\mu\text{m}$  in diameter, which is 2 to 4 times the wavelength of green light (see below). A similar grain size is found in photographic film for similar reasons,

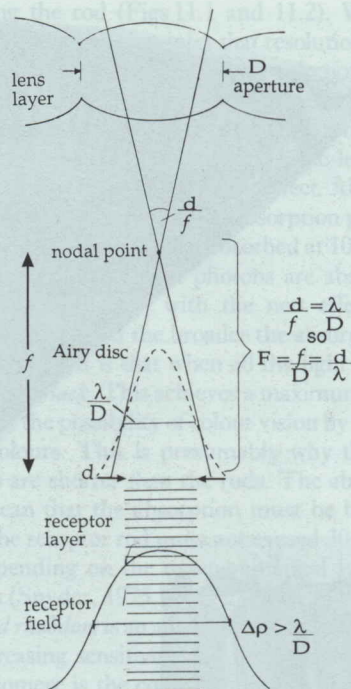


Fig. 11.3 A lens aperture  $D$  generates a blur circle of width  $\lambda/D$  radians at 50% intensity on a receptor of width  $d$  in the focal plane. For the blur circle width to match the subtense of the receptor,  $\lambda/D = d/f$  radians. The  $F$ -number of the lens is  $f/D$ . Therefore  $d = F/2 \mu\text{m}$ , because  $\lambda = 0.5 \mu\text{m}$ . The field width  $\Delta\rho$  of the receptor cannot be narrower than the blur circle on account of the finite width of the receptor.

because an  $F$ -number of about 2 to 4 is similar for eyes and cameras. This relation follows because the width of the circle of confusion (Airy disc) is  $\lambda/D$  at 50% intensity (Fig. 11.3), the  $F$ -number is  $f/D$  and the subtense of the photoreceptor of diameter  $d$  is  $d/f$  (all measured in radians), where  $f$  is the focal length and  $D$  is the lens aperture (Horridge, 1978a). Therefore, to match the diffraction properties of light, receptors must be only a few  $\mu\text{m}$  wide, and their fields must touch or overlap. Larger facets are accompanied by narrower rhabdoms (Fig. 11.4) and nocturnal eyes have huge rhabdoms that throw away resolution in exchange for sensitivity (Fig. 11.5).

#### Related Light-Guide Dimensions

From the relation  $d/f = \lambda/D$  (above) we can write  $\lambda/d = D/f = 1/F$  so that for  $\lambda = 0.5 \mu\text{m}$  and  $F = 2$  it follows that the receptor has a diameter of  $1 \mu\text{m}$ . Very conveniently, given the refractive index of the rhabdom material, raised above that of the surrounding cytoplasm by a high percentage of lipid, it can be shown that  $1 \mu\text{m}$  is also approximately the minimum diameter for a rhabdom to function as a light guide. Therefore we could say that the absorbing

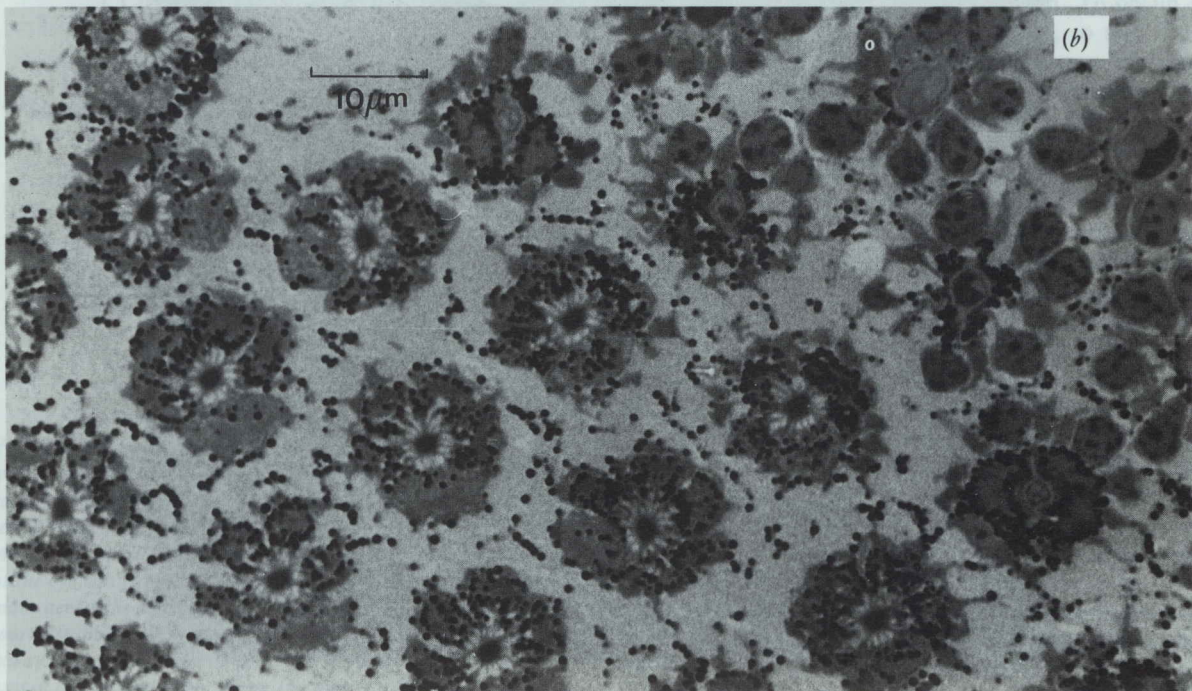
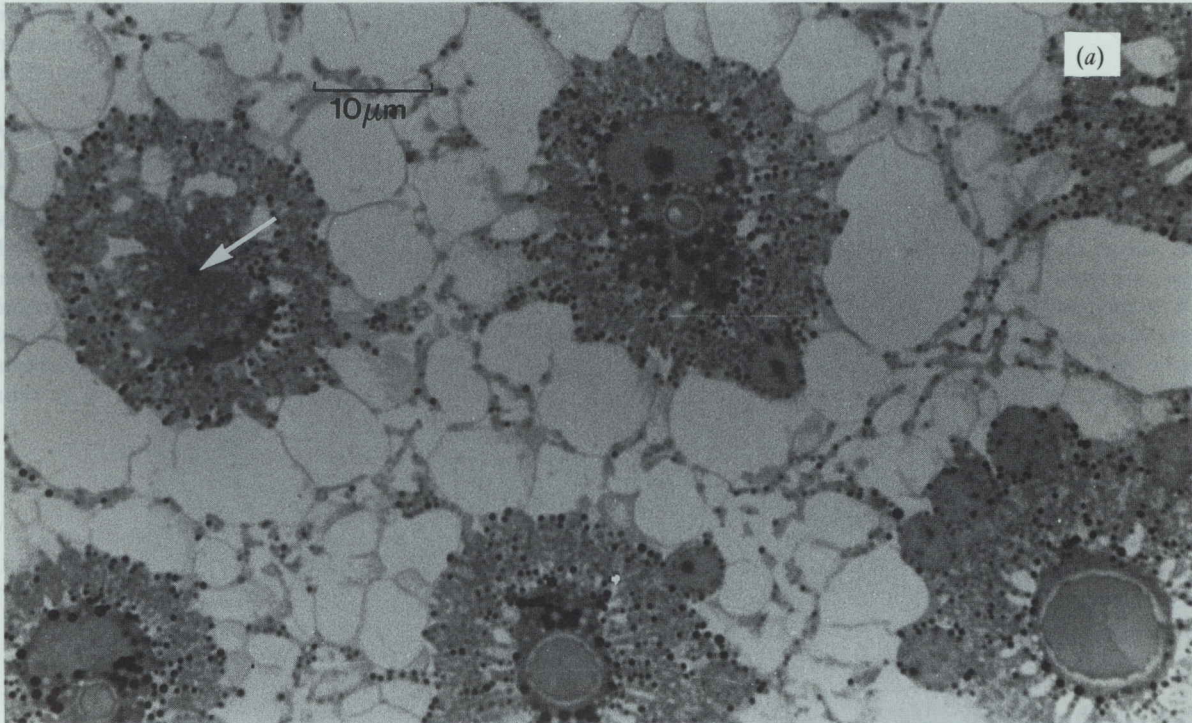


Fig. 11.4 Transverse sections through the region of the cone tip (see Fig. 11.1) of a dragonfly eye (*Austrogomphus*). (a) Through the foveal region looking forwards and upwards. The facets are large (from centre to centre) but the rhabdom (arrow) is only about  $1\ \mu\text{m}$  in diameter. (b) In the ventral part of the eye the facets have an area a quarter of that in the fovea and the rhabdom area is correspondingly four times as large as in (a), giving a nett uniform sensitivity to a diffuse source.

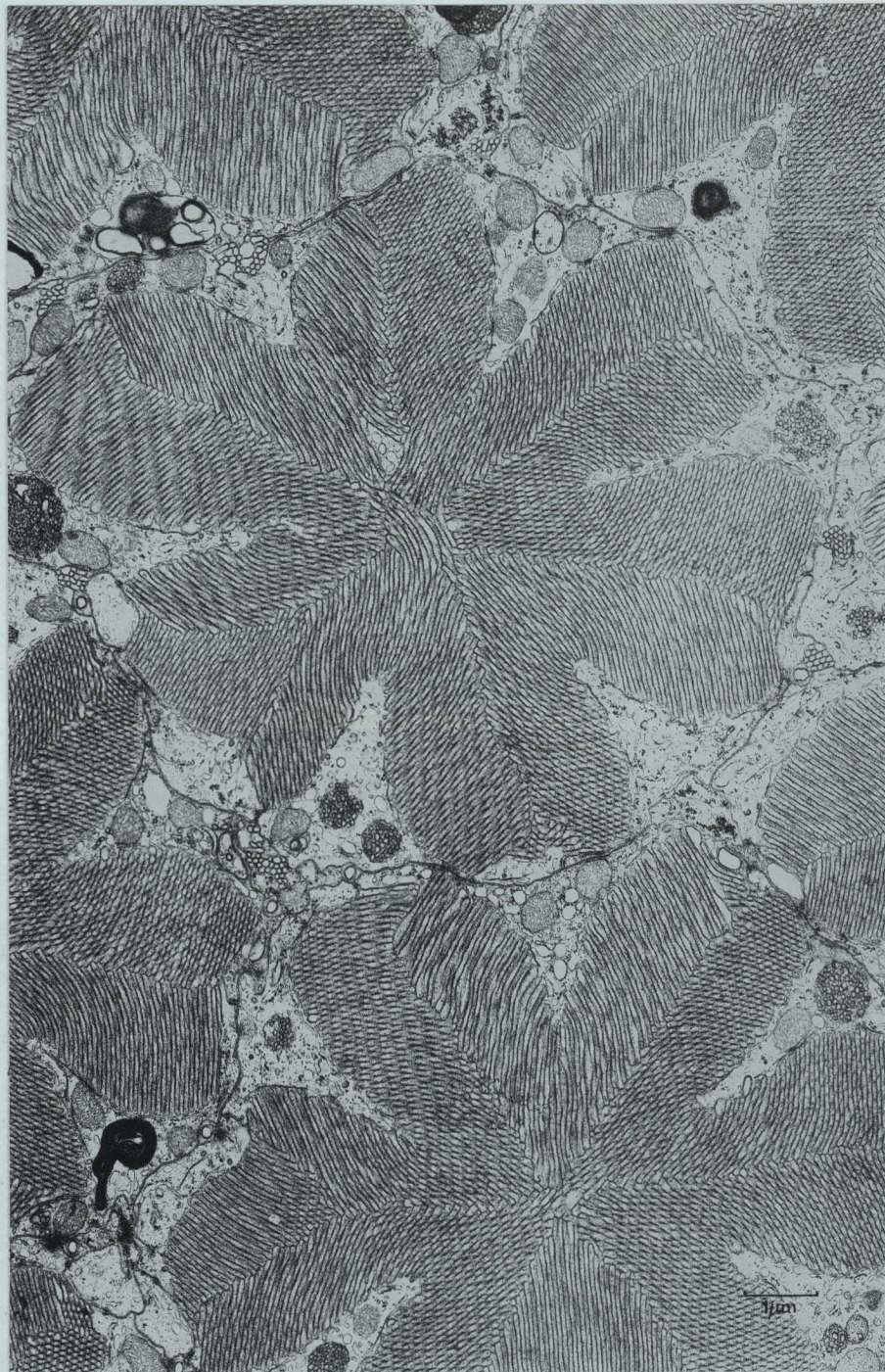


Fig. 11.5 *The rhabdoms of a nocturnal eye fill the cross-section within the eye to maximize the sensitivity to a diffuse source. This is a horizontal section through the eye of Chrysopa (Neuroptera). There are six large rhabdomeres to each rhabdom, but two other rhabdomeres at another level. The visual pigment molecules lie within the rhabdomere microvilli.*

light guides in eyes are admirably adapted to the  $F$ -number of the lenses that commonly occur. The geometry of the eye is governed by the laws of physics.

### Modulation

The main feature of a photoreceptor that governs the ability of the eye to see is the *field* of each individual receptor, which looks through this window at the outside world. The actual field is measured experimentally either by a moving, flashing point source (Fig. 11.6) or from the modulation of the potential caused by a range of regular striped patterns of different periods (Fig. 11.7). The stimulus to the receptor is the modulation of the light falling upon it as integrated in its field: the response is the

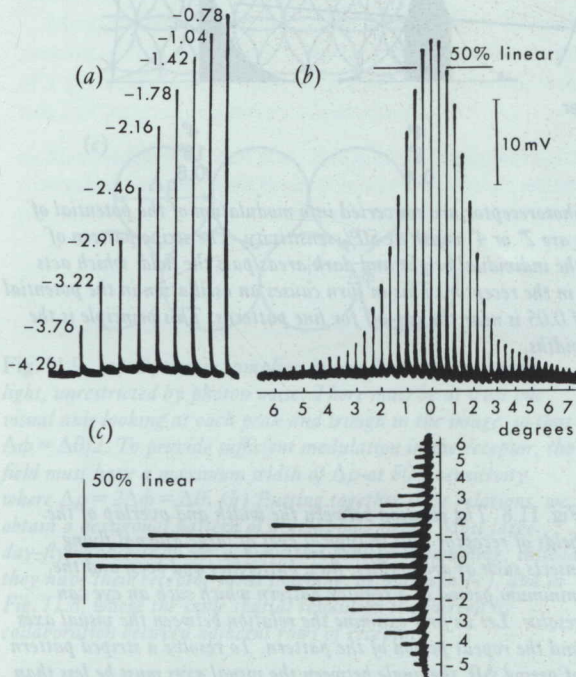


Fig. 11.6 The way that the visual field of a receptor is measured with a microelectrode in the receptor as a graded potential at each flash of light. (a) Calibration by means of a neutral density series. A light was flashed on axis at the log  $I$  values shown. (b) A constant small point source subtending less than half a degree at the eye was flashed at each half-degree step in the horizontal plane across the axis of the receptor. The angle subtended by this source at the eye is important. The brightness was selected to be always below saturation on axis. (c) A similar run through the axis in the vertical plane. From (b) and (c), angular sensitivity curves can be calculated from the calibrations in (a). From the angular sensitivities one measures  $\Delta\rho$  at the 50% sensitivity level for the light adaptation employed (Fig. 11.2), but note how close to the peak of the responses the 50% linear reduction in effective intensity is found.

modulated receptor potential, which is propagated to the second-order neurones (Fig. 11.10).

### Matching the Spacing of Visual Axes to the Lens Resolution

An array of receptors can resolve *in space* a spatial frequency up to a limit that is determined by the angle between the visual axes. Regular stripes, of period  $\Delta\theta$  subtended at the eye, can be distinguished when the angle between receptor axes  $\Delta\phi$  is smaller than the limit set by  $\Delta\theta = 2\Delta\phi$ , as illustrated in Fig. 11.8. In the same eye the stripes of period  $\Delta\theta$  can be resolved by the lens when  $\Delta\theta = \lambda/D$  from diffraction theory. Therefore  $\lambda/D = 2\Delta\phi$  when  $\lambda = 0.5 \mu\text{m}$  (Fig. 11.9).

We now have a relation between the spatial sampling resolution and the lens resolution, which in turn is linked to the receptor diameter, as applied to eyes with reasonable  $F$ -number and which can afford to operate near the diffraction limit because there is plenty of light. Of course, the processing mechanism behind the retina must be able to deal with this spatial resolution, which is the reason for postulating (later in this review) motion-detection templates *at this resolution*.

The above relation can be used to calculate the diameter of a compound eye for a given resolution. To obtain a spatial resolution of  $1^\circ$  we have  $\lambda/D = 1^\circ$ , so that  $D$  is approximately  $30 \mu\text{m}$  minimum (as in eyes of large dragonflies). The eye radius ( $R$ ) is therefore given by  $D/R = 1^\circ$  if the facets are adjacent, so that  $R$  is approximately 1.8 mm.

### Sacrifice of Resolution for Sensitivity

Many animals live in situations, or emerge at times of day, when vision is limited by lack of light, not resolution. Photons arrive at random, and so cause shot noise in the receptors at low light levels. Sunlight gives about  $10^{14}$  useful photons  $\text{cm}^{-2} \text{s}^{-1}$ , moonlight about  $10^7$  of them. We can calculate that if 1% of the photons are reflected from objects in full moonlight, a facet  $30 \mu\text{m}$  in diameter on an insect eye catches about  $1 \text{ photon s}^{-1}$ , which is hardly sufficient for vision even at 50% capture efficiency. Under the same illumination a camera-type eye with a lens 3 mm in diameter and low  $F$ -number has sufficient light in moonlight, even with receptors  $1 \mu\text{m}$  in diameter, e.g. rods in human eye.

The sensitivity of a receptor behind a lens is proportional to the area of the lens and the solid angle of the receptor area as subtended at the nodal point of the lens:

$$\text{sensitivity} = \frac{kd^2D^2}{f^2} = \frac{kd^2}{F^2}$$

where  $k$  depends on the efficiency of photon capture in the rhabdomere, and is approximately 1% per  $\mu\text{m}$  of depth.

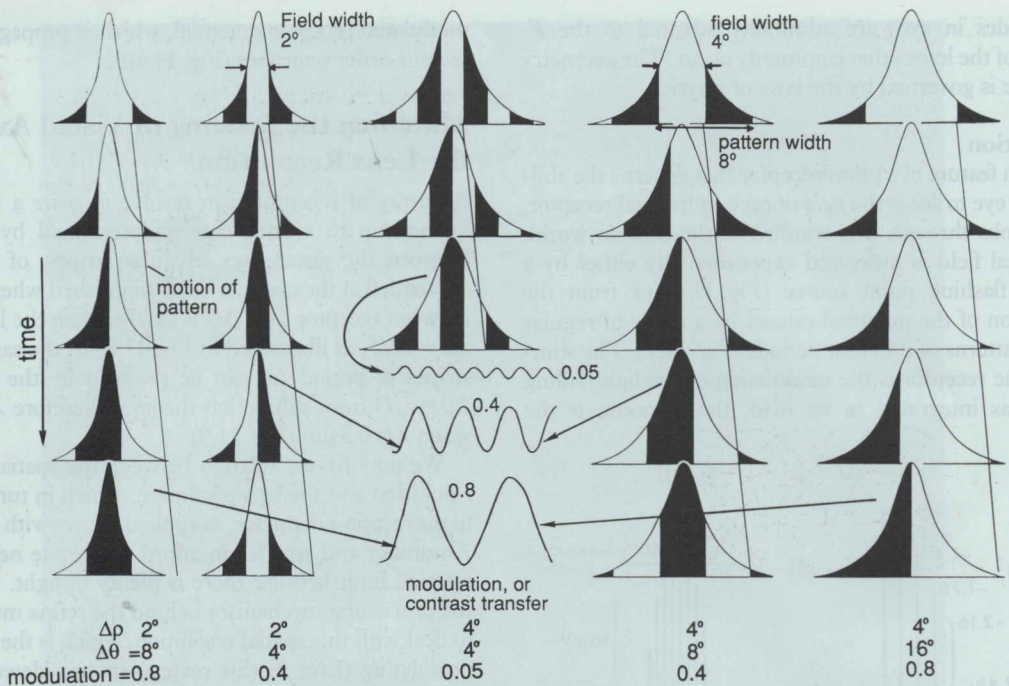


Fig. 11.7 The way that changes in intensity of the light falling on the photoreceptor are converted into modulation of the potential of the receptor, in relation to the field width. In this diagram the fields  $\Delta\rho$  are 2° or 4° width at 50% sensitivity. The stripe pattern of period  $\Delta\theta$  subtends 4°, 8°, or 16° at the eye and moves to the right. As the individual bright and dark areas pass the field, which acts like a shaped window, they cause modulation of 0.05 (5%), 0.4 or 0.8 in the receptor. This in turn causes an oscillation in the potential of the cell at the frequency at which the stripes go past. A modulation of 0.05 is near the cut-off for fine patterns. This principle is the basis of the loss of resolution of fine detail as a result of increased field widths.

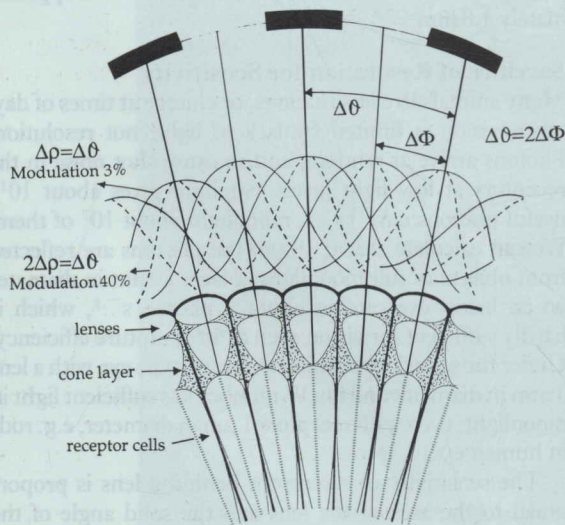


Fig. 11.8 The relation between the width and overlap of the fields of receptor cells in typical eyes of large diurnal flying insects such as dragonflies, flies, butterflies and bees, and the minimum period of a regular pattern which such an eye can resolve. Let us first examine the relation between the visual axes and the repeat period of the pattern. To resolve a striped pattern of period  $\Delta\theta$ , the angle between the visual axes must be less than  $\Delta\theta = 2\Delta\phi$  (at top right). To generate sufficient modulation for vision (3%) the field width of the photoreceptors must be less than  $\Delta\rho = \Delta\theta$ , as explained in Fig. 11.4. To generate 40% modulation, we have  $\Delta\rho = \Delta\theta/2 = \Delta\phi$ , which is a commonly occurring compromise that requires better lens resolution than that set by the diffraction limit. This principle is the basis of the compromise of spatial resolution as limited by the density of the visual axes and the overlap and width of the receptor fields.

Therefore the only way to increase sensitivity at constant  $F$ -number and capture efficiency is to increase receptor area (Fig. 11.5). This is why, in a camera for a given film type, more sensitive film has larger grains and therefore poorer spatial resolution. For the same reason, the evolu-

tion of eye performance is almost entirely related to the number of receptors of a given size that can be packed in.

The lower the light level is, the poorer is the spatial resolution that can be achieved, and the larger the lens aperture needed to catch the incident light. Therefore the



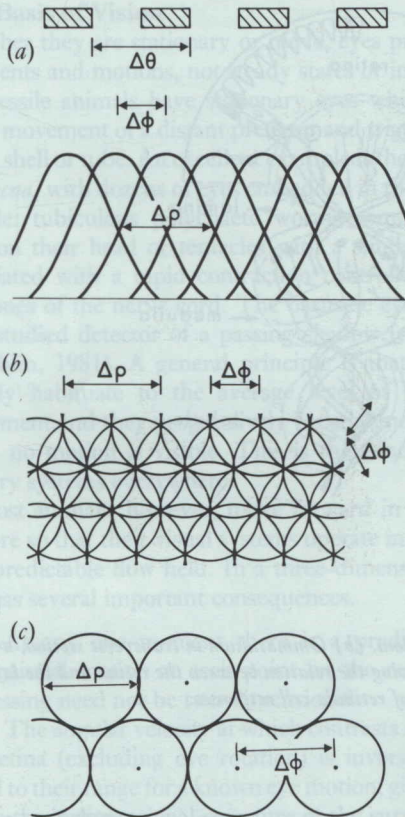


Fig. 11.9 (a) Optimum sampling in one dimension in bright light, unrestricted by photon noise. There must be at least one visual axis looking at each peak and trough in the image, so that  $\Delta\phi = \Delta\theta/2$ . To provide sufficient modulation in the receptor, the field must have a maximum width of  $\Delta\rho$  at 50% sensitivity, where  $\Delta\rho = 2\Delta\phi = \Delta\theta$ . (b) Putting together these relations, we obtain a hexagonal pattern of overlapping fields. Most large day-flying insects do not achieve this density of sampling: instead they have their receptor fields touching, as shown in (c), and in Fig. 11.8, where the same spatial resolution is achieved by collaboration between adjacent rows of receptors.

ratio of lens resolution to spatial resolution must increase as the eye evolves for vision in dim light. At one extreme, matching the diffraction limit to the spatial resolution in bright light (Fig. 11.9(b)), we had  $1/D = 2\Delta\phi$  so that  $D\Delta\phi = 0.25 \mu\text{m}$ .  $D\Delta\phi$  is called the *eye parameter* and it is half of the ratio of the spatial resolution to the lens resolution. A larger eye parameter, sacrificing some resolution for sensitivity, is shown in Fig. 11.9(c), where  $D\Delta\phi = 0.5 \mu\text{m}$ . To maintain the optimum resolution and sensitivity, the eye parameter must increase for low light levels to values between 3 and  $5 \mu\text{m}$  (Snyder, 1979, p 249). We see that much of the geometry of the eye is governed by the laws of physics.

## The Evolution of Vision

### The Old Problem of Eyes and Brains

For 130 years there has been discussion as to how the eye (of man) could have evolved together with the visual centres of the brain when every small improvement in one would be useless without an equivalent improvement in the other. The problem was well known to Darwin and the later proponents of evolution, and also to those who attacked the theory of evolution using arguments of this kind. A related problem, how the eye of a squid or octopus comes to resemble that of a man or elephant, was regarded as evidence that evolution limited to small steps selected from random mutations could never fully account for the facts. It was also clear that the theory of selection could *not at that time* be shown to be the only mechanism of evolution, even though the best scientists of Europe were convinced, and therefore the difficult example of the evolution of eye and brain was repeatedly quoted as support for alternative explanations of animal structure.

In hindsight we see that the problem was caused partially by the assumption that an eye with the structure of a camera must produce a two-dimensional picture, and because it was not realized that the evolution of an excellent eye is governed by the laws of optics, while the progressive evolution of the visual centres can follow along independently later, as the visual tasks become more complicated.

### Eyes Without Brains

Extensive texts describe the structure and histology of numerous eyes in primitive animals. Descriptions up to 1963 will be found in Bullock and Horridge (1965); the best special textbook is still Plate (1924), while modern work is reviewed in Ali (1984) and by Meinertzhagen and by Osorio (this volume).

Comparative anatomy tells us that excellent eyes, with numerous small photoreceptors and an apparently good optical system, evolved independently many times in lower invertebrates such as medusae and worms (Fig. 11.11) in the absence of a nervous system which could achieve much visual processing. A great many of these eyes generate no complex visual behaviour. The eye was clearly an organ of spatial resolution long before the nervous processing structures evolved to make more use of its potential. Many of these primitive eyes detect motion while they are stationary, with no suggestion that a scanning eye like that of *Copilia* (Gregory, 1967) or the fly larva (Fraenkel and Gunn, 1940) preceded them in evolution. The selective advantage of the lens resolution and spatial sampling resolution of these lower-invertebrate eyes is that an approaching predator or the direction of a mate is perceived as the smallest possible motion of the least pos-

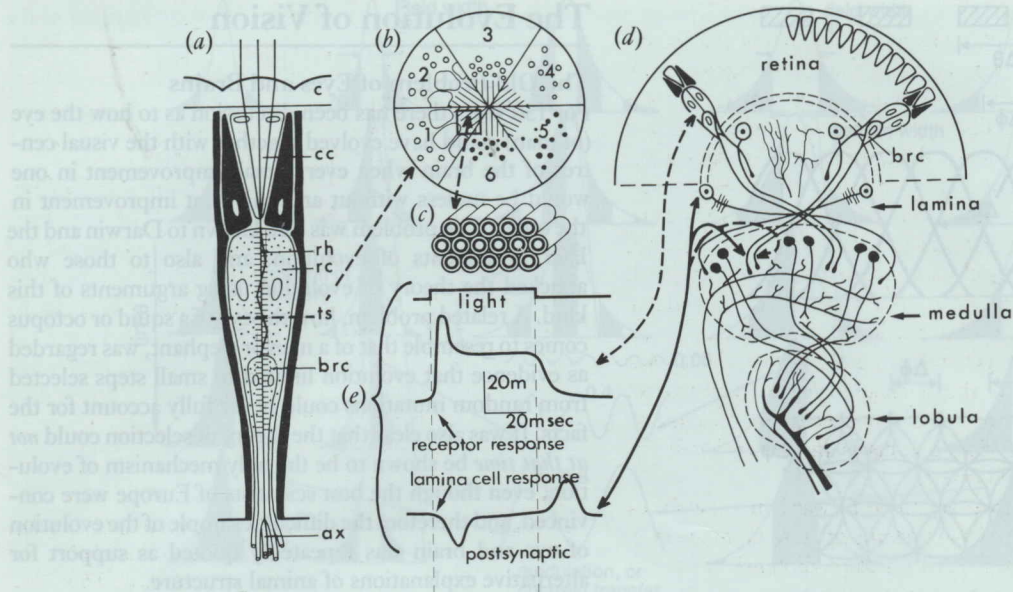


Fig. 11.10 Synopsis of retina-lamina relations. (a) Light is absorbed in the rhabdom. (b) Ommatidium in transverse section, with six receptor cells. (c) Rhabdomere microvilli. (d) Optic lobe in horizontal section showing the relation between the retina and the layers below. (e) Lamina cell responses are approximately inverted temporal derivatives of retinula cell responses.

sible contrast at the appropriate average light intensity.

The common visual behaviour of the medusae and worms which have excellent eyes is simply withdrawal from a moving shadow. Little can be learnt from them about visual processing without data on neurone functions. There are some examples where the behaviour depends on the angle at which the moving contrast lies, and many where motion of the whole visual field has little effect compared with local motion, but few where the directionality of motion is important until we come to the insects and crustaceans. For the practical reason that we know most about them, the insects illustrate very well a level of complexity where there are several different visual responses in one animal, and we also find a variety of visual behaviours when we look at different insects. It has become obvious over the past few decades that insects have several neuronal mechanisms in parallel behind every visual axis. Recording from insect optic lobe neurones shows that these mechanisms for visual detection of colour, direction of motion, polarity of an edge, stimulus orientation, flicker frequency, range and angle at the eye are subsequently channelled into partially separate but overlapping neuronal pathways. This is clearly a result of the history of their staged evolution. The model to be developed depends mainly on our understanding of insect visual mechanisms because this group is the best understood, but many of the concepts apply to all vision.

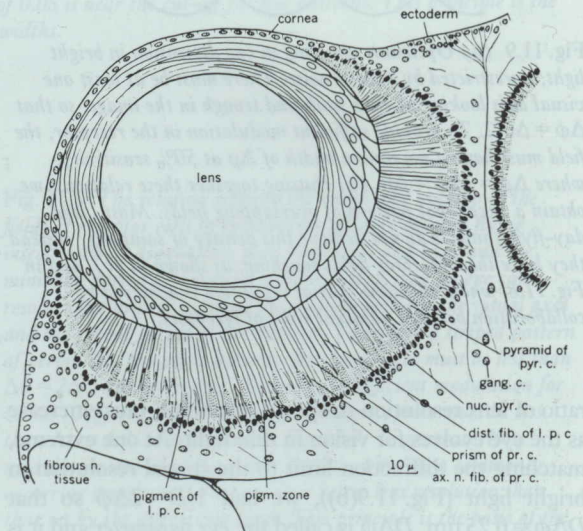


Fig. 11.11 The eye of the Cubomedusan, *Charybdea* (Coelenterata), in section, showing numerous photoreceptors and the typical arrangement of a camera eye, which is ideally a set of sampling stations of optimum spatial resolution and lens resolution. The only visual response of this animal is a reaction to moving shadows by inverting and swimming downwards away from the surface (Berger, 1900).

### The Basis of Vision

Whether they are stationary or move, eyes primarily see transients and motions, not steady states or intensities. A few sessile animals have stationary eyes which detect a small movement of a distant predator and trigger a retreat into a shell or tube. An excellent example is the giant clam, *Tridacna*, with dozens of eyes embedded in the edge of its mantle: tubicolous polychaete worms sometimes have eyes on their head or tentacles with a similar function, associated with a rapid contraction controlled by giant neurones of the nerve cord. The barnacle eye is another well-studied detector of a passing shadow (reviewed by Laughlin, 1981). A general principle is that these eyes rapidly habituate to the average level of background movement and they progressively become more sensitive when no motion is visible. This is the general rule for sensory systems everywhere.

Most animals, however, move forward in a stabilized posture so that their visual systems operate in the context of a predictable flow field. In a three-dimensional world this has several important consequences.

1. At any one moment there is a predictable one-dimensional motion at each point on the retina so that processing need not be two-dimensional.
2. The angular velocity at which contrasts move across the retina (excluding eye rotation) is inversely proportional to their range for a known eye motion, giving an idea of the three-dimensional structure of the surroundings if this angular velocity can be measured.
3. As the eye moves, nearer objects move across the background, causing sharp discontinuities of the flow field at their edges, here called parallax. It is tempting to think that visual processing mechanisms of freely moving animals have evolved to see parallax as a primary feature which provides information about the separation of objects in three dimensions. If so, parallax detection takes over from motion detection as the driving force for evolution of better resolution towards hyperacuity.
4. Vision is based on the motion of the image across the retina, which is projected like a map upon the fixed arrays of processing neurones inside. Therefore the moving features that the neurones at early stages detect at any one place (Fig. 11.15) are transients and there is a major problem as to how any part of the image is 'captured' as it moves across the neural array within the optic lobe.

### Field Sizes

The size of a neurone's field has evolved to match what the neurone is trying to see, especially at threshold. With reference to Fig. 11.12, if we search for an object that differs slightly from background, the signal-to-noise ratio is optimum when the field size matches the object size. In natural situations there has to be a compromise because

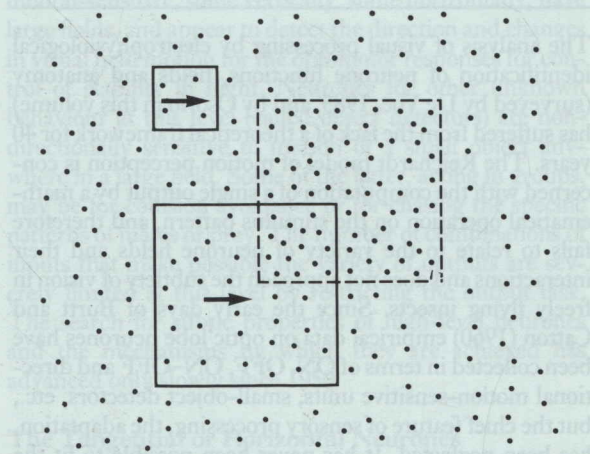


Fig. 11.12 *If a neurone responds to an object that differs slightly from background, the signal-to-noise ratio is optimized if the neurone's field is the same size as the expected object. This principle applies to all features of interest in the image in space and time, so that all visual neurones fields must be evolved towards a spatio-temporal optimum that is related to the visual task.*

the geometry of the image is never so predictable, and to detect the object location there must be many separately line-labelled fields. Therefore, fields cannot be specific. This is a basic principle for the distributed parallel processing within optic lobes.

Neurone field sizes in general in insects are as diverse as they can be, ranging from a single sensory cell input on the one hand to the whole of both eyes in the case of vision. Both extremes raise questions, in that fields corresponding to single receptors somehow have to be integrated spatially in order to abstract the significance of their combinations. Small fields miss the stimulus: large fields fail to locate it: moderate overlap is the usual compromise. One way to explain large fields is to suppose that they are alerting or inhibitory at their outputs, signalling a freeze. Possibly they act only in combination with small-field neurones which require a stimulus that is not specific with reference to direction, so that high resolution at every point is combined with a wide field. Large-field units sometimes relate to behavioural sensitivity towards the whole background in the visual world, as in optomotor responses, but they do not have to do so. Large fields may be a crude way to keep a moving contrast within the field of one neurone. On the other hand, small-field units, even though clearly related to high-resolution tasks, must always be tested on a variety of backgrounds because the detection of a moving contrast of any kind is essentially a detection against background.

## Insect Vision

The analysis of visual processing by electrophysiological identification of neurone functions, fields and anatomy (surveyed by De Voe, 1985, and by Osorio in this volume) has suffered from the lack of a theoretical framework for 40 years. The Reichardt model of motion perception is concerned with the computation of a single output by a mathematical operation on the stimulus pattern, and therefore fails to relate to the variety of neurone fields and their interactions and does not approach the subtlety of vision in freely flying insects. Since the early days of Burtt and Catton (1960) empirical data on optic lobe neurones have been collected in terms of ON, OFF, ON-OFF and directional motion-sensitive units, small-object detectors, etc., but the chief feature of sensory processing, the adaptation, has been neglected. It has never been possible to fit the electrophysiological data into a logical scheme and the known fields of neurones only increase the difficulty in understanding how the moving image is captured (if indeed it is) as it moves fleetingly across the eye (Fig. 11.15).

### The High-Resolution Tracks Through the Optic Lobe

Referring to Figs. 11.8 and 11.9, we observe that the visual fields of width  $\Delta\rho$  of adjacent receptors are narrow and overlap; and the angle  $\Delta\phi$  between visual axes sets the limit of spatial resolution of the eye as an angle-labelled sampling array. The eye cannot distinguish the separate bars of a striped pattern of period less than  $2\Delta\phi$  subtended at the eye, and single receptors receive no modulation from (are not stimulated by) a striped pattern of period narrower than about  $\Delta\rho$  subtended at the eye. We find that many eyes have evolved compromise optics and anatomy such that the field width  $\Delta\rho$  approximately equals the receptor spacing  $\Delta\phi$ . As more complex subretinal mechanisms evolve, the receptors are able to map the visual scene into them with the high spatial resolution that was evolved for directional and non-directional motion detection.

In insects three further stages of neurones as far as the lobula retain this topographical map at the highest spatial resolution (Fig. 11.13). First, the function of the lamina monopolar ganglion cells, so far as we know it, can be summarized as transforming intensity into the temporal derivative of intensity (Fig. 11.10) with minimal latency while retaining the maximum spatial resolution (Laughlin, 1981, 1987). The next level in insects, the medulla, is where the excitation spreads into a large number of small neurones arranged in columns on each visual axis. These neurones presumably respond in different combinations to different local details of the stimulus pattern. In the

small fly *Drosophila* there are about 30 intrinsic neurones in each of the vertical columns which correspond to a visual axis in the eye and a greater number of column neurones with axons to the lobula (Fischbach and Dittrich, 1989). This anatomical diversification into many small neurones recently led me to formulate the template model in which there are on each visual axis a number of

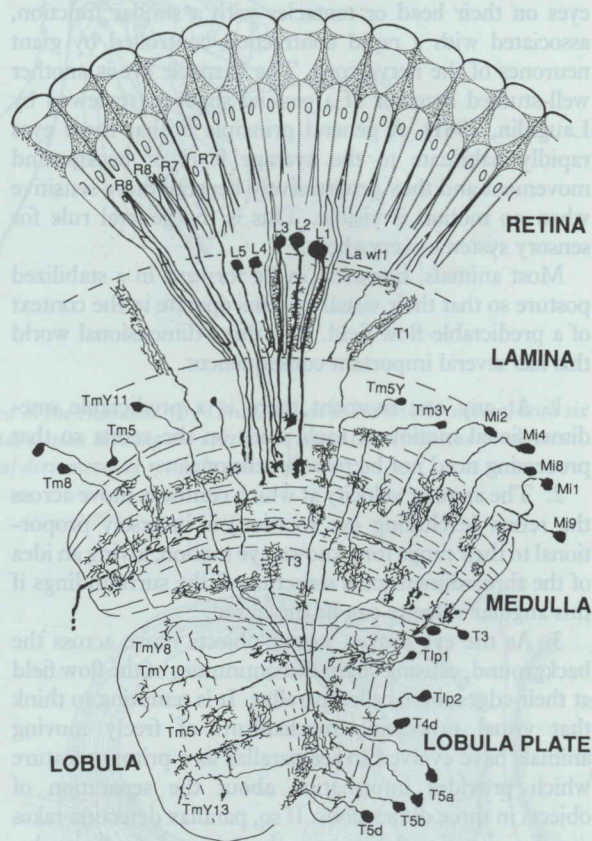


Fig. 11.13 *Neurones of restricted dendritic spread in the insect optic lobe, mainly based on the fly Drosophila.* Narrow field receptors of the retina feed into columns of the lamina. From here the axons of the lamina ganglion cell types (L1–L5) transmit high-resolution information to columns of the medulla. The long retinal fibres of receptors R 7 and R 8 run direct to the medulla. The chiasma between lamina and medulla has been omitted for clarity. So far all these neurones have no spikes, only electrotonic spread. In the medulla are 30–50 local neurone types in each column, which corresponds to a visual axis. Some of these project to the lobula, preserving much of the spatial resolution but the mapping of the retinal projection continues no further. Every level in this columnar projection has tangential fibres with long arborizations in specific layers and also medium-field neurones of several types. Large-field and small-field directional and non-directional motion detectors occur in the medulla and lobula but object detection appears to begin in the lobula (Redrawn from Fischbach and Dittrich, 1989).

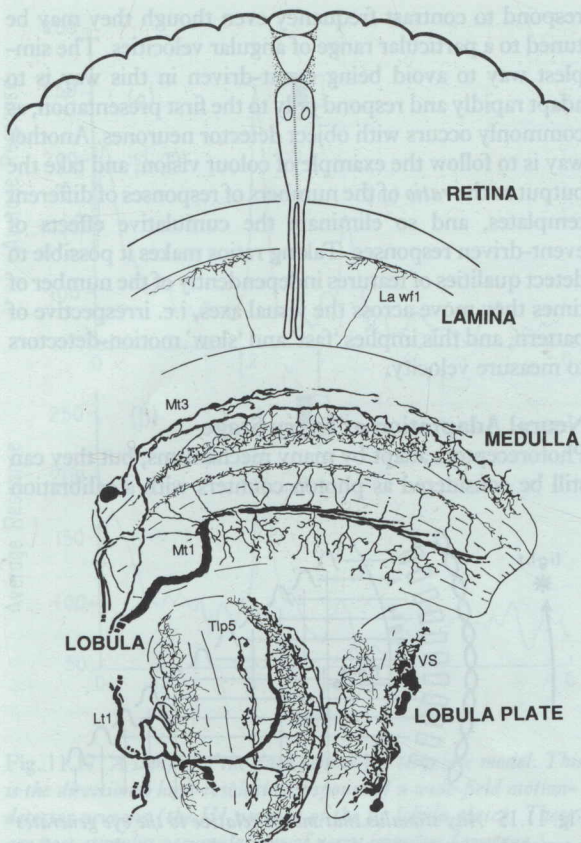


Fig. 11.14 Horizontal fibres with widely spread arborizations in tangential layers. There is a striking similarity to the collector neurones in the ensemble processing system of Fig. 11.28. Labels on neurones here and in Fig. 11.13 after Fischbach and Ditttrich (1989).

templates that respond differently to the different local spatio-temporal patterns in the visual scene (see below). At various horizontal strata in the medulla there are also wide-field neurones (Fig. 11.14) which are sensitive to motion, usually directionally. It is important to stress that the full retinal resolution appears to be used only for motion detection and not for colour vision, object vision, polarized light, or any other visual function that we know of. Only two types of simple behaviour, both obviously not restricted to insects, retain the full spatio-temporal resolution of the eye. They are the directional and non-directional responses to motion. All other vision has poorer resolution, as if many templates are involved. There is no evidence that the *positions* of edges are located by anything resembling a 'zero-crossing' mechanism.

Corresponding to the visual axes, anatomically small-field neurones arranged in columns continue into the third neuropile of the optic lobe, the lobula (Fig. 11.13). At this level the most obvious neurones collect from large num-

bers of the more peripheral units. They are directionally motion-sensitive, some vertically, some horizontally, have large fields, and appear to detect the direction and changes in visual field motion for the optomotor responses for control of stability in flight. Neurones for other unknown behaviour at this level (called object detectors) are non-directionally sensitive to motion of a small object anywhere in a large field. Some of the latter, acting in groups, may be for collision avoidance or for chasing the specific patterns of mates or prey. The number of combinations of inputs that make possible the variety of outputs are severely limited at this level by restricting the output task. The search for subtle properties of high-level neurones and the mechanisms by which they are achieved has advanced only slowly since 1950.

### The Tangential or Horizontal Neurones

Insects and crustaceans are typical in not having efferent fibres to the retina ending within the retina on the receptors. Insects have tangential fibres of unknown function spreading through the lamina, sometimes in more than one stratum. There have been suggestions that efferent fibres to the lamina control the relative gain of the inputs from different colour receptors, or local sensitivity, but no effects of this kind have yet been recorded from the lamina monopolar cells.

Numerous different neurones run in horizontal strata of the medulla, lobula and lobula plate (Fig. 11.14), some of them efferent. So far, all those recorded in the medulla have been wide-field collector neurones from large numbers of local motion-detectors, either directional or non-directional, with high spatial and temporal resolution. The directional neurones have a maintained response to continued motion, and are related to optomotor responses rather than to feature vision. Possibly they are gating neurones which allow other circuits to function only when the general motion is in the specified direction. These large-field optomotor neurones are especially well known from the lobula plate of the fly (Hausen and Egelhaaf, 1989; Strausfeld, 1989) where they respond as if they collect from a large number of high-resolution directional detectors separately for horizontal and vertical motions.

So far as we know, object-detector neurones tend to be complexly ramifying, not limited to one stratum, and are located in the lobula or brain. Almost nothing is known about identified neurones with small or medium-sized arborizations in the deep optic lobes (Osorio, Chapter 10).

## The Evolutionary Approach

### Evolution of Visual Processing

The progressive evolution of visual processing is the progressive addition of new neurones into a mechanism that

already operates at high spatial resolution with many receptors in parallel. Possibly, in compound eyes, there was also a reduplication of visual axes after some mechanisms of processing had appeared. The first stage, found in many lower invertebrates, has many photoreceptors backed by many local detectors of moving shadows, all feeding into one circuit that withdraws the animal into deeper water, or into a hole. Next, we have small-field motion-detectors that are a little more subtle than shadow detectors; then we find *directional* motion detectors which are asymmetrical. At about this level of evolution, we find an indifference to total background motion and different colour sensitivities appear, although directional motion-detection is apparently colour blind in insects.

Never can we contemplate a primitive starting point in which the visual processing mechanism was able to analyse or 'see' the whole picture or discriminate large numbers of patterns. Vision obviously evolved by progressive addition of task-directed ways of analysing the visual world, and the addition of functions implies the addition of new neurones. The first requirement was to detect a small number of relevant features, a task that can be done by a small number of neuronal templates that cover these features. Nor was it ever necessary for a primitive visual system to analyse the flow field caused by its own motion. Only a few aspects of the flow field are useful. It is tempting to imagine that the directional neurones first acted as non-adapting gates so that other channels function only during motion in the appropriate direction. This would fit in with their large fields and wide arborizations. A useful ability for a moving eye is the measurement of local angular velocity along one expected line of motion, to estimate range, to interpret the three-dimensional world and distinguish solid objects from patches of light and shade by use of parallax. Looming is also a convenient signal for impending collision, but, again, the analysis can be one-dimensional in each region of the eye.

The main problem with introducing templates matching the biologically significant features in the visual scene is that all template responses are 'event-driven', by which I mean that the template responds every time its combination of contrasts in the visual field passes its visual axis. If template responses are summed in any way in the animal or in an experiment (as in Fig. 11.17), the total response is dependent on the temporal frequency of the appropriate trigger features. At a higher level, a butterfly with templates that respond to flowers would need a mechanism to cope with a whole bed of flowers. This event-driven property has been known for many years in directional motion-detector neurones of the fly (Fig. 11.17), which respond to the passing of each edge: they therefore respond more to the movement of groups of edges than to single edges, so that the summed response detects direction of motion but cannot measure velocity independently of pattern. They

respond to contrast frequency even though they may be tuned to a particular range of angular velocities. The simplest way to avoid being event-driven in this way is to adapt rapidly and respond only to the first presentation, as commonly occurs with object detector neurones. Another way is to follow the example of colour vision, and take the output as the *ratio* of the numbers of responses of different templates, and so eliminate the cumulative effects of event-driven responses. Taking ratios makes it possible to detect qualities of features independently of the number of times they move across the visual axes, i.e. irrespective of pattern, and this implies 'fast' and 'slow' motion-detectors to measure velocity.

### Neural Adaptation at Every Stage

Photoreceptors adapt by many mechanisms, but they can still be considered as photon counters with a calibration

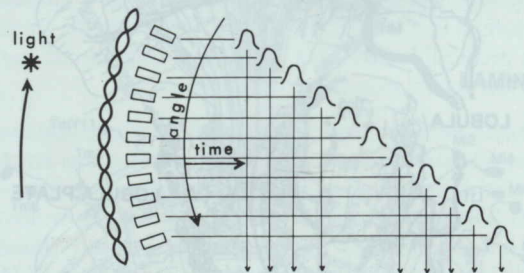


Fig. 11.15 Any stimulus that moves relative to the eye generates a series of responses spread out in time and space in the sampling array. Therefore visual processing has to be considered in spatio-temporal coordinates, and one of the main questions in vision studies is how the image is tracked across the spatio-temporal array within the nervous system.

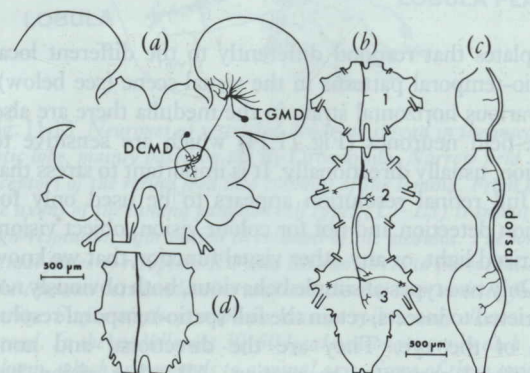


Fig. 11.16 The DCMD (descending contralateral motion detector) neuron of the locust and its feeder neuron the LGMD (lateral giant). These neurones are detectors of non-directional movement of any small contrasting object at the full spatial resolution of the retina but are little influenced by a large background motion. The peripheral connections run to flight motor neurones, and the probable function of this neurone is the rapid avoidance of obstacles when in flight.

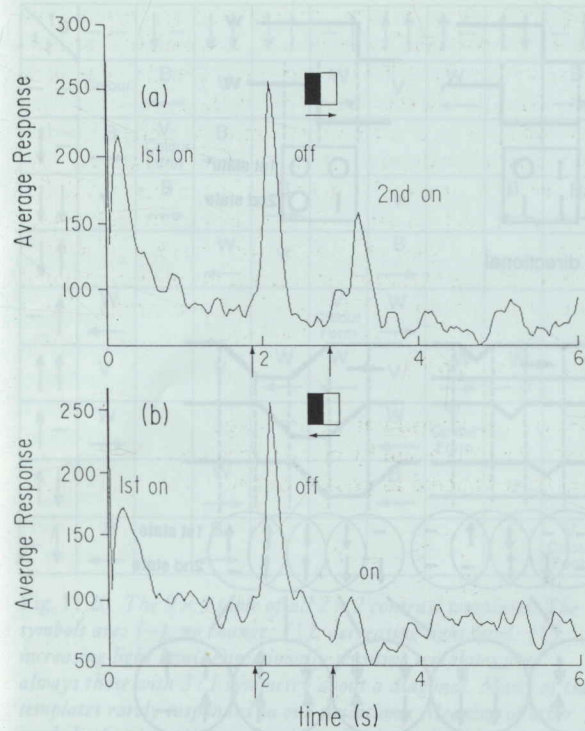


Fig. 11.17 A sample of the data behind the template model. This is the directional high-resolution response of a wide-field motion-detector neurone (the H1 neurone of the fly lobula plate). These are post-stimulus accumulations of nerve impulses (average responses). On the left at  $t=0$  the pattern with a single edge appears on the screen. At  $t=2$  s the pattern disappears and is replaced by plain grey, then 1 s later the pattern reappears but displaced by  $1.5^\circ$ . The response is shown (a) in the preferred direction, (b) in the antipreferred direction (note the inhibition of background impulses). This effect shows that the first position of the pattern is 'remembered' somehow for at least 1 s although there is a large OFF response at  $t=2$  s. Therefore opposite effects are propagated in opposite directions from the edge, causing the directionality, and in the absence of the stimulus these effects persist for a short time at the spatial resolution of the visual axes. Compare the templates in Fig. 11.18.

that can change. Some of the adaptation, caused by membranes and ions, is rapid; other adaptation, caused by movement of the receptor cells or of screening pigment, is slower, but all adaptation helps to keep the photoreceptor near the middle of the intensity range of the ambient light at the appropriate time of day. Photoreceptors can be modelled as photon counters with adaptation.

Lamina ganglion cells adapt so rapidly that they act like differentiating circuits with an output proportional to the temporal derivative of the intensity modulation in the photoreceptors on the same visual axis. Some photoreceptor axons bypass this stage and run direct to the next

neuropile, the medulla. So far as they are known (Mimura, 1972; Honegger, 1978; Osorio, 1987, 1990;), many of the small-field cells of the medulla also adapt rapidly (Osorio, this volume, chapter 10).

The higher-order neurones adapt even more obviously, especially those sensitive to motion of a small object anywhere in a larger field. Some respond only to a novel stimulus and, at least in restrained preparations, require a long recovery period before they respond again. The DCMD and LGMD neurones of the locust (Fig. 11.16) are examples of this type (Rowell, 1971; Rind, 1987). The directionally sensitive optomotor neurones of large insects (e.g. Hausen, in Ali, 1984; Rind, 1990) continue to respond for long periods to a steadily maintained motion, but they often respond much better to a change in velocity.

## The Template Model

### The Template Model of the Optic Medulla

Visual processing necessarily involves temporal and spatial correlations (Fig. 11.15) between photoreceptors that are adjacent on the retina, otherwise the temporal and spatial resolution is wasted. The second-order neurones of the insect lamina effectively take the temporal derivative of the photoreceptor output on each axis, at high gain and minimum latency. We also know that in motion detection there is a rapid saturation of contrast (Horridge and Marcelja, 1990b), although there may also be other separate mechanisms which actually measure contrast. Therefore we take a threshold temporal contrast of  $0 \pm 0.008$  on each visual axis, and base the model upon these threshold changes at adjacent visual axes at successive instants of time (Sobey and Horridge, 1990). The model is one-dimensional in space because the insect's own stereotyped motion generates one-dimensional motion at each point on the retina. An increase of intensity over the threshold at a single photoreceptor is ( $\uparrow$ ), a decrease is ( $\downarrow$ ) and indeterminate or 'no change' is ( $-$ ). There are nine possible pairs at adjacent axes in one dimension, namely ( $--$ ), ( $\downarrow-$ ), ( $\downarrow\downarrow$ ), ( $\downarrow\uparrow$ ), ( $-\uparrow$ ), ( $\uparrow\uparrow$ ), ( $\uparrow-$ ), ( $\uparrow\downarrow$ ), ( $-\downarrow$ ). When we take nine pairs at two successive times (Fig. 11.20) we obtain 81 spatiotemporal  $2 \times 2$  templates which are the smallest possible primitives for spatio-temporal analysis of the visual scene when contrast is thresholded. This is the simplest possible way that all quantized contrast changes can be included in time and space with full resolution.

The templates differ from neurones in that every group of two adjacent pixels at adjacent times causes the response of only one of the templates (Fig. 11.18), but real neurones in the columns behind one visual axis can respond in parallel in different numbers in various combinations simultaneously, and can yield graded responses as well. In

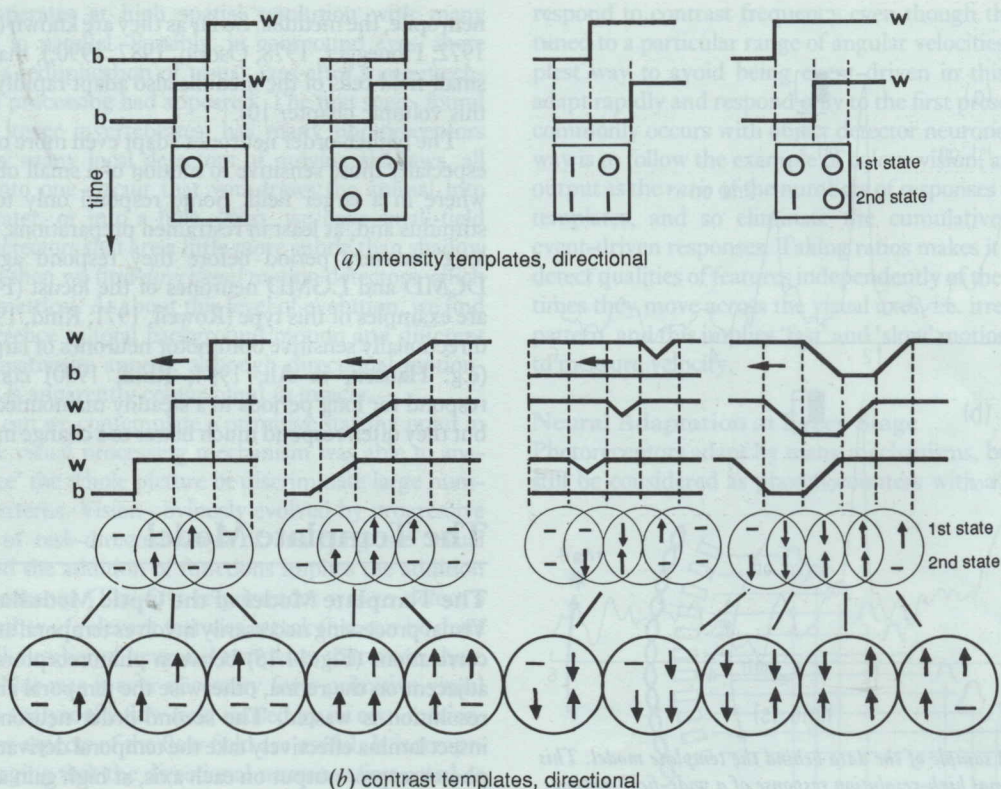


Fig. 11.18 The way that the template notation relates to the motion of a contrast in the outside world. *b*, black or dark level. *w*, white or light level. (a) Intensity templates; 0 represents dark and 1 represents light. The template consists of the state (either 0 or 1) at two adjacent visual axes at two successive instants. The diagonal 3 : 1 symmetry of the spatiotemporal (a) template is characteristic of directional sensitivity. (b) Contrast templates are preferred because they cope with the wide range of background intensity. Any increase in intensity is ( $\uparrow$ ), any decrease ( $\downarrow$ ) and 'no change' is (-). Sharp edges, as on the left, are uncommon in real eyes and natural images consist mainly of gradient edges. All of the templates shown here are directional for motion to the left, as shown by the symmetry about a diagonal (see Fig. 11.23, 11.24). A group of template responses are necessary to convey the nature of the moving contrast. The template idea shows how spatial image structure in motion is fed into an array of processing units.

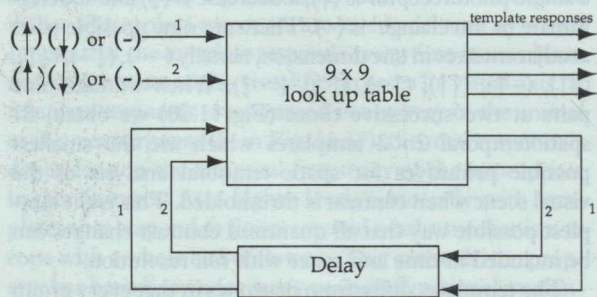


Fig. 11.19 One way to correlate adjacent pairs at successive instants is to arrange a state machine with a delay in the feedback loop. A neuronal circuit would presumably have a persistent transmitter with the delayed effect.

fact, I expect that many of the small neurones of the optic neuropiles are non-spiking neurones with mutual inhibitory interactions similar to those described by Burrows (1980) and his associates in the control of locust walking. The templates in the present model are therefore very simplified representations of parts of neurones or groups of neurones, but they illustrate the principles of parallel processing in discrete time steps by an artificial or natural system of this type. One way to visualize the state machine on each visual axis is shown in Fig. 11.19.

### The Significance of 'No Change' (-) and 'No Response'

In the development of the theory, the idea of 'no change' (-) arose naturally from the indeterminate region below



$t_0 \backslash t_1$	--	↓	↓	↓	↑	↑	↑	↑	↓	↓
--	Colour	B	V		W	V	W			B
↓	B	V Colour Form	B							
↓	V	B		B		V		B	B	
↓	↑		W	V		B				
↑	W				V Colour Form	W				
↑	V		V	W	W	V	W	W		
↑	W					W	V Colour Form			
↑	↓		W			B		V		
↓	B		B							V Colour Form

Fig. 11.20 The 9 × 9 table of all 2 × 2 contrast templates. The symbols are: (–), no change; (↓), decreasing light level; (↑), increasing light level. Directionally sensitive templates are always those with 3 : 1 symmetry about a diagonal. Many of the templates rarely responded in our conditions. Meaning of other symbols: (→), motion to the right; (←), motion to the left; (B), dark follows light in direction of motion; (V), templates that indicate velocity by their ratios.

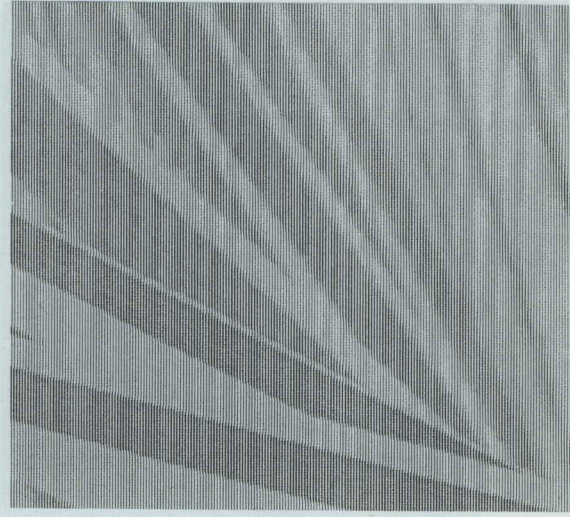


Fig. 11.21 The primary data of the one-dimensional image at successive times. This spatio-temporal printout is 200 pixels wide, covering a field of view of 10.5°, and 180 successive instants from top to bottom, over a period of 7.2 s at a sampling rate of 25 frames per second. In the picture are five targets plus a background, located at 180, 380, 740, 1210, 1640 and 2660 mm from the camera. This picture is a portion of the whole image 512 pixels wide.

threshold. Subsequently, because not all templates are employed, we may have many points in the spatio-temporal map with no template response. In general, absence of activity is significant in behaviour in three ways: firstly, so that a behaviour pattern can proceed in the absence of a veto or inhibitory gating; secondly, in the progressive increase in sensitivity with time that occurs in the absence of stimulation; thirdly, in leaving processing channels undisturbed for other types of discriminations to occur. These generalities apply to behaviour patterns and to single neurones. Conversely, repeated stimulation raises thresholds, sometimes precipitately, and can cause habituation for long periods, especially in higher-order neurones.

Also, 'no change' is significant because colour vision depends on the ratios of graded responses derived from different colour types of photoreceptors, and the ratios are more easily measured in undisturbed channels separate from those conveying temporal contrast.

### Implementation of the Model

To obtain data we take a real scene and scan it at 25 Hz with a single horizontal line of pixels in a moving CCD camera. The scene changes as a result of the horizontal motion of the camera, so that successive scans by the same

line of pixels generate a spatio-temporal picture of graded intensity in which motions of contrasts are seen as diagonal lines (Fig. 11.21). At this stage we can, if we wish, introduce some lateral inhibition which influences later processing (see below). To obtain temporal contrast we subtract the intensity at each pixel from the intensity in the same pixel in the previous scan and take a threshold contrast level so that

$$(\uparrow) \text{ is greater than } +0.008, (\downarrow) \text{ is less than } -0.008 \text{ and } (-) \text{ is } 0 \pm 0.008.$$

This threshold is arbitrarily found by adjustment above the noise level of the camera, so that small contrasts are detected but spurious responses are minimized.

The resulting digitized spatio-temporal map of contrast (Fig. 11.22) is then scanned with selected templates for particular primitives (Fig. 11.23). Let us examine the mechanism in detail. The motion of a graded contrasting edge to the right is represented pixel by pixel in  $x, t$  coordinates as a series of steps (Fig. 11.24(a)). The template (→)/(↓) responds to exposed corners on this profile and the template (↓-)/(↓) to inside corners: both of these templates have light-dark polarity and are directional because these particular corners occur normally with the motion of an edge with decreasing intensity towards the right. Directional templates are event-driven and therefore their total number of responses measures the number

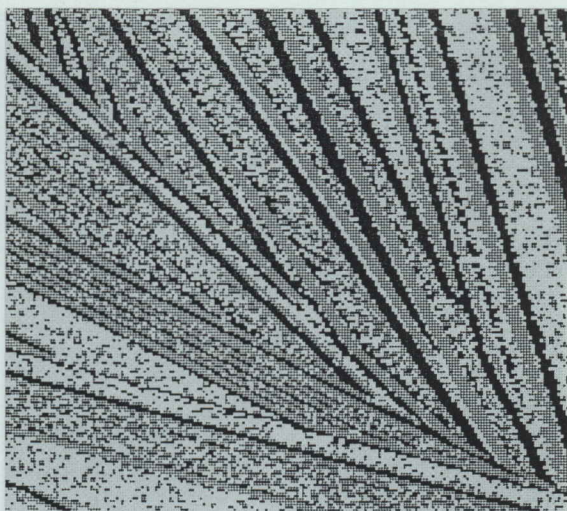


Fig. 11.22 The same spatio-temporal data as in Fig. 11.21, with the 8-bit intensity at each pixel subtracted from that at the previous instant on the same pixel. The threshold at a contrast of  $0 \pm 0.0008$  is then applied and the three resulting states are plotted; (·) 'no change' or indeterminate; (■) decreasing and (►) increasing intensity. This spatio-temporal map is then compared, pixel by pixel, with the templates in Fig. 11.20. Note that gradients at edges are often 3 to 5 pixels wide as a result of the preprocessing.

of edges that pass irrespective of pattern. They indicate the direction of motion reliably and respond instantly to transients. Directional templates such as  $(-)/(\downarrow-)$  and  $(\downarrow-)/(\downarrow\downarrow)$  respond in equal numbers (Fig. 11.24(a)), and their ratios do not measure velocity.

The templates  $(-)/(\downarrow\downarrow)$  and  $(\downarrow-)/(\downarrow-)$  respond differently, and are not directional. We see from Fig. 11.24 that the local ratio of the numbers of responses of these, and related non-directional templates, is a measure of the angular velocity. Setting the threshold low reduces the count of responses of templates such as  $(-)/(\downarrow-)$  and  $(- \downarrow)/(\downarrow\downarrow)$  with a 'no change' symbol. When this symbol is not used, vision is still possible, although only 16 possible template types remain. These 16 include directional ones such as  $(\downarrow\uparrow)/(\downarrow\downarrow)$  and non-directional ones such as  $(\uparrow\uparrow)/(\downarrow\downarrow)$ , so that measurement of direction and velocity in separate later channels is still possible. Also, templates for moving dark-light edges are necessarily the mirror images of those for light-dark edges, so that edges of opposite polarity are separately processed.

This model has many points of similarity to natural visual systems as seen in insects. Only a few types of templates need be utilized for limited vision. Evolution of such a system is easy because additional templates can be brought into use while others drop out. The observed greater number of neurones as evolution progresses is interpreted as the evolution of more templates. Template

responses can be gathered up in subsequent combinations to suit more complex visual tasks. The way to keep down the numbers of combinations of template responses in processing of this type is to make use of only the combinations needed for the behavioural output, the properties of the visual world, and the features to be abstracted.

### Properties of Templates

Template responses are highly non-linear, and the range of their necessary variety is compatible with our knowledge of optic lobe neurones. The effect of converting sharp edges to gradients by convolution with the visual fields of the receptors, and of emphasizing contrasts by adding some lateral inhibition, is to change the selection of templates that subsequently respond. These mathematical operations upon the image should be regarded as biologically significant, not because they reject high or low frequency in the spatial information of the visual scene, as stated for the past century or so, but because they allow better scope for separating template responses deeper in the processing mechanism.

The responses of individual templates are insufficient to convey much information and are ambiguous with reference to their stimulus, but this is an essential feature of a distributed processing mechanism. Groups of templates evolve together and templates are therefore incomplete in isolation. A system with a few essential templates on each visual axis could generate the behaviour that we observe in insects but, of course, would not analyse flow fields or stationary patterns in two dimensions. The response of a template is a unit of energy which is fed into the next stage of processing: a directional template generates a unit vector impulse at that point in the visual field of the whole eye. Preformed templates are an ideal mechanism for saving time in responses to transient presentations because all pathways act in parallel and any number can add their vector input at the same time. Templates avoid computation, do not take averages, and improve signal and reject noise by use of a greater number of templates, not by deeper computation.

To illustrate templates in action, we recorded an actual scene with a moving camera and one line of pixels so that the spatio-temporal data can be represented on a page with responses of certain templates superimposed on the primary data (Fig. 11.23). Templates of the form  $(- \downarrow)/(\downarrow\downarrow)$  and  $(\uparrow\downarrow)/(\downarrow\downarrow)$  respond directionally to moving edges of one or other polarity. Templates of the form  $(- -)/(\downarrow\downarrow)$  and  $(- \downarrow)/(- \downarrow)$  give a measure of velocity by the ratio of their responses. Templates of the form  $(- \downarrow)/(\downarrow-)$  rarely respond because sharp edges have become gradients (Fig. 11.18).

### Counting Line-Labelled Template Responses

The template operation separates significant primitives in the visual scene into different lines – a process called *line*

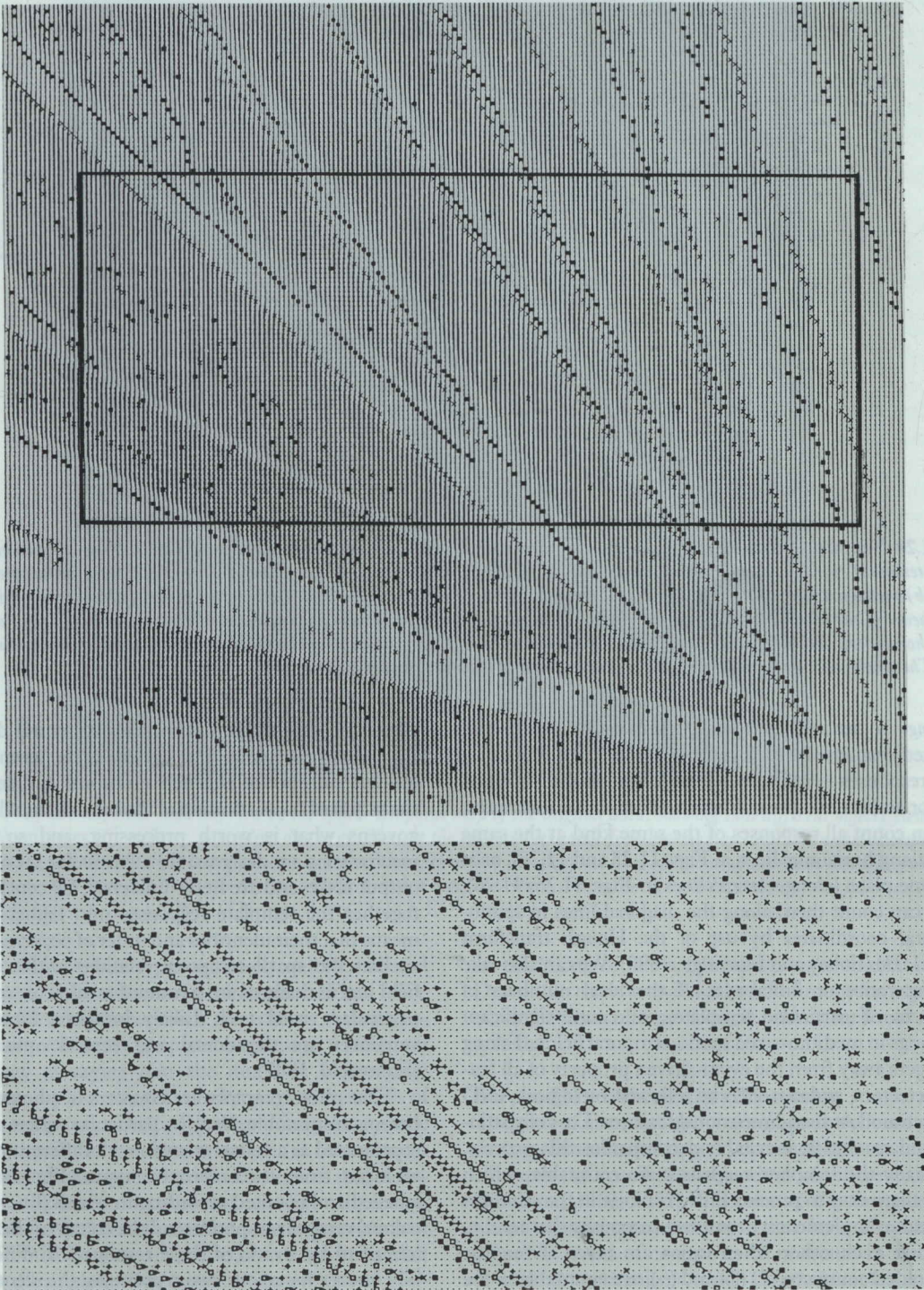


Fig. 11.23 The spatio-temporal scene as in Fig. 11.21 with some template responses superimposed. The lower part of the picture is the outlined area at higher magnification and with lateral inhibition added before the threshold stage. We observe that different edges have different ratios of template responses. Templates are as follows:  $\blacksquare = (\uparrow\uparrow/-\uparrow)$ ,  $\times = (\downarrow/-\downarrow)$ ,  $\succ = (\downarrow\downarrow/-\downarrow)$ ;  $\square = (\downarrow\downarrow/\uparrow\downarrow)$ ,  $\blacktriangleright = (\downarrow\uparrow/\downarrow\downarrow)$  and  $+$  =  $(\downarrow\downarrow/\downarrow\uparrow)$ .

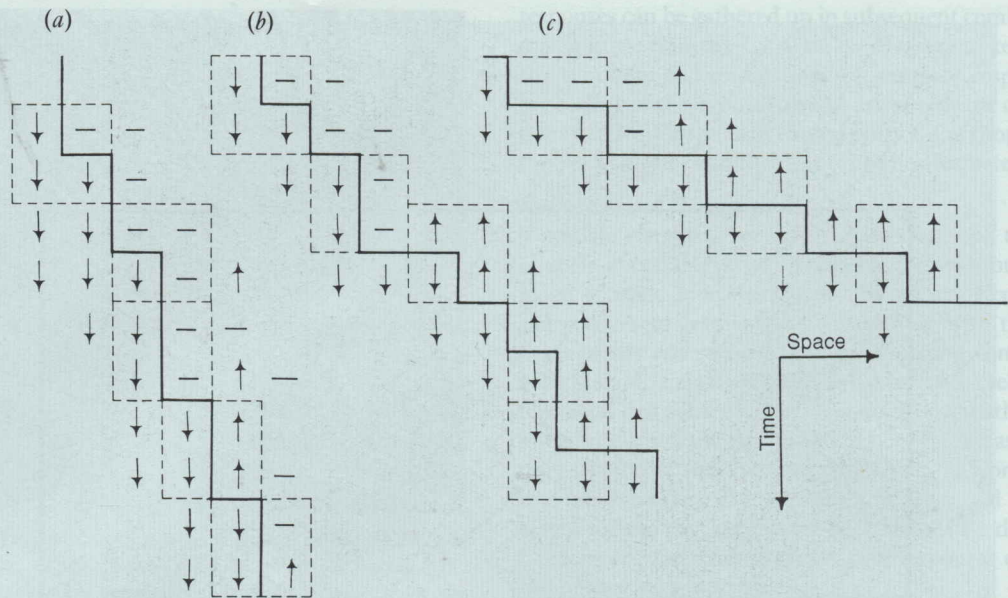


Fig. 11.24 The motion to the right of a contrast (an increased darkening) as seen at high magnification by a one-dimensional eye in spatio-temporal coordinates. ( $\uparrow$ ) increasing intensity, ( $\downarrow$ ) decreasing intensity, ( $-$ ) no change. Three different velocities are shown (a) slow, (b) medium and (c) fast. The directional templates with 3:1 diagonal symmetry detect the corners (which are an arbitrary consequence of the pixel-interval pattern). The ratios of the number of responses of the non-directional templates ( $\uparrow\uparrow/\downarrow\downarrow$ ) and ( $-/\downarrow\downarrow$ ) on the horizontals, to the ( $\downarrow-\downarrow-$ ) and ( $\downarrow\uparrow/\downarrow\uparrow$ ) on the verticals, gives a measure of the velocity but at a lower resolution in space and time. The whole spatio-temporal map can be broken down to template responses in this way.

labelling in sensory systems – and each line carries digitized responses which can be counted. To reduce all these responses in numerous lines to a simple set of decisions or actions, we can summarize them in various ways. We can count all responses of the same kind at the same time (along horizontal lines in Figs. 11.25 and 11.26), or at the same place (along vertical lines in Figs. 11.25, 11.26), or we can count them along the diagonal lines in spatio-temporal coordinates by use of the Hough transform (Fig. 11.33) (Sobey and Horridge, 1990). We can also take running ratios or differences of the local numbers of responses of particular templates, or count significant groupings in medium-sized spatio-temporal fields. Taking ratios gets away from the event-driven property of all feature detectors, and can then be followed by logical AND to accommodate two-dimensional features. Loss of spatio-temporal resolution inevitably accompanies this process. Essentially the template responses are countable, which makes them easy to process further. Noise in the stimulus generates complementary template responses that tend to neutralize each other when ratios are taken. Even the thresholded contrasts, ( $\uparrow$ ) ( $\downarrow$ ) and ( $-$ ), are countable in the same way. The higher-level neurones are now postulated to be leaky counters of template responses, just as the photoreceptors are leaky counters of photons. The number of different templates on each visual axis is a mea-

sure of the ability to discriminate structural diversity by the visual processing. The variety of the templates counted and number of alternative outputs is determined by the required sophistication of the behaviour, which in turn governs what is worth processing, and so semivision mechanisms are designed for a particular set of input and output tasks, and are never universal. Of course, bringing together the template responses into higher-level fields means that their identities and order of occurrence are lost, so that the spatio-temporal relations within the higher field are exchanged in favour of a decision relating to a particular visual task. Likely locations for these convergences are the dorsal and ventral optic glomeruli (Strausfeld, 1989).

That photoreceptors are counters of photons has long been a respectable idea, in particular when colours are represented as a colour triangle with different proportions of input from three receptor types. Photons are line-labelled and counted according to which (line-labelled by colour) receptor responds, and the colour of the stimulus is identified by a ratio of these responses in different lines. The theory for colour vision preceded the identification of the spectral types of receptors and their quantitative description. If we follow the same method to analyse vision of motion and form, by template response ratios, the next step is to identify which templates are used. We

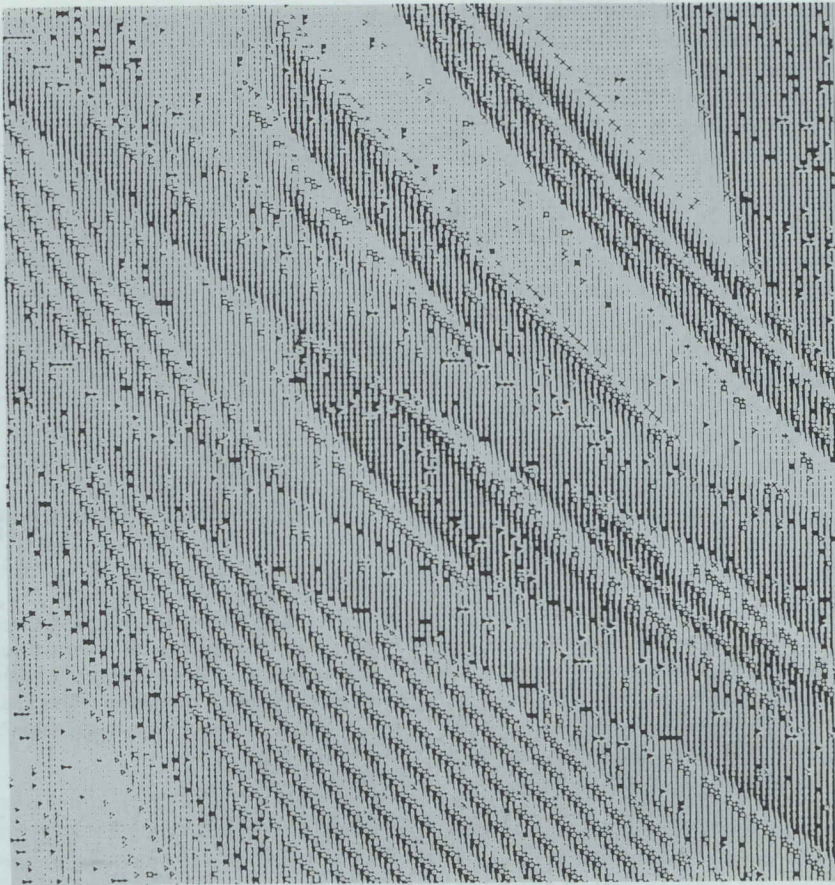


Fig. 11.25 *Template responses superimposed on the spatio-temporal map of the original data, showing that different objects yield characteristic mixtures of templates, depending on the gradient at the edge and its range from the moving camera. In this example the camera is moving towards the objects and passing by one of them; therefore the diagonal lines of motion of contrasts through space and time are curved. At bottom left the eye is passing an object with a regular texture. The meanings of the symbols are as follows: ■ = ( $\downarrow\downarrow/\uparrow\uparrow$ ),  $\succ$  = ( $-\downarrow/-\downarrow$ ),  $\blacktriangleright$  = ( $\uparrow\uparrow/--$ ); + = ( $\uparrow\uparrow/\downarrow\downarrow$ ),  $\square$  = ( $\downarrow/\downarrow\downarrow$ ),  $\times$  = ( $\uparrow-\uparrow\uparrow$ ).*

can represent four colour types or equally well four related template types at the corners of a tetrahedron (Fig. 11.27). Points at different distance from the corners then represent different ratios of template responses, which is a pictorial way of saying that the numbers of template responses can be processed in similar ways to numbers of photons in receptors, as well as by adaptation, antagonistic interaction, and temporal summation at subsequent stages. In the colour system we find receptors and neurones with responses that depend on light intensity, although final colour discriminations are independent of intensity. Similarly, the early directional neurones that respond to moving edges are strongly dependent on contrast frequency, being driven by each edge event that passes. Nevertheless the freely moving insect is able to discriminate features of edges irrespective of contrast frequency. In both cases the final visual behaviour depends

on the taking of ratios at a location *after* these neurones. Down this track we will find a lot of room for exploration of mechanisms of semivision in both natural and artificial systems.

#### Across-fibre Information Processing

Theories of chemoreception long ago introduced the idea that smells and tastes depend on simultaneous stimulation of clusters of specific combinations of different neurones (Erickson, 1963, 1982). Theories of the function of the mammalian hippocampus or cortex have also been presented in terms of ensembles of neurones in distributed processing mechanisms which carry a complex pattern in chemosense, touch or vision, by having information distributed in parallel in many neurones. The models usually envisage physically orthogonal arrays of two sets of neurones, usually considered to be in columns and in layers

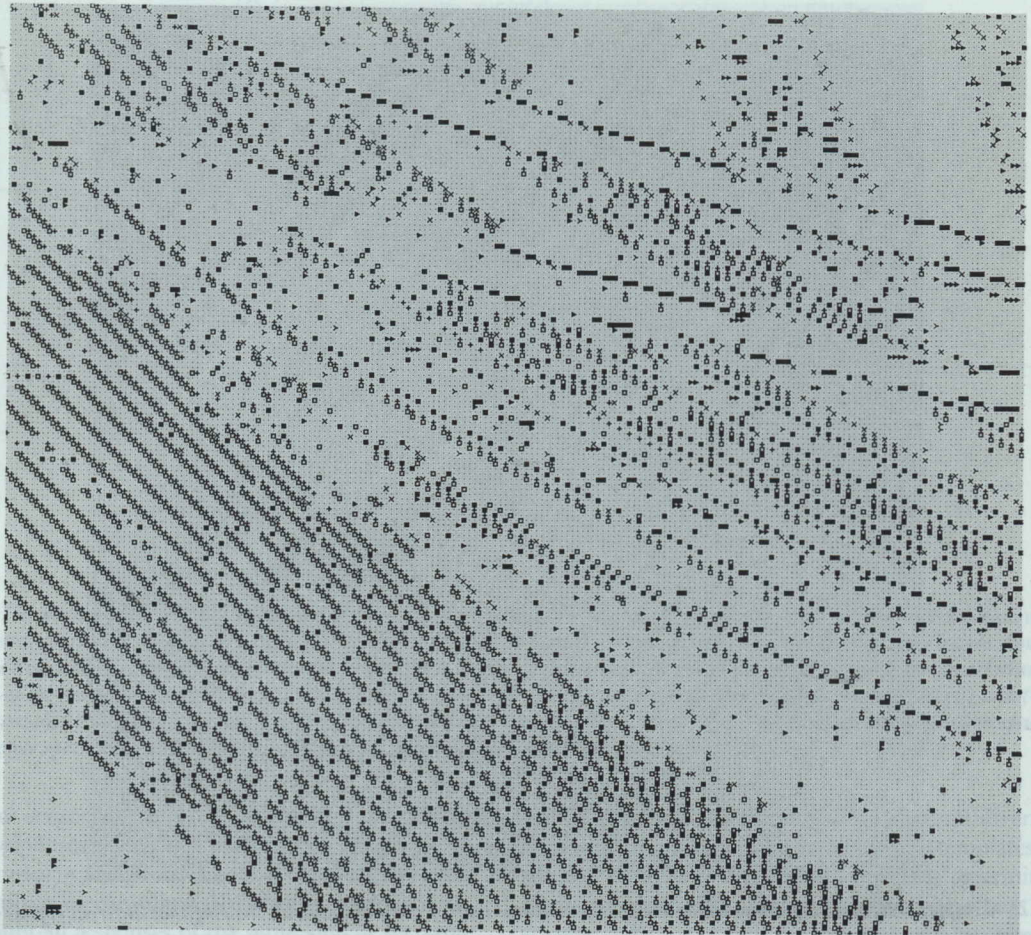


Fig. 11.26 *Template responses in a spatiotemporal plot as in Fig. 11.25, but for clarity the background of the original data (Fig. 11.21) is omitted. The meanings of the symbols are as follows:  $\square = (\downarrow\downarrow/\uparrow\uparrow)$ ,  $\triangleright = (\downarrow\downarrow/-\downarrow)$ ,  $\blacktriangleright = (\downarrow\uparrow/\downarrow\downarrow)$ ;  $\times = (\downarrow/-\downarrow)$ ,  $+$   $= (\uparrow\uparrow/\uparrow\uparrow)$ ,  $\blacksquare = (\uparrow\uparrow/-\uparrow)$ .*

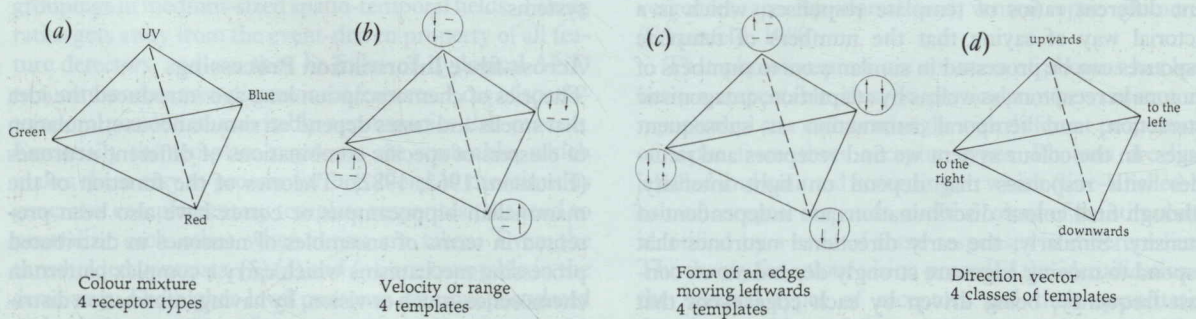


Fig. 11.27 *Because template responses are line-labelled and can be counted, groups of them can be represented in the same way as photon counts in receptors of different colour sensitivity. Here a few possibilities are illustrated ((a)–(d)).*

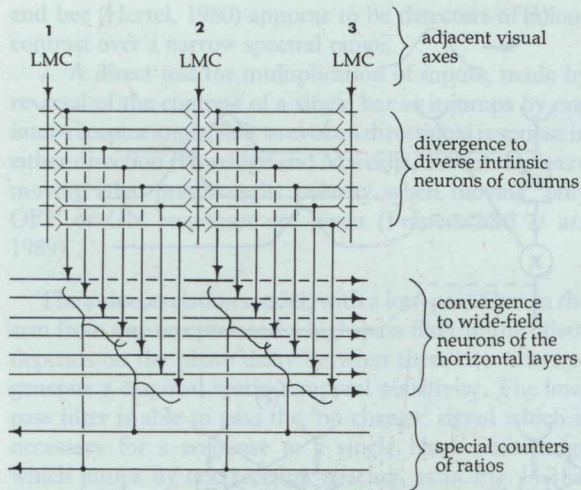


Fig. 11.28 *Ensemble processing by an array in the insect visual system. The inputs are the lamina monopolar cells which carry information about temporal contrast on each visual axis (L1 to L5 in Fig. 11.13). In each column these diverge to a number of intrinsic neurones which presumably abstract different aspects of the temporal contrast in various spatio-temporal combinations (templates). Templates of a given type then converge upon collector neurones, and other neurones are connected to selected groups as special counters of particular ratios.*

running cross the columns (Fig. 11.28). Where the two sets cross there are synapses which could be modified by a conditioning stimulus or by usage. Such a matrix responds with a specific output pattern of line-labelled neurones for each input pattern (reviewed in Rolls, 1987) and has the appropriate properties of tolerance to change in the input pattern, but still has hard-edged outputs. Such systems can complete an incomplete input pattern and can generalize stimulus patterns within limits. The insect optic neuropiles appear to be arrays of this type (Figs. 11.13, 11.14 and 11.28). Similar ideas are proposed by Strausfeld (1989).

The template model has been elaborated at length because it illustrates how a natural visual system could abstract spatial and temporal correlations from the moving image by a distributed mechanism. This is the new framework which allows us to think concretely about vision, both artificial and natural, and its evolution, in new ways.

## Other Models

### The Autocorrelation Model

For many years, data on motion perception in insects have been tested against the autocorrelation theory of Reichardt (1961), partly because the mathematics of systems analysis and filter theory have been conveniently available, partly

because no other likely system has been investigated, and partly because in the early work there was some positive evidence pointing towards multiplication of inputs. The basis of the original model was a series of observations on the optomotor response of the beetle *Chlorophanus* making choices at Y junctions in its path (Hassenstein, 1959).

1. The response was proportional to the square of the contrast in the stimulus.

2. During motion of a regular striped pattern, reversal of contrast caused the perception of direction to be reversed. In the original experiments, this was perhaps no more than a natural effect of the phase change.

3. Pairs of adjacent facets were sufficient for directional motion perception. (Later it was found in the fly that additional facets participate in the processing mechanism.)

4. The response to velocity was a bell-shaped curve, showing that time constants control the upper and lower limits of velocity.

5. Later it was shown that the relative phase of the first and third Fourier components of a regular pattern can be shifted without influence on the steady-state response.

6. It was also shown later that the response depends on the temporal frequency of the passing of stripes, independent of spatial frequency, i.e. independent of pattern. This, however, is a feature of any 'event-driven' mechanism that detects edges if the responses are summed, and is not a test of a particular processing mechanism.

On the above experimental basis, Reichardt (1961) proposed a mathematical model in which the overall response and its direction could be calculated by making an autocorrelation between the modulation in a receptor and the same modulation shifted in time to correspond with a receptor on the adjacent visual axis (Fig. 11.29).

### Deficiencies of the Correlation Model

The general difficulties are:

1. It is a mathematical operation in which the filtered image is multiplied with itself, whereas visual processing is an operation in which many neurone fields respond to a spatio-temporal pattern of input.

2. Motion perception involves much more than optomotor or fixation behaviour, and the correlation model has not proved useful for object vision or artificial systems.

3. This model does not suggest how the visual processing actually operates, or how to interpret the numerous neurones of the optic medulla with ON, OFF and ON-OFF properties.

4. It gives a single quantitative output where we know that even for optomotor behaviour there is a great deal of parallel processing by neurones, and we are interested in the mechanism rather than a calculation of the final summed effect.

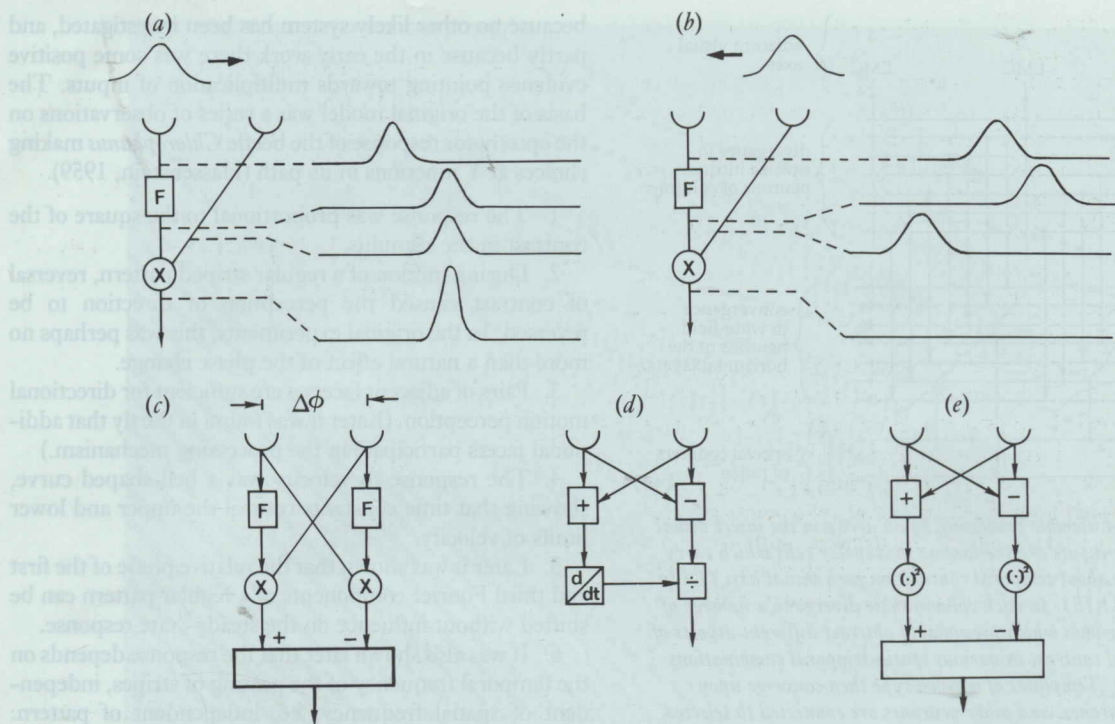


Fig. 11.29 Classical algorithms for motion detection. (a) Motion in the preferred direction over one-half of the Reichardt multiplication model (c); the filter  $F$  causes a phase delay which causes a coincidence with the spatially delayed signal. (b) In the antipreferred (null) direction there is no coincidence, so that multiplication yields zero. (c) The full multiplicative system which eliminates spurious responses to intensity change. (d) The gradient model. Velocity and its direction are obtained by dividing temporal contrast by spatial contrast, irrespective of pattern. (e) The energy model. The difference between the squares of the sum and the difference for two adjacent intensities. The final result is similar to that in (c) for some stimulus patterns.

5. Actual calculations from whole-animal responses yield arbitrary time and filter parameters that could have arisen from properties of the receptors, muscles or mechanical inertia, and are not necessarily related to the processing of motion.

6. In systems analysis, an equal number of gradients of opposite polarity are presented at the same time, generating the second harmonic in the responses.

In particular, there are other difficulties:

1. The finding that summed responses of large-field optomotor neurones or optomotor behaviour depend on the contrast (drift) frequency of the stimulus (e.g. the frequency of passing of edges) is not helpful in distinguishing processing mechanisms, and it makes us wonder how flying insects are independent of spatial frequency in the visual world.

2. The correlation theory predicts that white-black and black-white edges are treated together, but in fact we find that edges of opposite polarity are processed separately and differently (Kien, 1975; Franceschini *et al.*, 1989; Horridge and Marcelja, 1990a).

3. It predicts that (response) is proportional to (contrast)<sup>2</sup>, which is difficult to demonstrate in neurone responses. In fact, we find that the response is more nearly proportional to (contrast)<sup>1</sup> in flies, and that the response saturates at low contrast levels (Horridge and Marcelja, 1990b).

4. Responses to visual stimuli in early processing neurones are essentially phasic and immediate: they are not averaged responses.

5. Responses of visual processing neurones have overlapping response patterns in time and space which are not specific for individual features of the stimulus, suggesting that groups of neurones respond together in spatio-temporal combinations, as in other sensory and motor pathways in all nervous systems.

6. The behavioural responses and the individual-motion-sensitive neurones rapidly adapt: the high-resolution systems in insects appears to be phasic and designed to see direction and non-directional motion, change in velocity, temporal frequency and especially change in contrast. Similarly, in the colour domain, the opponent cells of the optic lobes of locus (Osorio, 1986)



and bee (Hertel, 1980) appear to be detectors of colour contrast over a narrow spectral range.

7. A direct test for multiplication of inputs, made by reversal of the contrast of a single bar as it jumps by one inter-receptor angle, fails to evoke a directional response in either direction (Horridge and Marcelja, 1990a). Unless a moving edge preserves its polarity when moving, only OFF or ON responses are given (Franceschini *et al.*, 1989).

The autocorrelation model, with a low-pass filter in the arm from one receptor and a high-pass filter in the other, depends on the phase delay between these two filters to generate a diagonal spatio-temporal sensitivity. The low-pass filter is able to pass the 'no change' signal which is necessary for a response to a single black-white edge which jumps by one receptor spacing, as in Fig. 11.18a. The spatio-temporal models illustrate two points with reference to directionality: (a) the diagonal spatio-temporal directional templates can be generated in many ways and (b) all that is needed for directionality similar to that given by the autocorrelation model is a 3:1 structure with diagonal spatio-temporal symmetry in any neurone field which can be simplified to a  $2 \times 2$  template.

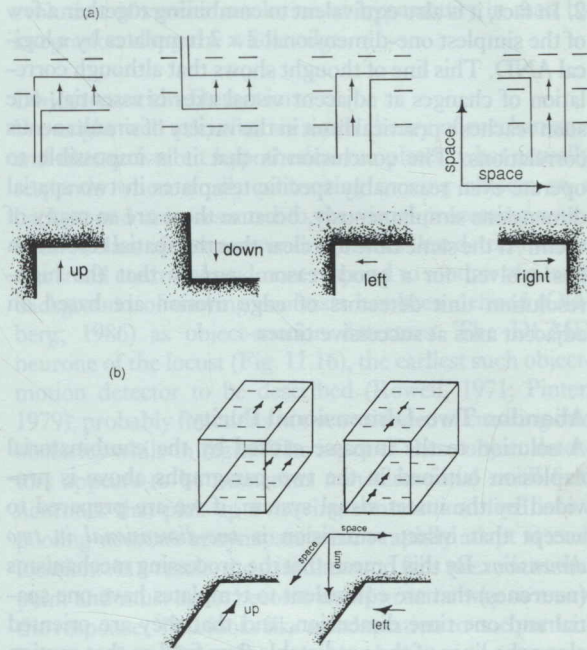


Fig. 11.30 How to construct templates for two dimensions in space: (a) corners that move in predictable ways generate patterns of intensity change; (b) adding the time dimension to (a) gives us three-dimensional templates.

## Two-Dimensional Vision

Let us try to work out what is meant by two-dimensional vision with distributed parallel processing as outlined by the model with templates as crude mimics of local neurones.

### Making Two-dimensional Templates

The basis of the one-dimensional model is that the flow-field is a predictable consequence of locomotion or scanning while the animal moves on an even keel, so that semi-vision in insects is concerned with the detection and discrimination of contrasts moving in predetermined directions. We can extend this to the detection of corners and edges by the simplest two-dimensional templates, which have a  $2 \times 2 \times 2$  structure (Fig. 11.30), but we can immediately anticipate several problems.

1. There are 6561 of these  $2 \times 2 \times 2$  templates, of which only a few may be useful, as determined by running tests with natural scenes.
2. The  $2 \times 2 \times 2$  templates appear to be insufficiently specific to detect two-dimensional features, and at the same time too numerous to collaborate together for discriminations because there would be far too many possible groupings of them.
3. The  $2 \times 2 \times 2$  templates are readily fooled by selec-

ted one-dimensional stimuli, as may be checked by testing with a variety of edge inclinations.

When we implemented this model, as before, but with successive two-dimensional images in spatio-temporal coordinates, we were met by complete failure with natural scenes. There are so many possible templates with combinations of three states at four adjacent axes at two successive times that the responses of almost all of the templates are scattered like noise across the scene. A few  $2 \times 2 \times 2$  templates show up moving edges that happen to be sharp and in line, but most edges turn out to be stepped or notched at the detailed level of individual visual axes, so that there is a chance for a variety of templates to respond here and there without revealing significant template groupings.

The situation is far worse when we try templates of the  $3 \times 3 \times 2$  type, of which there would be 387 420 489 on each visual axis. Many of these templates are reasonably specific in detecting moving corners, but they are so numerous that we cannot map any feature of a moving visual scene with the responses of any of them.

The situation can be simplified by combining templates together so that any one of several gives a response. This process is equivalent to making simpler templates such as  $3 \times 2 \times 2$  type, or reducing the number of states from 3 to

2. In fact, it is also equivalent to combining together a few of the simplest one-dimensional  $2 \times 2$  templates by a logical AND. This line of thought shows that although correlation of changes at adjacent visual axes is essential, one soon reaches a practical limit in the variety of simultaneous correlations. The conclusion is that it is impossible to operate even reasonably specific templates in two spatial dimensions simultaneously, because there are so many of them. At the same time it is clear that the spatial resolution has evolved for a good reason, namely that the high-resolution unit detectors of edge motion are based on adjacent axes at successive times.

### Abandon Two-Dimensional Purity

A solution to the impasse caused by the combinatorial explosion outlined in the two paragraphs above is provided by the insect visual system, if we are prepared to accept that insect semivision is *one-dimensional in two dimensions*. By this I mean that the processing mechanisms (neurones) that are equivalent to templates have one spatial and one time dimension, and that they are oriented along the lines of the predictable flow field so that motion and processing are one-dimensional in each channel at each point on the retina. The line can be vertical up the front and along the top of the eye. There are additional templates aligned vertically at the side of the eye, for optomotor correction of roll in flight; in fact, there can be a superimposed orthogonal or a three-axis array for one-dimensional vision along two or three axes in any part of the eye, but no early templates to detect features in two spatial dimensions. Templates in the vertical plane may abstract vertically moving contrasts as postulated for others in the horizontal plane, but there is no reason to take all combinations of the 81 possible horizontal and 81 possible vertical templates and get back to the 6561 combinations in two dimensions, which individually respond as frequently to noise as they do to real scenes. There is no reason why the 50 or 60 small local neurones on each visual axis in the medulla column should not be sufficient for some vertical one-dimensional vision during head nodding (as in butterfly flight) and also for some horizontal vision in head scanning.

Presumably we can devise tests presenting trained bees with moving targets to see whether all their visual discriminations are accounted for by scanning along predetermined lines by one-dimensional vision dependent on the bee's motions or whether simultaneous correlations in two spatial dimensions have to be accepted. If vision truly in two spatial dimensions simultaneously (whatever that means) occurs in insects, it sets a big problem for evolution and for implementation with a restricted visual system, but so far the difficulty seems to have been in deciding what we mean by tests for vision in two spatial dimensions.

## Semivision Performance

Semivision is the kind of vision that is inferred from the functional analysis of visual neurones and from visual behaviour of lower animals, with no suggestion of a conscious visual world or categorization of objects, but none the less providing excellent vision for mobility and recognition of features of mates, prey or obstacles in flight.

### What Insects See

Freely flying insects appear to see the angle on the eye of moving contrasting edges and manoeuvre relative to them. Somehow, as a result of their own eye motion, they get a measure of the range of contrasting edges at each angle on the eye (Lehrer *et al.*, 1988; Kirchner and Srinivasan, 1989). This gives them a crude representation of the surrounding three-dimensional world just at the time that they most need that information. Their vision of moving intensity gradients, shades of grey and smooth shading has hardly been studied, either at neuronal or at behavioural level. Insect colour vision is related to object detection and fixation rather than to motion. Object and colour discrimination seems to be associated with poorer spatial resolution than motion perception and (so far as we know) is usually associated with groups of visual axes feeding into specialized processing neurones in the part of the eye looking forward, together with visual fixation behaviour, sometimes fixating while scanning.

### Parallax

The word is from a Greek intransitive verb (*παράλλαχθην*) meaning 'to pass by one another' but for centuries has been used in astronomy to mean the angular displacement of an object relative to other stars as a result of the annual movement of the earth. Here, parallax means object motion against background as a result of eye motion.

Besides being able to measure the range of contrasts in different directions as a result of their own motion (Fig. 11.31), and to recognize the differences between contrasting edges, some insects can detect the three-dimensional structure of their surroundings (Fig. 11.32) by seeing the motion of edges against a structured background (Srinivasan *et al.*, 1990). In our spatio-temporal maps (Figs 11.23 and 11.25), parallax is detected as the place where a diagonal line of templates suddenly starts or stops. A general-purpose template to detect parallax is not easy to construct because the foreground and background may be moving in either direction across the eye, and the point of parallax may itself be moving in a different direction, but the sudden termination of a familiar mixture of template responses reveals the closing parallax where a distant contrast goes behind a nearer one, and the sudden novelty of a new mix of templates is a good sign of opening

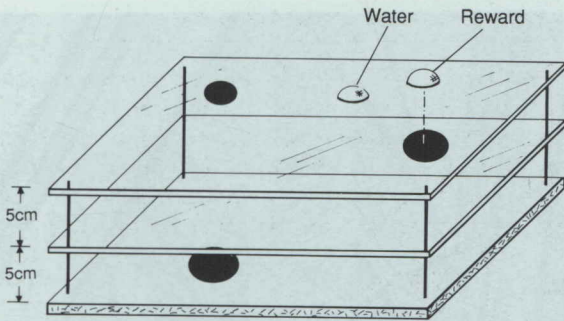


Fig. 11.31 *The apparatus used to demonstrate the ability of a bee to measure the range of objects irrespective of absolute size or position on the eye as they fly along. The bee flies over the three (or more) tiers of clear perspex sheet upon which discs of various sizes can be placed. The bees can readily discriminate and land over the discs of shortest, intermediate, or furthest range irrespective of their randomized sizes and locations. The target for the bee is the drop of sugar solution over the desired disc and other drops are only water to teach the bee not to look for the drop of sugar solution alone. The final discrimination tests are done with no reward drop present. Bees act as if colour blind in the measurement of range.*

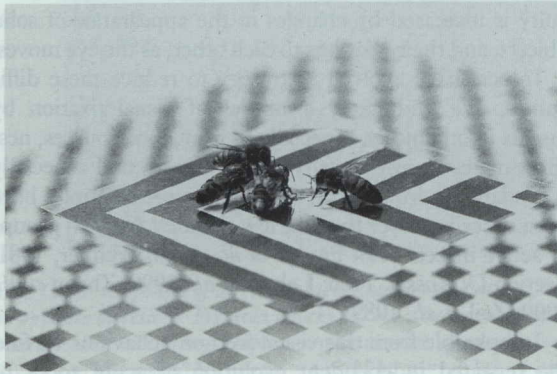


Fig. 11.32 *The apparatus used to demonstrate the ability of a bee to detect the raised patterned platform by the parallax generated against the patterned background as the bee moves. By use of a variety of patterns it can be shown which combinations of foreground and background are most effective for the bees. In these tests the bees come in to land at right angles to the near edge, which therefore generates closing parallax and would cause responses in a specific ensemble of templates (Photo kindness of Dr M. Srinivasan).*

parallax. If the insect assumes that distant large objects are stationary, parallax provides a powerful measure of range during a predictable movement of the eye. Parallax is a reliable signal that distinguishes a solid or separate object from a shadow or flat patch of colour, and it looks very much as if parallax is the driving force that has led to the

evolution of vernier resolution or hyperacuity in man by integration along the whole edge.

### Object-motion Detection

We might well ask, after so much theory of mechanisms and illustrative but hypothetical templates, what actually responds when small, contrasting moving objects on a background are presented to the visual processing neurones. A class of neurones in dragonflies that detect motion of a small contrast but do not respond to motion of a large background of moving contrasts has been defined (Oldberg, 1986) as object-motion detectors. The DCMD neurone of the locust (Fig. 11.16), the earliest such object-motion detector to be described (Rowell, 1971; Pinter, 1979), probably functions in the avoidance of crashing into obstacles while in flight. The input of these object detectors appears to be numerous non-directional small-field neurones that pick up a moving contrast and feed into a pooling neurone at synapses which rapidly adapt at each location. As a result, the stimulus soon fails to excite at that point and must again become novel in some way to renew the response. This looks like an adaptation to one particular mix of templates, but a renewed excitation as soon as the stimulus is shifted or a new mixture appears. Opening parallax continually generates new mixtures. At present we have no better way to interpret the electrophysiology of the detection of moving objects against a textured background.

A great deal of work remains to be done in uncovering the properties of anatomically identified neurones before we can make sense of insect visual processing at the lobula level where object vision may be based if it exists. One of the main difficulties is to identify the real visual fields and significant backgrounds because systems analysis, with stimuli containing black-white and white-black edges *always together* over a range of temporal and spatial frequencies, is not yielding the data we need for inferring neurone functions. If there are medium-level templates, how do we delineate them? So far, the known object-detector neurones do not reveal by their individual response profiles what specific targets they select, possibly because they have to work in groups. However, they suggest that insects do not have a relatively few complicated templates in visual processing, but have instead numbers of (ratios of) less specific template responses in parallel at every level. That makes the task of analysis even harder.

### Fixation, Foveas and Object Recognition

A major problem in visual processing is that the relative motion of the eye and the visual world causes the image to move across the receptor array, and therefore the representation of any particular contrast moves across the central map (Fig. 11.15). To detect this motion of contrasts across the eye it is necessary to look along diagonal

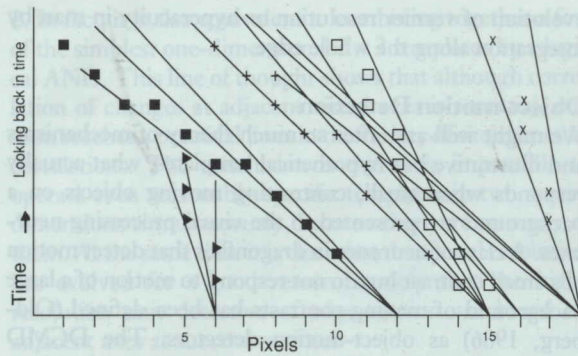


Fig. 11.33 *The Hough transform. Counting the different template responses along the diagonal lines in spatio-temporal coordinates for a moving eye. The slope of the diagonal line of template responses is a measure of the range if the eye scans laterally and does not rotate. So, from any point in time, we look back over the template responses and find the angles at which large counts of responses occur. These angles give the ranges of contrasting features in front of the eye. A vertical line of response represents a feature at infinity or at the fixation point.*

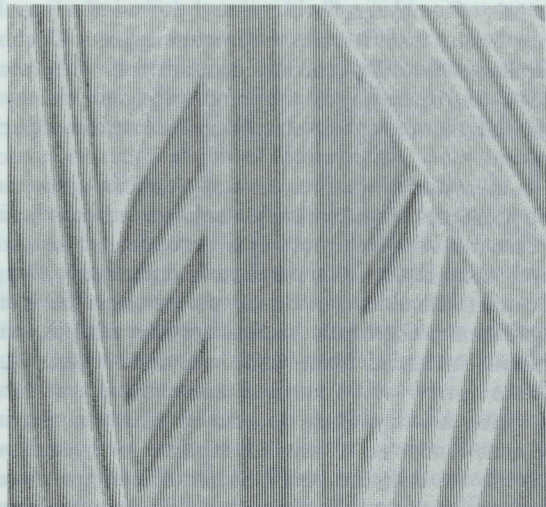


Fig. 11.34 *Fixation. The primary one-dimensional image (plotted along) at successive times (plotted downwards) in spatio-temporal coordinates when the eye scans and also fixates upon one feature in the visual scene. The fixation point is the apparently stationary vertical band down the middle. Objects in front of the fixation point move to the right and down, while those behind the fixation point move to the left and down; the steeper the slope the nearer the objects are to the fixation point.*

lines in spatio-temporal coordinates and to count the template responses to any particular visual feature as they move across the central projection (Fig. 11.33). A serious difficulty arises when these diagonal lines are not straight, i.e. when motions are not constant (Fig. 11.25). The diffi-

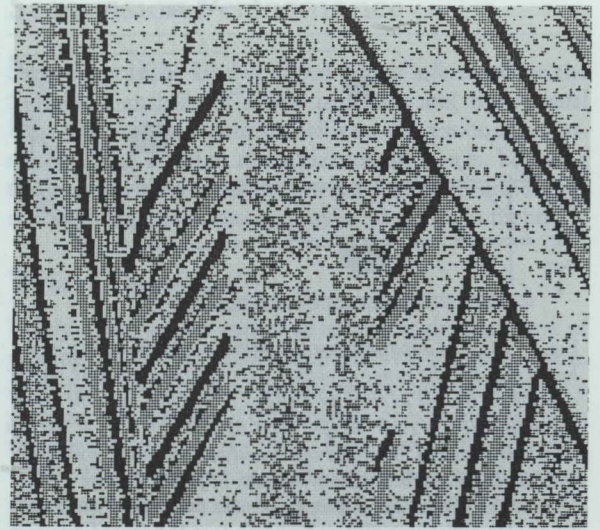


Fig. 11.35 *The spatio-temporal map of thresholded contrasts, pixel by pixel, from Fig. 11.34. (■) decrease in intensity, (▶) increase, (—) no change. We now see that the object which is fixated upon generates a lot of (—) 'no change', although it contains contrasts.*

culty is increased by changes in the appearance of solid objects, and their relations to each other, as the eye moves.

Insects have evolved a strategy to reduce these difficulties. We find many examples of visual fixation by insects upon objects, especially of prey, food sources, nest sites, mates or any novelty which moves, and especially when landing from flight. More and more examples have been described over the past few decades, from all groups of active insects (Van Praagh *et al.*, 1980; Wehner, 1981; Section D; Rossel, 1986; Lehrer *et al.*, 1990; Zhang *et al.*, 1990; Zeil *et al.*, 1989).

Inseparable from this visual fixation behaviour we find the repeated independent evolution of acute zones or foveas which have selective advantage only if there is visual fixation. When held at a predetermined position on the eye with the help of a fixation mechanism, a single contrasting feature can be processed over a reasonable time within a single neuronal field of reasonable size instead of giving fleeting phasic responses as it moves across the internal projection of the retina. However, the fixation behaviour requires some kind of attention-directing circuit so that the fovea is 'locked on' to one particular object and other objects are rejected. Foveas and fixation behaviour certainly evolved independently many times over, so that the advantage must be worth the investment in the control circuits. Acute zones are found in many groups of arthropods (Horridge, 1978a) and always associated with fixation. Predetermined neurones behind the fovea have been described in the male housefly

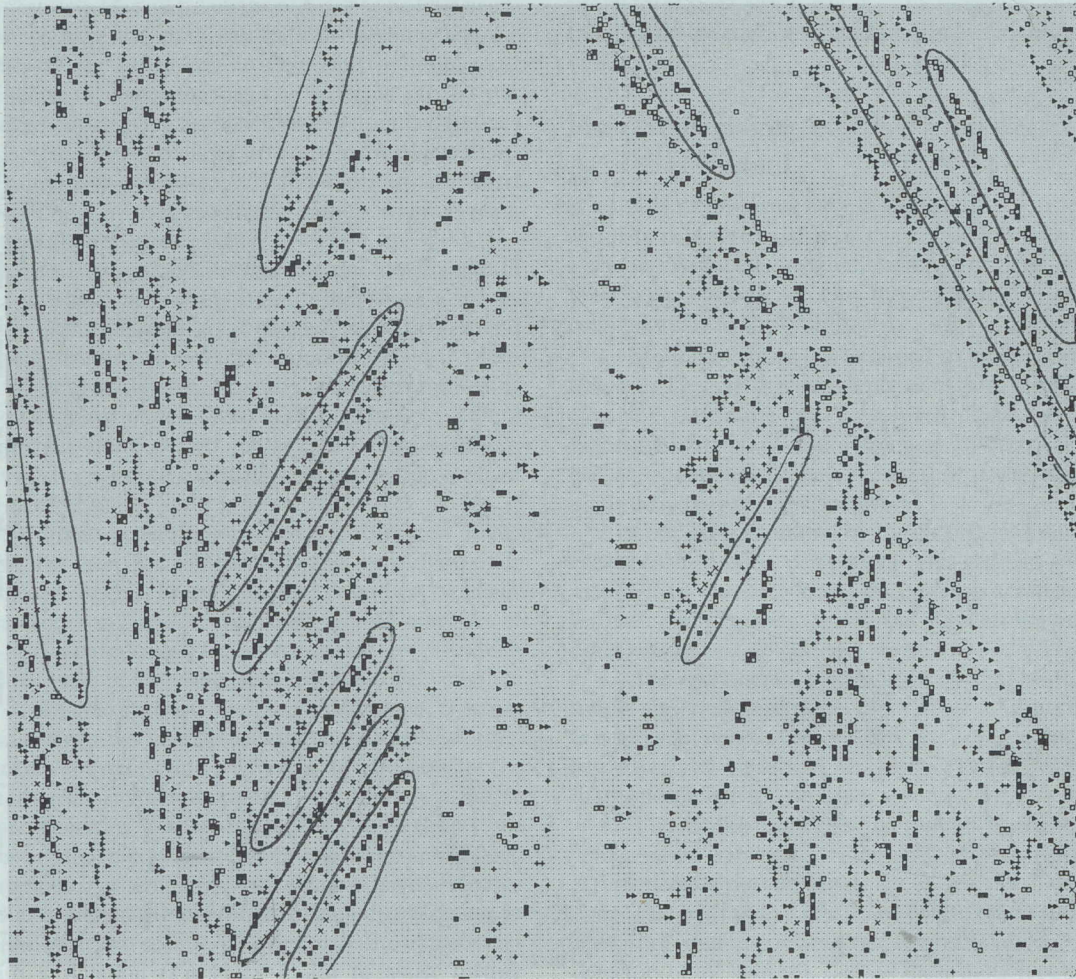


Fig. 11.36 Responses of directional templates looking at the fixated scanned feature. The background and foreground stand out, moving in opposite directions. Fixation knocks out the directional templates from the fixated object. Each moving edge in the scene has its own characteristic mixture of template responses, some of which are outlined. Templates are as follows:  $\blacksquare$  =  $(-\uparrow/\uparrow)$  and  $(\uparrow\uparrow/-\uparrow)$ ;  $\square$  =  $(\uparrow/-\uparrow)$  and  $(\uparrow\uparrow/-\uparrow)$ ;  $\times$  =  $(\downarrow\uparrow/\uparrow\uparrow)$ ;  $+$  =  $(\downarrow\downarrow/\downarrow-)$  and  $(-\downarrow/\downarrow\downarrow)$ ;  $\blacktriangleright$  =  $(\downarrow\downarrow/-\downarrow)$  and  $(\downarrow/-\downarrow\downarrow)$ ;  $>$  =  $(\uparrow\downarrow/\uparrow\uparrow)$ .

and are related to the chasing behaviour by the males (Strausfeld and Nässel, 1981, p 102).

Fixation upon one particular contrast does not necessarily imply that the insect is stationary relative to the contrast. It may be moving forwards or scanning sideways at that time, as if coming in to land, pursuing a target or inspecting a contrasting feature. Fixation usually includes a relative movement of the eye as it wanders about the target. Therefore parallax becomes more important when combined with fixation. On a spatio-temporal mapping, the contrasting feature which is fixated upon generates a vertical band (Figs 11.34–11.37). Considering higher-level neurones as leaky counters of template responses, with the advantage of fixation they can count over longer

times and with greater spatial resolution, i.e. by neurones with smaller fields than would be possible if the image were moving across the retina.

The effect of fixation upon the counts of template responses opens up further interpretation of the function of foveas in the visual process. Moving the eye sideways while fixating upon a contrast in the middle distance causes contrasts in the foreground and background to move in opposite directions (Figs 11.34–11.36) and they move faster the further they are from the point of fixation (see Sandini and Tistarelli, 1990, for recent references).

Fixation also assists colour vision. For obvious reasons few templates for motion respond around the fixation point. Instead, templates carrying two or more adjacent

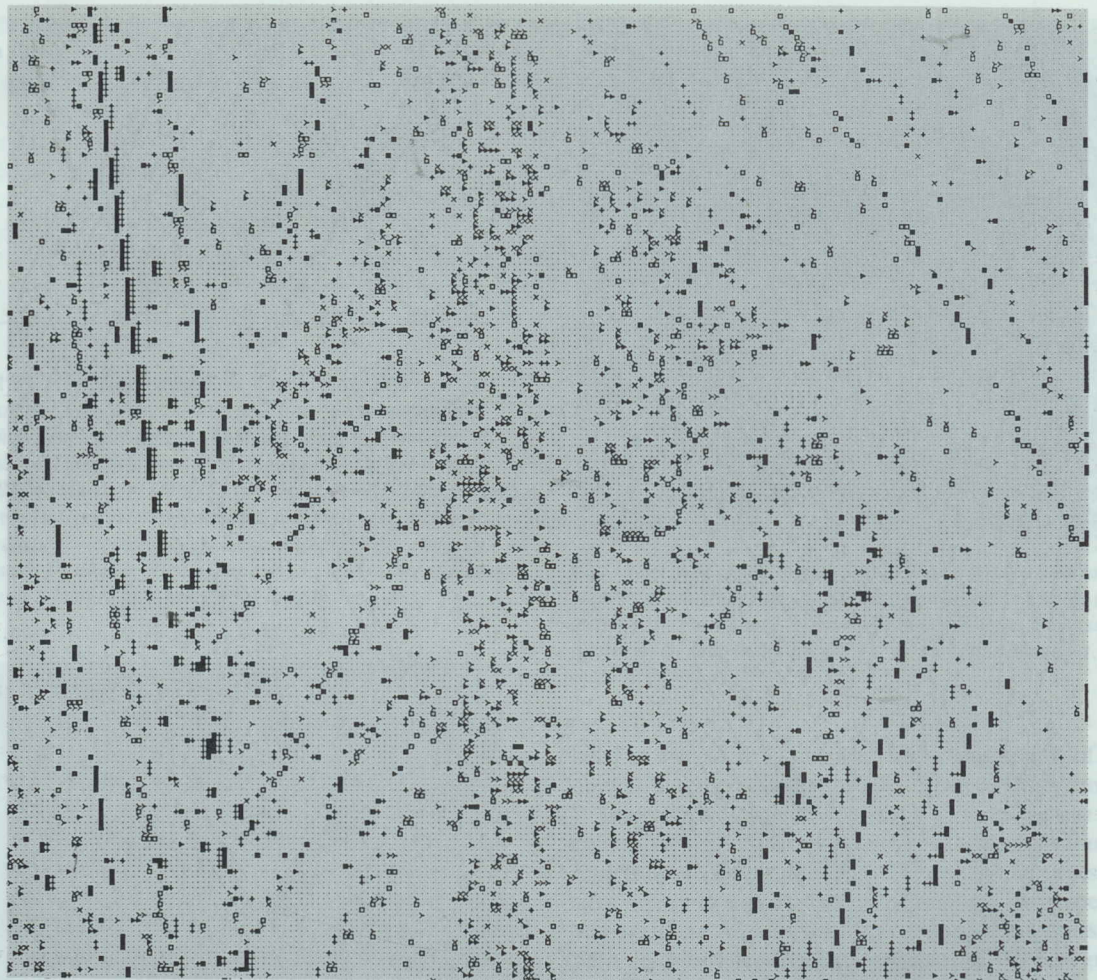


Fig. 11.37 Responses of non-directional templates to a scene which includes the fixated, scanned feature in Fig. 11.34. The jitter in the fixation causes noise which appears as non-directional motion within the fixated object. Combinations of these templates detect certain velocities but are not particularly useful for identifying other features. Symbols:  $\blacksquare = (-\downarrow/-\downarrow)$  and  $(\downarrow-\downarrow/-)$ ;  $\blacktriangleright = (\downarrow\downarrow/\uparrow\uparrow)$ ;  $\times = (\uparrow\uparrow/\downarrow\downarrow)$ ;  $+$   $= (-\uparrow/-\uparrow)$  and  $(\uparrow-\uparrow/-)$ ;  $> = (-\downarrow/\downarrow)$ ;  $\square = (\downarrow\downarrow/-)$ .

'no change' pixels respond frequently in the blank areas of the spatio-temporal map (Fig. 11.35). Regions of the visual field where adjacent pixels are not changing are exactly those where opponent colour mechanisms and spatial interactions for colour discrimination can function undisturbed by motion or flicker. We already know that the high-resolution motion pathway is colour blind (Lehrer *et al.*, 1990); we will probably find that pattern discrimination and colour vision are 'motion blind' and that colour vision has its own channels with templates containing adjacent 'no change' pixels.

### Why Templates Are Proposed

This theory, based on small, high-resolution motion-detection templates in early vision, extends beyond pre-

vious theories. The optic lobes of insects which process the motion information do not function by multiplication of the stimulus intensity with itself: they function by convolution of the pattern as it moves over the eye, with the fields of many neurones located behind every visual axis. There are 50 or more of them in each column of the *Drosophila* medulla (Fischbach and Dittrich, 1989); see Fig. 11.13). There is essentially an instantaneous phasic response to the local spatiotemporal features from the neurones within, as the image of the visual world flows across the central projection.

Insects execute active manoeuvres, especially when flying, which give them information on range, parallax and, in some cases, some kind of rudimentary form of vision and memory of individually distinct landmarks in different directions. The neurophysiologist in search of

the optimum fields and functions of small optic lobe neurones must have a framework of ideas about what the neurones might be doing, so that electrophysiological experiments to determine neurone patterns can be put into action as soon as each neurone is found. Systems analysis, steady-state theories or mathematical operations on the stimulus pattern are not useful in this experimental context. They are not much use, either, when making inferences about neurone fields.

The effort to proceed beyond the anecdotal stage in studies of visual behaviour must be based on a theory which is useful in so far as it suggests experiments and appropriate controls to make the experimental results definitive. We can readily see, by study of old descriptions of visual discrimination tests with bees, for example, a lot of experimental effort wasted for lack of a useful theory. There is also a lot of speculation in vision for want of useful data.

These broad generalizations led me to search at the interface between electrophysiology and visual behaviour for a theory which is based upon large numbers of small-field neurones in parallel. We expect to find numbers of small-field neurones with properties that reveal which templates have in fact been selected from the complete range of possible spatio-temporal combinations. If vision depends on contrast changes at adjacent pairs of receptors at successive instants, and we have a thresholding mechanism, then  $2 \times 2$  spatio-temporal templates are a simple way of generating responses that can be counted like photons. Actual neurones are, of course, more flexible and overlapping than templates, but we now have a scheme which exemplifies the parallel distributed processing in operation with real scenes. The template theory provides us with a notation (Fig 11.23–11.26) to assist further thinking about visual processing. The template theory also turns out to be a useful way to think about the evolution of visual processing, and it generates new ideas about colour vision, discrimination of features independently of their number, the effect of fixation and the significance of one-dimensional vision with predictable motion of the eye.

The old lessons from biological processing of information are that neural mechanisms are sloppy and inaccurate, that they respond rapidly to transients by preformed expectant sensors and few successively higher level neurones, and that they avoid extensive sequential computation. The new lessons are that several very simple, specialized pathways can work in unison to achieve specificity and discrimination, that parallel distributed systems cannot be analysed as if they are single channels and that prior knowledge of the task is built into every aspect of structure and function by progressive evolution of earlier processing mechanisms.

The idea of templates helps us to visualize the mechanisms and is a useful theory for many approaches to the

problem of how vision works. The templates show how the spatial organization of contrast can be fed into an array of neurones.

## How Semivision Avoids Problems in Artificial Vision

A number of problems have been identified by those who simulate computer vision or try to build artificial seeing systems. With the insights gained from the template theory, we can now see how insect semivision has overcome or avoided these difficulties.

1. Motion of the eye is assumed to be horizontal and forward. This yields a one-dimensional analysis of motion at each point on the eye. To control roll and pitching in flight there are additional systems in the vertical planes, but no evidence that these are used for two-dimensional picture analysis of the visual world.

2. Contrast is saturated by means of a high gain in the channels for motion perception. This reduces the information load with little loss of the image structure in return for seeing any moving contrast at high spatial and temporal resolution.

3. The visual world is composed of edges so that  $2 \times 2$  spatio-temporal templates are useful in groups. This generates a useful compromise between ambiguity of responses and variety of templates. Templates that are more specific are too numerous.

4. The template system for motion is colour blind: colour vision is in separate channels. Maybe colour vision is directionally motion blind.

5. The high spatial resolution of the retina and lamina continues into directional and non-directional motion detectors of the medulla. Each of these motion pathways is then separately summed into large-field neurones which control vital locomotory reflexes. There is no evidence of other kinds of vision, e.g., of objects, at the highest resolution.

6. In particular, there is no evidence that the position of edges are *located* with any kind of mechanism downline that can give the coordinates of a 'zero crossing', so such operations are irrelevant.

7. To obtain a simple three-dimensional map of the world, the range in each direction is measured from the relative motion on the eye caused by the predictable flow field at each point. There is no need to analyse the whole flow field.

8. The three-dimensional map of the world is augmented by the parallax caused by motion of nearby edges over a patterned background. Parallax is detected where different motions of edges meet. Parallax detection is incompatible with the 'smoothness constraint' that is sometimes assumed in the analysis of flow fields in artificial visual systems.

9. The variety of outputs is limited and visual behaviour relies on a few specific visual cues for each task. This helps reduce the enormous number of template combinations that would be essential for vision in two dimensions.

10. Picture analysis and categorization of objects are replaced by a reliance on trigger features or predictable cues.

11. If possible, during a difficult discrimination task, the image is held stationary in the central projection by fixation while scanning (Figs 11.34–11.37).

12. The use of one-dimensional processing along predictable flow lines means that the classical problems of two-dimensional vision do not arise.

13. The 'aperture problem' arises because the true motion of a moving straight edge cannot be determined when it is seen through a single-bounded window such as a neurone field. The 'smoothness constraints' assume that edges of objects are continuous, that velocity is constant over small areas of the image, and that objects are not elastic. These assumptions are not relevant to one-dimensional vision.

14. Contrary to the smoothness constraint, discontinuities in the flowfield are useful indicators of parallax, and freely flying insects interacting in flight cannot assume that the surrounding world is rigid.

15. The enormous number of combinations of inputs that makes picture analysis computationally heavy and slow is reduced by one-dimensional vision, built-in templates, limitation of input and output tasks, ensemble processing and the use of large fields for high-level channels.

16. Learning is a separate process which changes the weighting in ensemble processing, i.e. changes the preferred *ratios* of template responses, which applies equally to vision of features and colours.

## Larger Templates

### Recognition of Significant Images by Semivision

An obvious feature of insect vision is the ability to recognize and chase a mate or prey, fly away and return to the same twig or leaf, repeatedly visit flowers of one kind, or use landmarks to return to a nest hole. How are these features recognized?

So far I have outlined a theory of elementary colour-blind templates which abstract direction of motion, polarity of contrasts and non-directional motion at the maximum level of spatial resolution that the design of the visual axes allows (Fig. 11.18). Those templates drive the retina towards the evolution of better resolution. The same templates can provide ratios to give measures of contrast qua-

lities, just as is done by receptor inputs in colour vision (Fig. 11.27). After that, the ratios, now independent of pattern, can be put together by logical AND to trigger deeper functions for specific tasks to avoid the combinatorial explosion. Is there more?

Certainly, at the whole-eye level there is more, as outlined in examples below; but we should carefully consider a group of mechanisms of intermediate scale that detect biologically significant images which are commonly called 'trigger features' or 'sign stimuli' following early popularization of these terms with work on the visual behaviour of fishes, spiders and birds (Tinbergen, 1951). To be as critical as possible, I would like to see discriminations that cannot be accounted for by ratios or logical AND of responses of templates such as those in Fig. 11.18 and 11.20. Let us examine a few examples before trying to analyse the possible processing mechanisms.

### Recognition of Mate, Prey or Predator

Examples abound; one is the preference of stick insects which, when they fall on the ground, walk towards any object that looks like a bush (Fig. 11.38). Another is the threatening posture which is displayed by praying mantis of the genus *Stagmatoptera* when encountering an insectivorous bird. The bird is recognized visually. The behaviour is restricted to this genus of mantis which have large imitation eyes on the prothoracic femur, and the response is fairly specific to certain species of birds (Crane, 1952), but is apparently not learnt. The mantis rears up, exposes

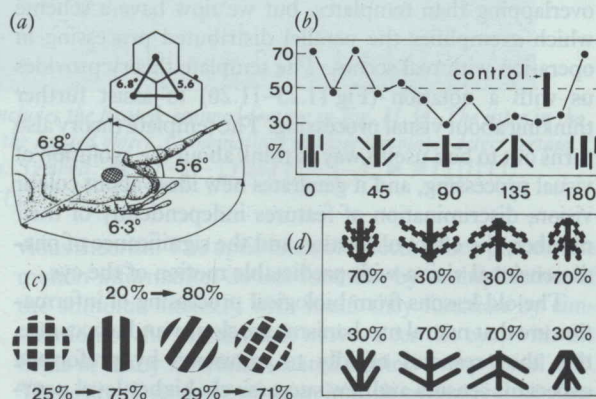


Fig. 11.38 The relative attractiveness of different images to a hungry stick insect when given a choice. (a) The large interommatidial angles of the eye. (b) The angle at which the side branches project has a large effect on the visual preference. (c) Preference is always to the right-hand member of the pair. (d) Roughening the branches makes them more attractive when they are pointing upwards. Inbuilt visual behaviour of this kind suggests that predetermined groups of appropriate templates occur in the optic lobes (Modified from Horridge, 1978, after Jander and Volk-Heinrichs, 1970).



the mock eyes, elevates its wings, stridulates violently and sways from side to side, often frightening away the bird. With one compound eye covered over, the mantis can still recognize a bird, or a film of one, but cannot estimate its range (Maldonado, 1970).

The significance of visual shape in arthropod behaviour patterns is reviewed at length, with numerous examples, by Wehner (1981). The main categories of images are the shapes of flowers, the recognition of mates and prey, courtship behaviour (especially in hunting spiders and fiddler crabs) and the pattern on the wings of butterflies, dragonflies and others. The basic problem is that we are unable to obtain much information about mechanisms of visual processing in these cases; therefore we turn either to the analysis of object-detector neurones (if they can be found) or to the visual behaviour of the trained bee.

### Eidetic Images

The word is from the Greek root  $\text{E}\text{I}\Delta\Omega$  (latin, *video*) or  $\Sigma\text{i}\delta\omega\lambda\omicron\upsilon$ , meaning an image in the mind, which has come into English as 'idol'. The word 'eidetic' means an imprint of the image in the visual processing mechanism, but authors are usually not clear whether the eidetic image can float about in the spatial array of the visual projection or whether it is burnt in at one location, requiring congruence at the same location before recognition can occur.

As mentioned above, experimental analysis is almost restricted to bees. The honeybee can be trained to come to a food source that is recognized visually, and then, being trained for one target pattern, can be tested with the same or other patterns to reveal something about what has been learned. Three warnings are essential when considering this topic. Firstly, the bee must be given a frame of reference in the form of her own motion, including horizontal scanning, and the direction of gravity; she is then able to discriminate many shapes, angles of inclination and locations relative to patterns or landmarks. Most experiments before Wehner's work, starting about 1970, were done with patterns laid flat, probably copying Von Frisch's early experiments with colours. The bees, having no reference axis while flying over the flat patterns, showed that they could discriminate *degrees of disruption* in the patterns, but little more. Only experiments on vertical surfaces are useful. Secondly, the bee must be forced to make her decision at some distance from the target, and target positions must be randomized regularly during training. Bees that are allowed to examine targets closely before making a choice can concentrate on regions of the target, e.g. the top corner, so that it is then impossible to know what part of the target the bee is looking at. Thirdly, all aspects of the pattern except the detail that is the topic of the experiment must be randomized repeatedly during the experiment to teach the bee what features *not* to look at and learn. This experimental design of controls is so frequently ignored

that it is impossible to be sure whether the bees were looking out for some totally irrelevant detail. Many other controls are sometimes essential, such as restricting the approach path of the bee, or photographing the bees in flight, because the cue that the bee uses to locate the reward may be one that the experimenter does not anticipate. By randomizing all other features, Lehrer *et al.* (1988) showed that bees can measure range to objects independently of size, and can discriminate a selected range (Fig. 11.31) to get a reward; Van Hateren *et al.* (1990) and Srinivasan (unpublished data) showed that bees can discriminate a difference of  $45^\circ$  in the slope of parallel lines in a pattern of random parallel stripes (Fig. 11.39), but they cannot discriminate one random pattern from a similar one at the same slope; Srinivasan *et al.* (1990) showed that bees can use parallax (Fig. 11.32) to get an idea of the separateness of an object from background. These experiments can all be interpreted readily by templates responding to horizontal and vertical motion with scanning (Fig. 11.20). Apparently, pattern disruption interferes with pattern discrimination by bees, as would be expected from a theory based on ratios of template responses. To make inferences about more complex tem-

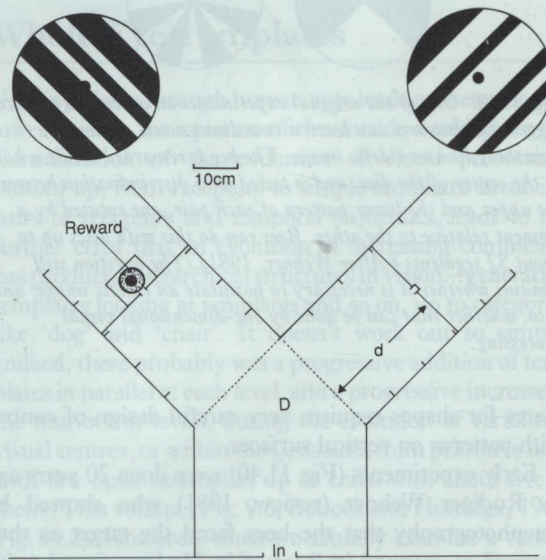


Fig. 11.39 The apparatus used to test visual discrimination of forms and angles. The bee enters through the hole, and in the middle of the Y must make a choice between the two targets which can be placed at various distances ( $d$ ) in the arms of the Y. The reward is found in the hole in the centre of the pattern. The two examples shown were from a selection of randomly striped patterns. It was found that from a distance the bee can discriminate a difference of  $45^\circ$  angle irrespective of pattern but bees cannot discriminate these two patterns if they are inclined at the same angle (Redrawn from Van Hateren *et al.*, 1990; Srinivasan, unpublished data).

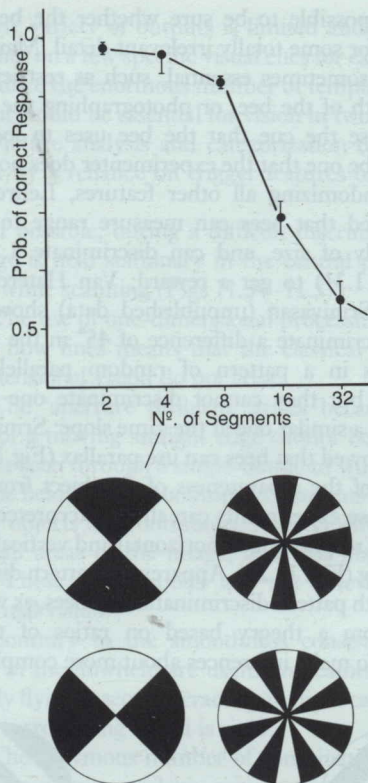


Fig. 11.40 One of the original experiments on a vertical surface suggesting that bees can learn a two-dimensional pattern by something like an eidetic image. The bee is rewarded inside a hole at the centre of the disc, and is tested on a discrimination between the upper and the lower pattern of each pair, one rotated by a segment relative to the other. Bees can do this with discs up to about 32 segments (After Wehner, 1981). The question still remains whether it is necessary to postulate an eidetic image, and is so, whether that can be done by one-dimensional visual processing.

plates for shapes requires very careful design of control with patterns on vertical surfaces.

Early experiments (Fig. 11.40) were done 20 years ago by Rudiger Wehner (review, 1981) who showed by cinemography that the bees faced the target as they hovered or scanned in front of it. He later found (with Flatt) that they could not recognize the target if the eye region that learnt it was painted over, and suggested that something is fixed in the map behind the eye. For this to be so it is evident that the image cannot be at the full resolution of the eye—it must be fuzzified by neurones with large fields so that the image can be recognized *within a region*, albeit less precisely. Taking ratios of template responses would be one way to enlarge those spatio-temporal fields.

Over the past decade the standard way to think of pattern recognition in insects, by a combination of old and

new results, was to imagine accurate measurement of flicker sequences as images swept across the retina and also as a less precise recognition of a pattern that is briefly lined up with an internal representation like an eidetic image (e.g. Collett and Cartright, 1983; Gould, 1985). The fovea of the insect eye was seen as the most effective region for picking up the eidetic images and responding to their recurrence, so that fixation behaviour was tied into object discrimination.

### Template Theory Applied to Object Vision

First, setting aside colours, let us refer to the colour-blind high-resolution templates in Fig. 11.20, noting that most respond to *polarity* and directional or non-directional *motion* at adjacent visual axes. As discussed already, making larger templates *at this resolution* is futile on account of the combinatorial explosion, but taking ratios of template counts, followed by logical AND in local spatio-temporal regions, looks promising, although two-dimensional shape is thrown away in the process. We add the possibility that there is also a crude, one-dimensional semivision of this type in the vertical plane, even if it is less effective.

Within a brief period in any local region of the eye, as the bee scans horizontally with one-dimensional horizontal and vertical directional detectors, the ratio of template responses for upwards, downwards, left-and-right motion (Fig. 11.27(d)) could be a measure of the angle of inclination of the edge, independent of form, and it could also distinguish edge polarity. If the direction of the bee's own scanning motion is monitored by large-field directional neurones, then a stationary angle could be discriminated by a freely flying bee.

Bees can in fact discriminate the angle of inclination of randomly arranged stripes (Fig. 11.39) when the form of the pattern presented is randomized in the tests so that the bee is taught to look only at the angle. In these tests, however, bees cannot discriminate different examples of random stripes except by trivial properties such as average brightness or contour density. When the tests are controlled in this way, the bee cannot make an eidetic image *from a distance* (Fig. 11.39).

The experiments that most strongly support the idea of an eidetic image are those of Wehner (1981, p 477) done in 1972. Bees can detect the change from white to black at the top of a disc with 16 alternating black and white segments (Fig. 11.40). At first sight it is hard to explain the bee's ability to discriminate a small angular shift of the radial pattern, because the sum of all the angles is independent of the angular displacement. In these experiments, however, the bee comes close to the hole in the centre of the pattern, which therefore occupies a large part of the eye. The experiment seems designed to tell the bee that she should look to see whether the horizontal midline of the two eyes

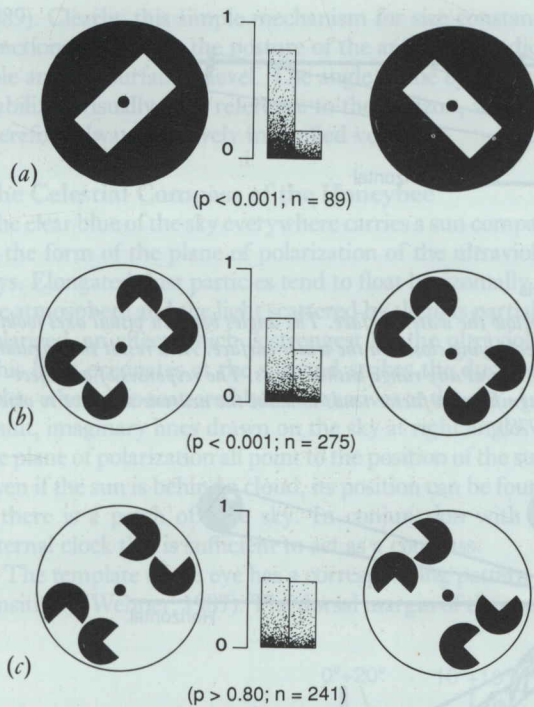


Fig. 11.41 Bees can easily discriminate between the two figures in (a) and fairly easily between those in (b), but cannot distinguish the two patterns in (c). It is an interesting question whether their failure to discriminate in (c) is caused by the disruption in the patterns or by the loss of the visual illusion (Redrawn from Van Hateren *et al.*, 1990).

is in line with black or white. Another puzzling discrimination test, done recently in Canberra, is illustrated in Fig. 11.41. Bees easily discriminate between the patterns in (a). The patterns in (b) are more difficult, but are still distinguished correctly more often than not. The same patterns as (b), but rearranged as in (c), so as to eliminate the visual illusion, cannot be discriminated by bees. Whether the failure to distinguish in (c) is simply because the pattern is more disrupted than (b), as seen in horizontal scans, or whether the bees see the illusion as we do, cannot be decided. Certainly it is hard to design crucial tests for eidetic images which have only one interpretation.

All of the patterns recently employed for tests for pattern vision on bees, except the randomized use of selections of random patterns (Van Hateren *et al.*) can be criticized in the light of the discussion of proper controls. It is certainly possible to devise template ratios that discriminate the patterns and provide alternatives to eidetic images in the supposedly definitive studies (e.g. Gould, 1985, 1986). However, there are other significant natural situations, akin to patterns, that may depend solely on spatial (possibly one-dimensional) correlations over large

visual fields: for example landmarks, and whole-eye or two-eye templates.

### Landmarks

Insects apparently use visual landmarks so that without retracing their steps they can return to a known site that they have previously learned, often having made exploratory flights. The extensive literature on navigation by landmarks is reviewed by Wehner (1981). Experiments with movable landmarks show that bees learn what the distribution of landmarks should look like, as seen from the point where they desire to be. They search around until the bearings of the landmarks each lie on the retina where they have been accustomed to seeing them and at their normal apparent size subtended at the eye. However, because bees look forward but fly in any direction we are faced with the question of how the eidetic images of the landmarks can rotate within the bee. The best solution offered is that the bee has several internal snapshots which can be used when looking in the appropriate directions (Collett and Cartright, 1983). If this is so, we have no idea where they are located (Strausfeld, 1989).

### Whole-Eye Templates

Viewing vision through human eyes leads to the error that lower animals see a picture of the outside world as we do. Consideration of the evolution of visual processing, bottom up, from receptors to simple templates involving pairs of receptors and temporal sequences, leads to the further error that an evolution of increasing complexity has produced hierarchical structures in vision – generating templates looking at templates and so on, up to categories like ‘dog’ and ‘chair’. It doesn’t work out so simply. Indeed, there probably was a progressive addition of templates in parallel at each level, and a progressive increase in the number of levels, during the evolution of vertebrate visual centres, or within the Crustacea from primitive ones with few optic neuropiles up to crabs with about five of them (This volume, Fig. 9.6; Bullock and Horridge, 1965, Fig. 19.2). However, when we actually examine a variety of examples of complex visual discriminations done by lower animals, it turns out that the sampling array itself is structured around a restricted set of tasks, or dedicated to a single task, so that eye structure and at least one channel of processing act together like a single template (Wehner, 1987). The following are three examples.

#### Size Constancy in an Object-Motion Neurone

The pond backswimmer *Notonecta* hangs in a predetermined position below the water surface, waiting for prey such as damaged insects that struggle as they float within

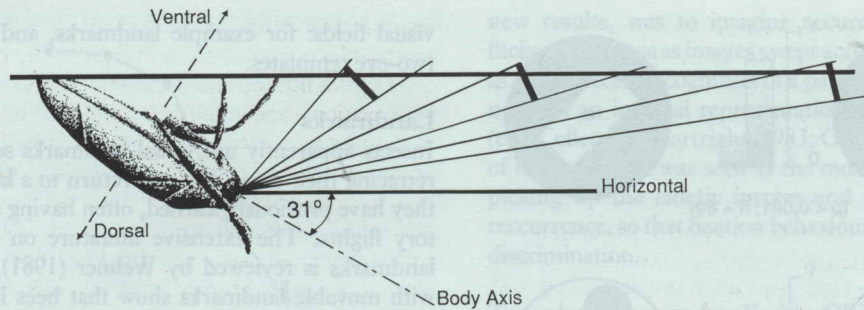


Fig. 11.42 *The backswimmer Notonecta hangs in a predictable posture below the water surface. The angles between visual axes looking horizontally to the distant water surface are less than those looking at objects lying closer on the water surface. As a result the angular subtense, measured in number of visual axes, is similar for an object of one size at any range within limits. The responses of an object-detector neurone of the optic lobe show that this gradient in the spatial magnification factor tends to make the neurone sensitive to objects of a given absolute size irrespective of range (Redrawn after Schwind, 1978).*

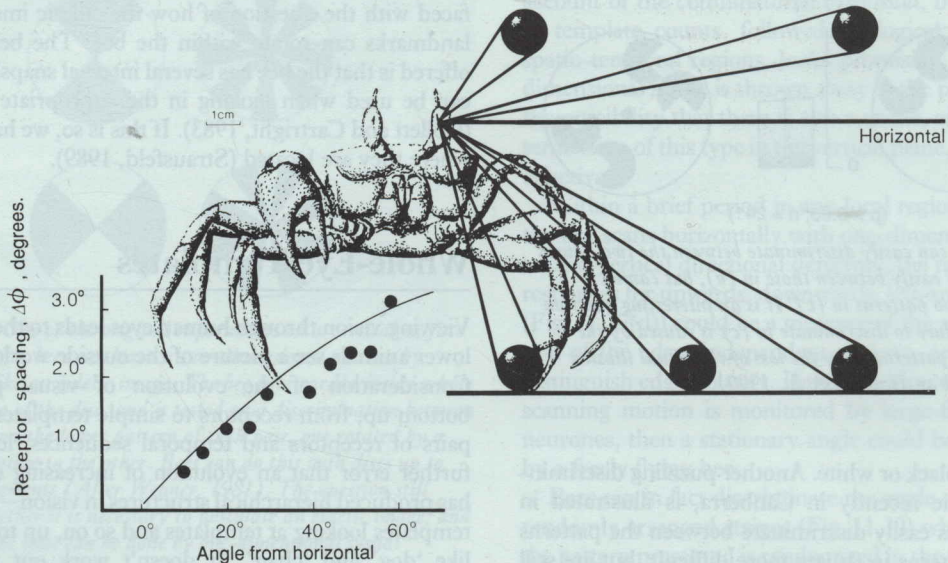


Fig. 11.43 *The ghost crab Ocypode has a vertically elongated eye with small angles between the visual axes looking at the horizon, grading into larger angles between axes looking up or down (inset). This built-in gradient of magnification in sampling could have the effect that an object approaching along the ground cuts the same number of visual axes irrespective of range. Similarly, an object above the horizontal is sampled by a constant number of visual axes irrespective of its distance, over a limited range, but the absolute size cannot be measured because the altitude is not known (Redrawn from Wehner, 1987).*

range. A large neurone in the optic lobe responds to motion of any small, black object within a field which includes the forward-looking part of both eyes. The neurone's response is reduced if only one eye sees the object. Motion of a large patterned background elicits no response and reduces that to a superimposed small object. The interesting feature is that objects of the size that produces the maximum response do so irrespective of their distance along the water surface. The preferred object size of the eye region looking at the horizon is smaller than that looking at a point near to the animal (Fig. 11.42). In the figure, the angular subtense of an object decreases as the object

recedes, but the neurone (and perhaps the retina) compensates in such a way that the preferred absolute dimension of an object (at the water surface) is constant. In fact, the most preferred object is one-fifth the size of the *Notonecta*, a convenient size for a prey (Schwind, 1978). The same principle has been extended to ghost crabs (*Ocypode*) and other flat-world crabs, and related in a specific way to the eye anatomy (Fig. 11.43). The visual axes pointing to the horizon are closer together than those looking down to the ground, with intervals between axes graded in such a way that an object subtends a similar number of visual axes whatever its range, over limited distances (Zeil *et al.*,

1989). Clearly, this simple mechanism for size constancy functions only when the posture of the animal is predictable and the surface is level. The angle of the eyestalks is stabilized visually with reference to the horizon, which is therefore always relatively magnified vertically.

**The Celestial Compass of the Honeybee**

The clear blue of the sky everywhere carries a sun compass in the form of the plane of polarization of the ultraviolet rays. Elongated dust particles tend to float horizontally in the atmosphere and the light scattered by them is partially polarized, an effect which is strongest for the ultraviolet. This light originates at the sun and strikes the dust particles, where it is scattered before it arrives at the eye. As a result, imaginary lines drawn on the sky at right angles to the plane of polarization all point to the position of the sun. Even if the sun is behind a cloud, its position can be found if there is a patch of blue sky. In conjunction with an internal clock this is sufficient to act as a compass.

The template of the eye has a corresponding pattern of sensitivity (Wehner, 1987). The dorsal margin of the com-

pound eye of the bee (and of many other insects) has ommatidia unlike those in the rest of the eye. They have poor optics but the photoreceptor cells are sensitive to the polarization plane in the ultraviolet. The axis of maximum sensitivity points towards the dorsal pole of the eye, in the same way (radially) as the lines at right angles to the planes of polarization in the blue of the sky point to the sun. Therefore, as the bee turns round, each blue part of the sky will appear maximally bright at two positions 180° apart. The pole of the eye then gives the direction of the sun (or the direction away from the sun). The ambiguity of the sun's position is overcome by use of the general brightness of the sky and by the integrative action of the bee's eye as a whole, because the sun is not expected to be below the horizon.

**Fly-Grabbing by a Hungry Mantis**

The praying mantis has a foreleg modified to flick out and catch a fly in its tarsal-tibial joint. To be effective the mantis must have the fly at the range that suits the length of its leg. Normally the mantis lies in wait and looks

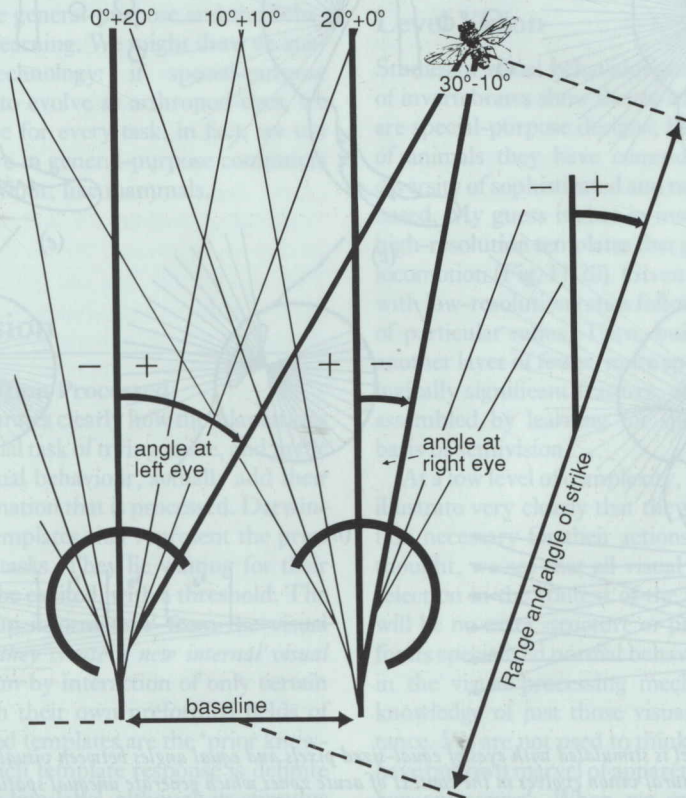


Fig. 11.44 The praying mantis eyes are able to control the strike at a fly. The angle for the strike by the leg depends on the mean of the angles subtended from the midlines of the two eyes (with signs as shown). The difference between these angles gives the range and is approximately constant for a given range, as shown. An interesting corollary is that as the mantis grows and its leg lengthens, the visual angles and processing must also change to maintain the accuracy. (Redrawn after Rossel, 1986).

towards an approaching fly. The visual fields of mantid eyes have a binocular overlap that is abnormally large for an insect – about  $45^\circ$ , with a wide, shallow, forward-looking fovea (see Fig. 11.1). In a series of ingenious experiments with prisms placed in front of the mantis's eyes, Rossel (1986) has found that the mantis strikes at the apparent range of the fly as seen through the prisms by both eyes. The visual axes radiating out from the two fixed eyes therefore form a fixed lattice (Fig. 11.44) which is able to locate the fly and indicate its range. The mechanism still functions quite well when the fly is up to about  $20^\circ$  from the mid-line. An approximate range can be calculated by adding together the angles subtended by the fly at the two eyes and taking the reciprocal, but knowing the way insect nervous systems function, it seems likely that there are

internal templates which are set for just those combinations of angles that give the correct range, as suggested in Fig. 11.44.

#### Conclusions from Whole-Eye Templates

Although directional motion detection has the full spatial resolution of the retina, there is no evidence of invertebrate semivision mechanisms that locate the position of a stationary edge with the same accuracy. Although the minimum displacement of a moving edge is less than the interommatidial angle in crabs and insects, the only indications that the location is measured relative to the eye are the very ones where a fixed eye geometry measures the direction for tasks such as grabbing prey (Fig. 12.45g,h).

A new principle shows up clearly in these whole-eye

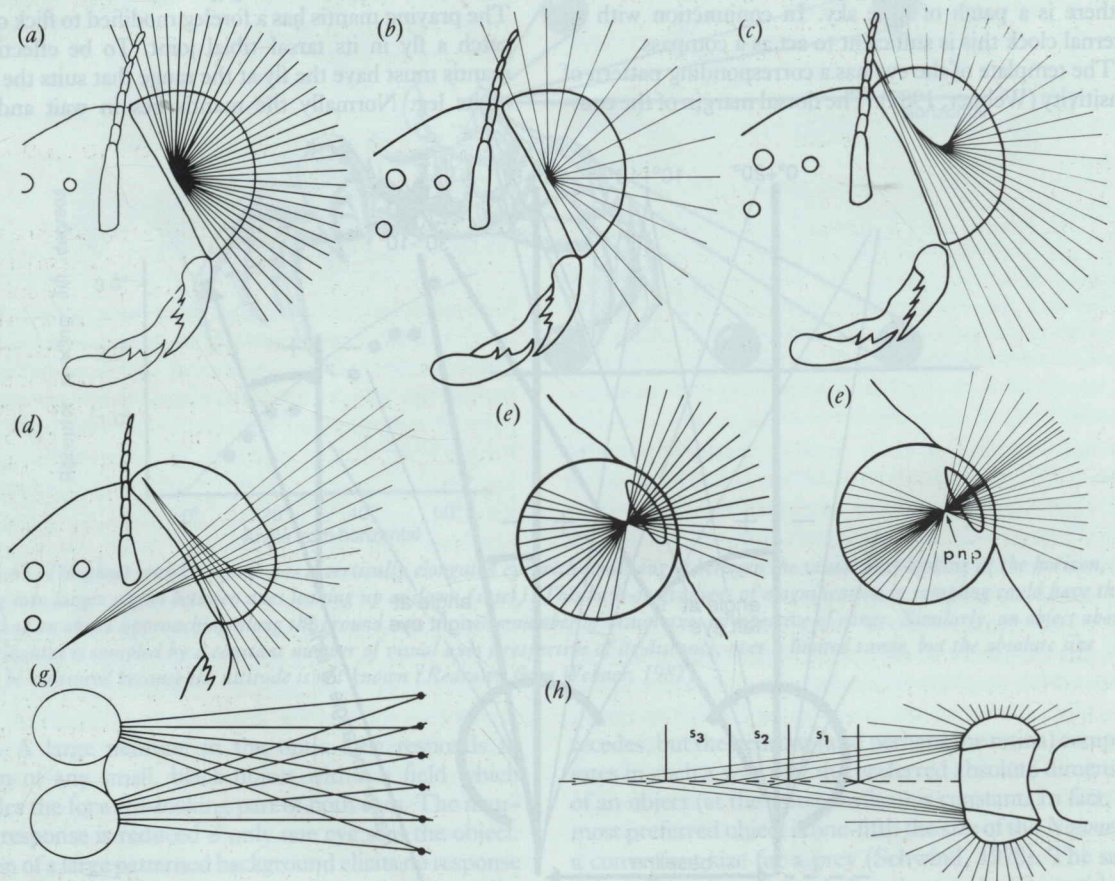


Fig. 11.45 The template model is stimulated with eyes of equal-sized pixels and equal angles between visual axes. However, as illustrated in this diagram, natural vision evolves in the context of acute zones which generate unequal spatial magnification in the central projections, clearly improving the visual processing of the fixated object. (a) A diurnal eye with small facets and dense sampling array. (b) A nocturnal eye with larger facets and fewer visual axes. (c) A forward-looking fovea with increased eye radius. (d) Two foveas, as in some dragonflies. (e) A camera-type eye without an acute zone. (f) A camera-type eye with a fovea of increased sampling density and therefore narrower receptors. (g) Mantis type of binocular overlap for control of the strike of the foreleg off the midline. (h) Binocular overlap along the midline, as in dragonfly larvae, for control of the strike by the mouthparts.

templates. There is a contribution from each part of the eye acting as a sampling array. When the whole of the bee's polarization mechanism sees the sky, it all functions together, but if only a small patch is visible then that is sufficient to give the sun's direction. There are many examples in insect visual behaviour, especially in chasing or turning to look at an object, or when fixating upon an object, when the strength of the response depends on the sum of a number of inputs subtended at different points on the eye.

These whole-eye templates, *which allow stationary animals to measure range*, clearly depend on hard-wired circuitry that is already fully functional when the imago insect emerges from its pupa or larva, and yet we find this exact correspondence to the structure of the outside world and the visual task. A job for the future is to combine the template model with the acute zone, and implement artificial visual systems with gradients of angle between the visual axes as illustrated in Fig. 11.45. In terms of evolutionary adaptations, the insect or crab displays its visual behaviour in its sampling array. Such adaptations are more characteristic of arthropods than (say) mammals, where the eye tends to be general-purpose and the behaviour to be perfected by learning. We might draw an analogy with modern technology: if special-purpose computers were as easy to evolve as arthropod eyes, we would make a special one for every task: in fact, we use rapidly adaptable software in general-purpose computers which have flexible behaviour, like mammals.

## Beyond Semivision

### Adding to the Information Processed

The template model illustrates clearly how the adaptations of the neurones to the visual task of trying to see, and to the action of the ensuing visual behaviour, actually add their contribution to the information that is processed. Darwinian selection promotes templates that represent the prior knowledge of the visual tasks. They lie waiting for their spatio-temporal fields to be excited, with a threshold. The templates do not 'pick up information' from the visual world. In their clusters *they create a new internal visual world* in their own domain by interaction of only certain aspects of the input with their own preformed fields of sensitivity. The preformed templates are the 'prior knowledge' put into vision. Each template response is definite and it contributes its own line-label, although its stimulus feature must have been fuzzy. The same process of specifying what combinations of low-level templates can exceed threshold takes place at the next higher level of templates, even though the mixture of low-level templates

was only approximately satisfactory. This process is repeated from one layered neuropile to the next and repeated in parallel, with a threshold at each level down to the motor neurones. The output actions are clear-cut and sharp, although the message is a wobbly transient at every level, and noise is prevented from evoking error by having many circuits in parallel. There is no reason why these principles should not apply to both man and bee.

Responses of the whole organism can depend on mechanisms which count large numbers of template responses in parallel to obtain reliable detailed information, but which can also respond to a single template response that threatens survival, or any intermediate between these extremes. Each different insect species has its own selection of simple templates which are counted by higher-level neurones and diversity of visual behaviour is presented to us by a diversity of insects. The mammalian cortex appears to have an additional mechanism which makes its own templates by flexible circuitry.

## The Wider Context of Adaptive Low-Level Vision

Studies of visual behaviour and eye structure in a variety of invertebrates show that to a large degree visual systems are special-purpose designs; but clearly in any one group of animals they have common features upon which a diversity of sophisticated and rapidly evolved examples are based. My guess is that in insects this basis is the set of high-resolution templates that go with the visual control of locomotion (Fig. 11.20). Given these, the rest can be done with low-resolution ratios followed by logical coincidences of particular ratios. Then, built upon this foundation is another layer of fewer, more specialized templates for biologically significant features, either genetically built in or assembled by learning for specific tasks. That was the basis of semivision.

At a low level of complexity, invertebrate visual systems illustrate very clearly that they process only the information necessary for their actions. If we extend this line of thought, we see that all visual systems evolve by natural selection in the context of the animal's activity and there will be no extra structure or process which is not needed for its specialized normal behaviour. Each of the templates in the visual processing mechanism represents a prior knowledge of just those visual tasks of selective importance. We are not used to thinking in this way because we have our own marvel of apparently universal vision 'before our open eyes'. When we see things they are already endowed with meaning based on memory; we have size constancy, colour constancy, visual illusions and hallucinations of visual images. Here we leave the realm of semivision.

### Addition of a Huge Memory with Immediate Access

Meaning in human vision or hearing depends on categories, which in turn depend on associations and a long history of learning. The process of mixing, filtering and abstracting pattern from sense data starts at the eye or ear. The recognition of categories is done with the aid of an enormous recall memory before the level of consciousness. This is the only way to explain how thoughts and sensory perceptions arrive in consciousness already coloured by everything which memory recalls for the occasion. The visual or auditory categorization is an unconscious action of the brain, learned in childhood, clearly cultural but also related to the empirical world against which it is continually rechecked. Although it is operationally successful and based on data and reality, much of what we humans think we see is a series of hallucinations generated by memory, in which the detail is regularly updated by eye movements, and we learn to see objects which are then given names and significance before they arrive in our conscious visual world. An excellent example is reading, as you are now doing. Clearly the idea of successive layers of templates which create a new visual world in their own domain is still useful for explanations of vision up to the highest level, but we are no longer concerned with on-line semivision.

In primitive eyes where the relation between complexity of processing and complexity of visual behaviour is more direct, we see that they cannot be expected to 'see' the visual world as we do, or 'see' the whole visual flow field as they move. Although an insect may have 360° vision, perhaps only the nearest relative motion, or one set of contrasts resembling a mate or prey, is of interest to it at any one time, and whatever is of interest has already been installed as the prior knowledge embodied in mechanisms of processing via Darwinian evolution of something like templates. In human vision the prior knowledge is installed by far more complex processes of maturation and learning.

### Conclusion

In the study of natural visual processing at a higher level, we find that prior knowledge has already been incorporated in a way that is impossible with mathematical operations on the stimulus alone. In the vernacular we say that you have to know the relevant features of the visual world before you can see what you are looking at. In man, these 'top-down' effects include long training in infancy, memory and rational inference, but in primitive vision the equally essential requirement for prior knowledge is met by the choice of processing structures that are the result of

Darwinian selection – what I have called the choice of templates. Therefore, to design a simple artificial visual system it is essential to start with what we want to see and the directions in which we expect contrasts to move at each point on the retina. Then we insert in parallel in repeated columns only as many simple templates as are required. The choice of templates represents the prior knowledge. In deciding what to do with the template responses, we count only sufficient combinations of them to generate the variety of outputs needed for the ensuing action and we count in fields that are as large as possible. Visual fixation helps by limiting the processing to one place on the map behind the retina. By copying natural vision in these ways it may be possible to overcome some of the problems of low-level artificial vision, but we will only begin to understand vision by drawing upon many different approaches involving many scientific disciplines.

### Acknowledgements

I would like to thank many who work or have worked on these and related questions in Canberra over the past few years while these ideas have been fermenting. The comprehensive theory has depended on research on the visual behaviour of the bee, much of it done by Miriam Lehrer, on the electrophysiology of insect optic lobes studied by Ljerka Marcelja over many years, and on computer modelling done recently by Peter Sobey. There are many others besides, working on insect visual processing. Especially, however, I would like to thank my colleague Mandyam Srinivasan for his interest and unfailing support, and for the ideas and critical comments that emerge from our discussions. Numerous drafts of the MS have been typed by Elizabeth Watson. The work on object vision in insects is supported by a grant from the Fujitsu Company of Japan, and by funds from the Centre for Visual Sciences and the Centre for Information Science in ANU.

### References

- (Works marked with an asterisk (\*) are general accounts or parts thereof.)
- \*Ali, M. A. (1984). *Photoreception and Vision in Invertebrates*. London and New York: Plenum.
  - Berger, E. W. (1900). Physiology and histology of the Cubomedusae, including Dr F. S. Conant's notes on the physiology. *Mem. Biol. Lab. Johns Hopkins Univ.*, 4 (4), 1–84.
  - \*Bullock, T. H. and Horridge, G. A. (1965). *Structure and Function in the Nervous Systems of Invertebrates*. San Francisco and London: Freeman.
  - Burrows, M. (1980). The control of sets of motoneurons by local interneurons in the locus. *J. Physiol. Lond.*, 298, 213–233.
  - Burt, E. T. and Catton, W. T. (1960). The properties of unit dis-



- charges in the optic lobe of the locust. *J. Physiol. Lond.*, **154**, 479–490.
- Collett, T. S. and Cartright, B. A. (1983). Eidetic images in insects: their role in navigation. *Trends Neurosci.*, **6**, 101–105.
- Crane, J. (1952). A comparative study of innate defensive behaviour in Trinidad mantids (*Orthoptera: Mantoidea*). *Zoologica*, **37**, 259–293.
- Dartnall, H. J. A. (1953). The interpretation of spectral sensitivity curves. *Br. Med. Bull.*, **9**, 24–30.
- \*De Voe, R. (1985). The eye: electrical activity. In: *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, vol. 4: *Nervous System: Sensory*. eds Kerkut, G. A. and Gilbert, L. I. pp. 277–354. Oxford: Pergamon.
- Erickson, R. P. (1963). Stimulus encoding in topographic and non-topographic modalities: on the significance of the activity of individual sensory neurones. *Psychol. Rev.*, **75**, 447–465.
- Erickson, R. P. (1982). The across-fiber pattern theory: an organizing principle for molar neural function. *Cont. Sens. Physiol.*, **6**, 79–110.
- Fischbach, K. F. and Dittrich, A. P. M. (1989). The optic lobe of *Drosophila melanogaster*. I. A Golgi analysis of wild-type structure. *Cell Tiss. Res.*, **258**, 441–475.
- \*Fraenkel, G. S. and Gunn, D. O. (1940). *The Orientation of Animals*. Oxford: Clarendon.
- \*Franceschini, N., Riehle, A. and Le Nestour, A. (1989). Directionally-selective motion detection by insect neurones. In *Facts of Vision*. eds Stavenga, D. and Hardie, R. pp. 360–390. Berlin: Springer.
- Gould, J. L. (1985). How bees remember flower shapes. *Science Wash. DC*, **227**, 1492–1494.
- Gould, J. L. (1986). Pattern learning by honeybees. *Animal Behav.*, **34**, 991–997.
- Gregory, R. L. (1967). Origin of eyes and brains. *Nature Lond.*, **213**, 369–372.
- Hassenstein, B. (1959). A cross-correlation process in the nervous center of an insect eye. *Nuovo Cimento*, **X13**, 617–619.
- Hausen, K. and Egelhaaf, M. (1989). Neural mechanisms of visual color control in insects. *Facts of Vision*. eds Stavenga, D. and Hardie, R. pp. 391–424. Berlin: Springer.
- Hertel, H. (1980). Chromatic properties of identified interneurons in the optic lobes of the bee. *J. Comp. Physiol.*, **137**, 215–231.
- Honegger, H. W. (1978). Sustained and transient responding units in the medulla of the cricket *Gryllus campestris*. *J. Comp. Physiol.*, **125**, 259–266.
- \*Horridge, G. A. (1978a). The separation of visual axes in apposition compound eyes. *Phil. Trans. R. Soc. Lond. B*, **285**, 1–59.
- \*Horridge, G. A. (1978b). A different kind of vision: the compound eye. In: *Handbook of Perception*, vol. 8. eds Carterette, E. C. and Friedman, M. P. pp. 3–82. London: Academic.
- Horridge, G. A. and Marcelja, L. (1990a). Responses of the H1 neurone to jumped edges. *Phil. Trans. R. Soc. Lond. B*, **329**, 65–73.
- Horridge, G. A. and Marcelja, L. (1990b). Responses of the H1 neurone to contrast and moving bars. *Phil. Trans. R. Soc. Lond.*, **329**, 75–80.
- Jander, R. and Volk-Heinrichs, I. (1970). Das strauschspezifische visuel Perceptorsystem der Stabheuschrecke (*Carausius morosus*). *Z. vergl. Physiol.*, **70**, 425–477.
- \*Kien, J. (1975). Motion detectors in locusts and grasshoppers. In: *The Compound Eye and Vision of Insects*. eds Horridge, G. A. pp. 410–422. Oxford: Clarendon.
- Kirchner, W. H. and Srinivasan, M. V. (1989). Freely flying honeybees use image motion to estimate object distance. *Naturwissen.*, **76**, 281–282.
- \*Kirschfeld, K. (1976). The resolution of lens and compound eyes. In: *Neural Principles in Vision*. eds Zettler, F. and Weiler, R. pp. 354–370. Heidelberg: Springer.
- \*Laughlin, S. B. (1981). Neural principles in the visual system. In: *Handbook of Sensory Physiology*, vol. VII/6B: *Invertebrate Visual Centers and Behavior*. pp. 133–280. Heidelberg: Springer.
- Laughlin, S. B. (1987). Form and function in retinal processing. *Trends Neurosci.*, **10**, 478–483.
- Lehrer, M., Srinivasan, M. V., Zhang, S. W. and Horridge, G. A. (1988). Motion cues provide the bee's visual world with a third dimension. *Nature Lond.*, **332**, 356–357.
- Lehrer, M., Srinivasan, M. V. and Zhang, S. W. (1990). Visual edge detection in the honeybee and its chromatic properties. *Proc. R. Soc. Lond. B*, **238**, 321–330.
- Maldonado, H. (1970). The deimatic reaction in the praying mantis *Stagmatoptera biocellata*. *Z. Vergl. Physiol.*, **68**, 60–71.
- Mimura, K. (1972). Neural mechanisms subserving directional selectivity of movement in the optic lobe of the fly. *J. Comp. Physiol.*, **80**, 409–437.
- Oldberg, R. M. (1986). Identified target-selective visual interneurons descending from the dragonfly brain. *J. Comp. Physiol.*, **159**, 827–840.
- Osorio, D. (1986). Ultraviolet sensitivity and spectral opponency in the locust. *J. Exp. Biol.*, **122**, 193–208.
- Osorio, D. (1987). The temporal properties of non-linear transient cells of the locust medulla. *J. Comp. Physiol.*, **161**, 431–440.
- Osorio, D. (1990). Inhibition, spatial pooling and signal rectification in the sub-units of non-directional motion-sensitive cells in the locust optic lobe. *J. Neuroscience*, submitted.
- Pinter, R. B. (1979). Inhibition and excitation in the locust DCMD receptive field; spatial frequency, spatial and temporal characteristics. *J. Exp. Biol.*, **80**, 191–216.
- \*Plate, L. J. (1924). *Allgemeine Zoologie und Abstammungslehre*. Part 2: *Die Sinnesorgane der Tiere*. Jena: Gustav Fischer.
- \*Reichardt, W. (1961). Autocorrelation: a principle for evaluation of sensory information by the central nervous system. In: *Principles of Sensory Communication*. Rosenblith, W. A. pp. 303–317. New York: Wiley.
- Rind, F. C. (1987). Non-directional, movement-sensitive neurones of the locust optic lobe. *J. Comp. Physiol. A*, **161**, 477–494.
- Rind, F. C. (1990). A directionally selective motion-detecting neurone in the brain of the locust: physiological and morphological characterization. *J. Exp. Biol.*, **149**, 1–19.
- \*Rolls, E. T. (1987). Information representation, processing and storage in the brain: analysis at the single neurone level. In: *Neural and Molecular Bases of Learning*. eds Changeux, J. P. and Konishi, M. Chichester and New York: Wiley.
- Rossel, S. (1986). Binocular spatial localization in the praying mantis. *J. Exp. Biol.*, **120**, 265–281.
- Rowell, C. F. H. (1971). The orthopteran descending movement detector (DMD) neurones: a characterization and review. *Zeit. vergl. Physiol.*, **73**, 167–194.
- Sandini, G. and Tistarelli, M. (1990). Active tracking strategy for monocular depth inference over multiple frames. *IEEE Trans Patt. Anal. and Mach. Intel.*, **12**, 13–27.
- Schwind, R. (1978). Visual system of *Notonecta glauca*. A neurone sensitive to movement in the binocular visual field. *J. Comp. Physiol.*, **123**, 315–328.
- \*Snyder, A. W. (1975). Optical properties of invertebrate photoreceptors. In: *The Compound Eye and Vision of Insects*. ed. Horridge, G. A. pp. 179–235. Oxford: Oxford University Press.
- Snyder, A. W. (1979). The physics of vision in compound eyes. In: *Handbook of Sensory Physiology*, Vol VII/6A: *Vision in Invertebrates*, ed. Autrum, H. pp. 255–314. Heidelberg: Springer.
- Sobey, P. J. and Horridge, G. A. (1990). Implementation of the template model of vision. *Proc. R. Soc. Lond. B*, **240**, 211–229.
- Srinivasan, M. V., Lehrer, M. and Horridge, G. A. (1990). Visual figure-ground discrimination in the honeybee: the role of motion parallax on boundaries. *Proc. R. Soc. Lond. B*, **238**, 331–350.
- \*Strausfeld, N. J. (1989). Beneath the compound eye: neuroanatom-

- mical analysis and physiological correlates in the study of insect vision. In: *Facets of Vision*. eds. Stavenga, D. G. and Hardie, R. C. pp. 317–359.
- \*Strausfeld, N. J. and Nassel, D. R. (1981). Neuroarchitectures serving compound eyes in crustacea and insects. In: *Handbook of Sensory Physiology*, vol. VII/6B: *Vision in Invertebrates*. ed. Autrum, H. pp. 1–132. Berlin, Heidelberg: Springer.
- \*Tinbergen, N. (1951). *The Study of Instinct*. Oxford: Oxford University Press.
- Van Hateren, J. H., Srinivasan, M. V. and Wait, P. B. (1990). Pattern recognition in bee: orientation discrimination. *J. Comp. Physiol.*, in press.
- Van Praagh, J. P., Ribi, W., Wehrhahn, C. and Wittmann, D. (1980). Drone bees fixate the queen with the dorsal front part of their compound eyes. *J. Comp. Physiol. A*, **162**, 159–172.
- \*Wehner, R. (1981). Spatial vision in arthropods. In: *Handbook of Sensory Physiology*, vol. VII/6C: *Vision in Invertebrates*. ed. Autrum, H. pp. 287–616. Berlin, Heidelberg: Springer.
- Wehner, R. (1987). 'Matched filters' – neural models of the external world. *J. Comp. Physiol. A*, **161**, 511–531.
- \*Zeil, J., Nalbach, G. and Nalbach, H. O. (1989). Spatial vision in a flat world: optical and neural adaptations in arthropods. In: *Neurobiology of Sensory Systems*. eds. Singh, R. N. and Strausfeld, N. pp. 123–136. New York: Plenum.
- Zhang, S. W., Wang, X., Liu, Z. and Srinivasan, M. W. (1990). Visual tracking of moving targets by freely flying honeybees. *Vis. Neurosci.* **4**, 379–386.