Why can't the eye see better?

H. B. BARLOW

Some people say that we hardly understand how the eye works at all, so it is perverse and premature to ask why it does not work better. I am a great admirer of this direct approach, but it is incomplete and liable to give unsatisfactory answers. Everybody must at some stage have dismantled a clock to find out 'how it works', but if this was your only question your answer was probably misleading. Thus, if it was an old-fashioned lever-escapement clock you probably found the mainspring and the chain of gears and said to yourself, 'Aha! the spring drives the hands through these gears'. That is quite right but what are the other gears and the escapement for? 'Oh they just slow it down', you probably said to yourself, perhaps with an uneasy feeling that there was more to it than that. There is, of course, more to it than that, for the whole evolution of timepieces has centred round the escapement, persuading it to divide up time into more exactly equal segments regardless of perturbing influences. You reach this heart of the matter, not just by asking, 'How does it work?', but by asking, 'Why doesn't it keep better time?'. The visual system is also complex; it performs a difficult task and we need to identify the difficulties before we can understand much about the system.

In this chapter I shall follow the same general plan as in an article I wrote 20 years ago on the physical limits of vision (Barlow, 1964b), but I shall deal only briefly with the parts that have stood the test of time and further experiment, shall correct some errors, and shall add material on new limiting factors, not
necessarily physical ones, that have become evident since that time.

What is needed for seeing?

Seeing is getting information about the environment from the light that enters our eyes. It is convenient to divide up this operation into three steps. The first step is to split up the entering light so that the amounts coming from different parts of the field, or from different parts of the spectrum, are absorbed preferentially in different structures. The next step is to generate signals dependent upon the amounts of light in each of these structures, and to transmit these signals to wherever the information is needed. This has to be done continuously, so that the signals refer to the amount of light absorbed in the recent past, for speed of response is an important requirement and photoreceptors cannot integrate indefinitely over time as a photographic plate can; this amounts to the additional requirement that the light be split up temporally, as well as spatially and spectrally. The final step is to collect and compare the signals from different structures, sometimes using signals that occurred at earlier times, and to make decisions about what there is in the visual environment on the basis of these assembled signals.

In my 1964 article (Barlow, 1964a), the primary emphasis was on the second of these three steps, generating signals dependent on the amounts of light after it has been subdivided. It was argued that the quantity of light in each of the subdivisions was small because the eye, unlike most physical instruments, splits up the light in space, colour and time simultaneously, and cannot increase the amount available by integration along any of these dimensions. Because the quantities were small, noise from quantum fluctuations was likely to be important, and might indeed limit the extent to which splitting up was worthwhile. There are three problems with this argument. The first is that something other than quantum fluctuations becomes of dominant importance at levels of illumination only slightly above threshold; however, this does not completely invalidate the argument because, whatever its origin, this extra noise is likely to be more serious in small subdivisions than in large ones. The second objection is that the splitting up and the signalling do not necessarily occur serially; for instance the signals from a rod appear to be partially combined with those of its neighbours before its synaptic signal is generated (Fain, 1975; Fain, Gold & Dowling, 1976). This is certainly an interesting fact, but it does not effect the conceptual validity of the argument.

The third objection is more important: I overlooked the possibility that systems might integrate
1. Why can't the eye see better?

along one dimension while maintaining good resolution along another, and that parallel pathways might do this in different ways. This is certainly important; for instance, Mullen (1985) has shown that the colour system has poor spatial resolution and can therefore increase the amounts of light available by integrating in space (see Fig. 1.1), but this does not affect resolution based on luminance changes in the image, because this is mediated by a system operating in parallel. Similar considerations apply to the motion system (Burr & Ross, 1982). With these qualifications, the previous article seems to have been correct about the limits involved in generating these graded signals, but the evidence is now enormously improved and is worth summarizing in the next sections.

Generating graded signals

The psychophysical evidence 20 years ago suggested that rods responded to single quanta (Hecht, Shlaer & Picard, 1942), that equivalent signals occurred in the dark (Barlow, 1956), possibly from thermal isomerization of rhodopsin, and that the rate of such noise events was increased following lights that bleached small quantities of the photopigment, thereby elevating the threshold (Barlow, 1964a). Elegant experiments on photoreceptors have confirmed these conclusions. Fig. 1.2 shows directly that rods give easily detected responses to single quanta (Baylor, Lamb & Yau, 1979). Furthermore, similar responses occur (Baylor, Matthews & Yau, 1980) at a very slow rate in total darkness and there is an increased rate of occurrence of these spurious signals (Lamb, 1980) early in dark adaptation (Fig. 1.3); however, this is unlikely to be the only cause of the elevation of threshold (Barlow, 1972). It was also thought that the lower sensitivity of cones might be attributed to the higher rate of thermal isomerization of their red-shifted pigments (Barlow, 1957). Quantal fluctuations cannot be detected in cones because another source of noise usually masks them; it is not known whether this is caused by thermal isomerization, but it is at least a possibility (Simon & Lamb, 1977). The situation in blue-sensitive cones is not clear; being shifted to the blue compared with rods, one might expect their pigments to be more stable thermally.

Little progress has been made in finding why measurements of quantum detection efficiency show that the visual system performs further and further from the quantal fluctuation limit at the luminance level increases (Barlow, 1962). Psychophysical measurements of quantum efficiency based on the desensitizing effect of adding noise to the visual stimulus lead to much higher values than those based on detection (Pelli, 1983), but low values of detective efficiency have been reported at the level of retinal ganglion cells in the cat (Levick et al., 1983). The quantum efficiency of

Fig. 1.2. Responses of a rod to the absorption of single photons. Forty flashes of an intensity that caused an average of 0.53 isomerizations of rhodopsin per flash were delivered just before the times indicated by dots. In some trials no isomerizations occurred, in others one, and in others two or more; the frequencies of these fitted the Poisson expectations. (From Baylor et al. 1979.)

Fig. 1.3. The top two traces show events, recorded as in Fig. 1.2, occurring at a slow rate in total darkness; Baylor et al. (1980) have shown that these are probably due to thermal isomerization of rhodopsin. The lowest trace was taken 3 min after a bleach calculated to isomerize 9 × 10^6 rhodopsin molecules in the rod (about 0.4%); the increased fluctuations are thought to be caused by an increase in the rate of the spontaneous events (from Lamb, 1980).
receptors when signalling suprathreshold light levels has not yet been measured.

Limitations to splitting up the light

In vertebrate eyes the light is split up according to (1) the position in the visual field from which it comes (i.e. its direction of entry), (2) its spectral composition, and (3) its time of arrival. There have been interesting developments in understanding the first and second of these, though the limits on the third are not fully understood.

Direction of entry

The ability to separate light according to its direction of entry determines the spatial resolution or acuity of an animal’s vision. In vertebrate eyes, as opposed to the compound eyes of arthropods which will not be considered further in this chapter, splitting according to direction of entry is achieved by forming an image on an array of receptor cells. The quality of this image is one obvious limiting factor, but it has quite recently become clear that there is another less obvious one, namely the ability of a receptor to restrict its sensitivity to the light that falls directly upon it. Let us first consider the image itself.

Image quality can be expressed in the form of the line spread function, which simply shows the distribution of light in the image of a very thin line. Alternatively, nearly the same information can be expressed in terms of the modulation transfer function (MTF), which shows the extent to which spatial sinusoids of different frequencies are demodulated, or lose contrast, in the image. Direct estimates of human image quality have been made by Krauskopf (1962), Weisheimer & Campbell (1962) and Campbell & Gubisch (1969), using the technique of analysing the light reflected from the fundus and emerging from the eye. Examples of these for the human eye are shown in Figs 1.4 and 1.5, which show the best line spreads and MTFs achievable at three different pupil diameters.

Fig. 1.4. Line spread functions of intact human eyes at four pupil diameters obtained from the light reflected from the fundus through the pupil and corrected for double passage through the optics. The line inner curves show the expected performance with diffraction-limited optics, while the points and the smooched denser curves show the actual performance. Errors would broaden the curves, but there is very little room for improvement in the results for the 2.0-mm pupil; with larger pupils the eye performs substantially below the diffraction limit. (From Campbell & Gubisch, 1966.)
A factor which causes an inevitable loss of high spatial frequencies, and thus sets a physical limit to acuity, is diffraction at the pupil aperture. The cut-off, or highest spatial frequency, that can pass the pupil and be represented in the image, is simply $d/\lambda$ cycles/radian, where $d$ and $\lambda$ are, respectively, the pupil diameter and the wavelength of the light. How closely do actual eyes approach this limit? A pupil diameter of 2.5 mm has a cut-off frequency at 78 cycles/degree for light of wavelength 550 nm; this is approached closely by human grating resolution in bright lights and also by Australian birds of prey (Reynolds, 1984). One is led to entertain the hypothesis that image quality is the sole determinant of human resolution, and that it is close to the diffraction limit.

Figs. 1.4 and 1.5 show the best estimates of the actual line spread functions and MTFs, compared with their theoretical limits. The actual spreads are not of course as narrow as the optima, nor do the MTFs lie on the theoretical curves, but they are quite close to them provided that the pupil diameter is small; at pupil diameters larger than 2 mm, the diffraction-limited line spread functions narrow and the cut-off spatial frequencies increase, but the actual line spread functions broaden and spatial resolution gets worse.

What about the first part of the hypothesis, that image quality is the main determinant of resolution? The comparison made in Fig. 1.6 shows that this hypothesis is wrong, for if psychophysical contrast sensitivity is limited by image quality at 10 cycles/degree, it falls below what image quality would allow at both higher and lower frequencies. There must be other limiting factors to account for the failure of the retina and postretinal mechanisms to make use of all the information available in the retinal image. At high frequencies, the probable factors are the difficulty of confining light within a single receptor, and the losses engendered by having too few sampling stations with large receptive fields that attenuate high frequencies. The latter factors must be responsible for the greatly reduced sensitivity to high spatial frequencies in the parafovea and periphery of the retina, but at the fovea centre the difficulty of confining light to a single receptor is the most likely candidate and is considered next. The reduced sensitivity at low frequencies probably results from lateral inhibition, which will be considered later.

Fig. 1.5. Modulation transfer functions of intact human optics obtained by the same method as in Fig. 1.4. For each pupil diameter, the spatial frequencies have been normalized by dividing by the diffraction limit for that diameter. (From Campbell & Gubich, 1966.)

Fig. 1.6. The continuous inner curve shows the contrast sensitivity of the human eye at high luminance levels plotted on double logarithmic coordinates. The dotted curve touching this at 10 cycles/degree is the MTF of the human eye. Clearly, the contrast sensitivity at higher and lower frequencies must be degraded by factors other than the quality of the image. (From Barlow & Mallet, 1982.)
Receptor resolution

In 1964 I thought that chromatic aberration was also important in setting a limit, since it prevented acuity being improved by enlarging lens aperture. There may still be something in this, but Enoch (1961) has pointed out that light must be confined within one receptor to prevent blurring within the retina, and Snyder (1972) has shown how receptors must cease to act as effective waveguides if they are below about 2 μm in diameter. Perhaps the most convincing indication that receptor size is important came from the examination of the eyes of a range of hawks done by Snyder & Miller (1977). The overall size of these eyes varied greatly, and in general every structure within them was scaled precisely in accordance with the overall size; but the foveal receptors were close to 2 μm in diameter and did not vary at all; the presumption is that they were already at the smallest effective size, even in the largest eye. If chromatic aberration had been the limiting factor, receptor diameter should have decreased in proportion to the square root of the eye's focal length, which was certainly not the case. The fact that receptor diameter, not chromatic aberration, sets the limit to acuity has important implications for the expected effect of eye size on acuity (see Table 1.1 below).

Why is it so important for the eye to split up light as finely as possible according to direction of entry? One can see a great deal, even with a very blurred image, and in some cases blurring actually aids visibility (Harmon & Julesz, 1973). But everything that can be seen in a blurred image could in principle be detected by performing the appropriate spatial

Table 1.1. The effect of eye size

<table>
<thead>
<tr>
<th>Type of performance</th>
<th>Limiting factor</th>
<th>Measure used</th>
<th>Expected value</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acuity</td>
<td>Diffraction at pupil</td>
<td>Grating cycles per radian</td>
<td>$d$</td>
<td>$d$</td>
</tr>
<tr>
<td>Absolute sensitivity</td>
<td>Retinal noise</td>
<td>$1$ min. no. of quanta</td>
<td>$\propto D^2 / F$</td>
<td>$1.5 D^3 / F$</td>
</tr>
<tr>
<td>Acuity and speed at low luminance</td>
<td>Quanta fluctuations</td>
<td>Cycles per radian or flicker fusion frequency</td>
<td>$\propto D$</td>
<td>$D / 5$</td>
</tr>
</tbody>
</table>

$D =$ maximum pupil diameter; $d =$ minimum pupil diameter; $\lambda =$ wavelength; $F =$ posterior nodal distance.
summatin operation on a sharp image, while the reverse is not the case. Splitting up as finely as possible therefore gives vision its versatility.

One way to make this clear is to calculate the amount of information per unit waveband, using Shannon's measure, at different spatial frequencies. Such calculations are plotted in Fig. 1.7, and these curves show immediately which spatial frequencies carry most information, while the areas under the curves show the total amounts of information available. The top curve is for a diffraction-limited image at 1000 cd/m²; information about polarization is not used, and spectral information has been neglected. The next curve is for 100 cd/m² luminance; the curve is lower because the smaller number of quanta reduces the highest attainable contrast sensitivity. The third curve down represents an estimate made for the retinal image with a 2.4 mm pupil, as described by Campbell & Gabisch (1966), and the lowest curve is an estimate made for the contrast sensitivity function in bright light (from Campbell & Robson, 1968). The points to notice are as follows: (1) reducing light intensity by a factor of 10 has only a moderate effect, for it only reduces the dynamic range by \(\sqrt{10}\), and information is proportional to the log of the dynamic range; (2) the eye’s optics lose a lot of high-frequency information but preserve quite a large proportion of the low-frequency information that passes the 2.4 mm pupil; (3) the low-frequency cut shown in Fig. 1.6 has a negligible effect on the total information, because the amount available per unit waveband decreases so rapidly with spatial frequency; and (4) in contrast, the loss of high frequencies has a large effect in reducing the total amount that is available centrally. It is presumably the fact that the light entering the eye contains so much information at higher spatial frequencies that has limited the evolution of acuities almost up to the diffraction limit.

**Colour**

Within a central circle 1 degree in diameter, the entering light is split into almost 10000 spatial subdivisions, whereas the trichromacy of colour vision implies that it is split into only three spectral subdivisions. The eye is obviously much less interested in colour than in spatial position and it is worth asking, 'Why not split the light into four or more spectral subdivisions?'

Fig. 1.8d shows the Fourier transforms of the spectral sensitivities of the red, green and blue components of the human colour system, using the derivations of Smith & Pokorny (1975). The physical significance of these curves can be visualized as follows (Barlow, 1982). Imagine a suitably dispersed spectrum formed, for instance, by a prism. Now place in this spectrum a comb with teeth having a sinusoidal profile, and recombine the spectrum with a lens. If the sinusoidal comb had a low frequency, one tooth might block the blue and green while the yellow and red passed through the neighbouring gaps and the recombined light would obviously look orange; shifting the position of the comb would reverse what was blocked and passed, and the recombined light would change to greenish blue. If, on the other hand, the comb was of high frequency or short wavelength, the teeth would obstruct some light of all colours, and the hue of the recombined light would be almost unaffected, whatever the position of the comb. We now see the relevance of Fig. 1.8d, for it shows how individual colour mechanisms would respond to comb-filtered spectra of various comb-frequencies. The ordinate shows the predicted percentage modulation of each system's output as the comb (assumed to pass 100% at the peaks in the gaps and to pass 0% at the troughs caused by the teeth) was moved through the spectrum. The scale of the abscissa is unfamiliar; recall that human colour vision is good only over the range 435 to 650 nm, which corresponds to 689 to 461 THz (Terahertz; 1 THz = 10¹²Hz) or a bandwidth of 228 THz. The abscissa gives cycles/1000 THz bandwidth, so to obtain the number of cycles of the comb within the range of good human colour discrimination one should divide the abscissa value by 4.4. It now becomes clear that the spectral sensitivities of the human colour mechanisms are so broad that they attenuate comb-filtered spectra by 40% even when there is only one complete cycle of the comb within the range of good colour discrimination; for higher comb-frequencies even more of the modulation is lost.

It is a well-known proposition that only \((2N + 1)\) sample points are needed to obtain all the information from a segment of waveform containing \(\pi\) periods of the highest frequency present. We have seen that frequencies higher than 4.4 cycles/1000 THz bandwidth, corresponding to one cycle in the range of good colour discrimination, are demodulated more and more strongly, so the three channels of our trichromatic system look like a reasonable choice; more channels, corresponding to more sample points, would pick up
more information, but the higher comb-frequencies they might inform us about would be strongly demodulated by the broad spectral sensitivity curves. That seems a plausible justification for trichromacy, but the argument has a corollary. If the spectral sensitivity curves could be narrowed, then high comb-frequencies would be less demodulated and tetrachromacy would be a better proposition. The dotted curve in Fig. 1.8d shows the Fourier transform of the spectral sensitivity for a chicken cone containing a coloured oil droplet, calculated by Bowmaker & Knowles (1977). The principal effect of the droplet is to narrow the sensitivity curve, thus enabling it to respond to higher comb-frequencies. Fig. 1.8c shows how a narrowed red-sensitivity curve might be fitted in with the other three human mechanisms. Liebman (1972) has suggested that turtles, which also have oil droplets in their receptors, have five spectral channels, droplets might therefore make tetrachromacy worthwhile, and (c) shows how such a narrowed curve might fit in with the three known human mechanisms. The roach has the four pigments shown in (b); these curves are no narrower than those of the human pigments, but they span a bigger spectral range; this is another way of making tetrachromacy worthwhile, though there seems to be no behavioural evidence of tetrachromacy in any animal. (From Bowmaker, 1983.)

![Graphs showing spectral energy distribution and Fourier transforms](image-url)
though I know of no behavioural demonstration of more than three channels in any animal.

Another case where tetrachromacy might be worthwhile is in animals which have a greater spectral range of useful vision than human beings. Bowmaker (1983) has evidence that this may occur, for he has found that the roach has the four different cone pigments shown in Fig. 1.8b; each of these has a receptor band as narrow as that of the human pigment, but the positions of their peaks range from 355 to 620 nm, compared with the human extremes of 430 and 565 nm. Four different cone pigments in one retina do not prove tetrachromacy, but they certainly prompt one to look for behavioural evidence.

Another possible way of making polychromacy worthwhile would be to narrow the absorption spectra of the receptor photopigments. Little seems to be known about possible limiting factors here, but an interesting fact has recently come to light. If the absorption spectra of pigments with peaks at different positions are plotted on a scale of $(\lambda)^{1/3}$, then they become very nearly superimposable by a simple lateral shift (Barlow, 1982). This implies that the breadth of an absorption spectrum of this class of pigment varies with the $1/4$-power of the wavelength of its peak absorption. The Vitamin A2-based photopigments appear to follow the same rule, but their absorption spectra are a little broader than those of Vitamin A1 pigments. It would be interesting to know the physical basis of these rules.

Can the contrast sensitivity function for combined spectra be measured psychophysically? Preliminary results (Barlow et al., 1983) suggest that it does not follow the form predicted by the Fourier transforms of Fig. 1.8. Low comb-frequencies are attenuated and there is a peak at intermediate frequencies; this is presumably a direct manifestation of colour opponency, just as the attenuation of low spatial frequencies is thought to be the result of spatial opponency. This will be discussed further in the section on collecting and comparing.

**Temporal factors**

Fig. 1.9 shows the responses of a monkey rod to brief flashes of light of increasing intensity (Nunn & Baylor, 1982). Notice that they are extremely slow, especially at low intensities. Cones are about three times faster, but it is still surprising to find psychophysical evidence that changes in the timing of events can be detected down to less than 200 μs (Burr & Ross, 1979). Faster receptors would seem to have obvious advantages for animals as well as sportsmen, so the question 'Why aren't they faster?' needs an answer.

The first point to make is that the prolonged responses to brief flashes shown in Fig. 1.9 provide a mechanism for temporal summation, which will improve the sensitivity of the eye to long-lasting stimuli. Anybody who has examined single frames of cinema film or video tape will appreciate how effective temporal averaging is in improving signal/noise ratios.

It is not the only possible mechanism for doing this, for if the receptors generated rapid brief pulses, there would be nothing to stop these being integrated temporally at some central site. This would have the advantage that it would be optional; a parallel system that required speed rather than high signal/noise ratios could take advantage of the receptors' rapidity. One should notice that the responses become faster for strong flashes, and especially at high adaptation levels, which fits in with the psychophysical evidence, but one is left with the feeling that faster receptors would be advantageous in much prey-catching and evasive behaviour. Lamb (1984) found that the toad rod...
kinetics had a temperature coefficient \( Q_m \) of 2.2, and R.A. Cone has suggested that the speed of response is related to receptor size. Perhaps fundamental limits to photoreceptor speed will emerge when the steps between the absorption of light and changes in membrane currents are better understood.

The effect of eye size
The best that an eye can achieve in splitting up the light is strongly dependent on its size, for a larger pupil catches more quanta and allows a higher diffraction-limited acuity. Increasing pupil diameter by increasing numerical aperture alone will improve sensitivity but will not improve acuity for two reasons: (1) chromatic aberration will get worse; and (2) if there was an improved retinal image, this would require smaller receptors but, as we have seen, receptors have a lower limit for their useful diameter at about 2 \( \mu m \) (Snyder, 1972). The Abbé formula for microscope resolution (minimum resolvable distance = \( 0.61 \lambda / n \sin \beta \)) suggests that to resolve 2 \( \mu m \) requires a numerical aperture \( n \) of 0.17, which implies that the pupil should subtend about 15 degrees at the receptors. It seems that there can be no selective pressure to increase numerical apertures much above this value for reasons of acuity, though it would be advantageous to do so to improve sensitivity.

Taking these limiting factors into account, one can draw some conclusions about the likely visual performance of an animal simply by observing the size of its pupil, as shown in Table 1.1. The diameter of the pupil in bright light \( (d) \) is the value relevant for the highest acuity attainable, and the fourth column gives the highest attainable spatial frequency in cycles/radian. One often wishes to compare an animal's performance with that of man, so this ratio is given in the fifth column: thus a hawk with a pupil diameter in bright light of 4 mm would be expected to have a grating acuity about twice that of mankind, whereas for a lizard with a pupil diameter of 0.5 mm the expected acuity would be only one-quarter that of man. Note that the expected acuity is almost certainly not reached in the many species of animal that are not specialized for high resolution: dogs and lions, for instance, are probably not as good as people, though they have large pupils and the table suggests that they might be better; the same must be true for the large herbivores like cows, which have enormous pupils.

For sensitivity, the diameter at low luminances \( (D) \) is relevant. A proportional sign is given because many specializations, such as reflecting tapeta, would influence sensitivity, and the body temperature would also be important where thermal decomposition is the main determinant of retinal noise. For simplicity these other factors have been neglected in the final column and it has been assumed that the proportionality factor is substantially the same as for man. Note that for the cat this column suggests a sensitivity some 10 times that of man, which is probably about right, but again the expected performance would not be reached in animals not well adapted to low luminances. The lowest row of the table is more tentative, for there are few conditions where vision is truly limited by quantal fluctuations. Nevertheless, this row indicates that animals with large pupils have scope for considerable improvement over people.

Collecting and deciding
Splitting up the light and producing graded signals for each of the subdivisions is a necessary prerequisite for making use of the light that reaches the eye, but it is only the first step towards finding out what is in the visual environment. The rest of the job is probably more difficult and is certainly less well understood. Furthermore, the limits of the next steps, as far as one can discern them, are not set by hard physical laws such as those governing quantal fluctuations, diffraction at the pupil, or the problem of holding light within the waveguide-like outer segment of a photoreceptor. Instead, they arise from the logic of the process of recognition and decision, and from the limited capabilities of neurons for performing the kinds of task that we can see are necessary. Let us consider the former first.

Logic of recognition
Our eyes identify familiar objects in the world around us and tell us where they are. Of course they also warn us of unfamiliar objects, guide our eye movements, tell us when to blink and allow us to admire sunsets, but object recognition is surely the basic skill and probably the most difficult one. Consideration of this topic usually starts with the invariance problem: the object has to be recognized in different three-dimensional orientations, over a range of positions and sizes, under varying conditions of illumination, against different backgrounds, and so forth. To meet this requirement it is thought that representations are
formed and stored using descriptors and coordinate systems such that the various confounding factors do not change the representation, or only change a small number of its parameters which can be ignored if required. This sounds plausible but difficult, and I'm not sure how much success has been achieved in finding and implementing such representations. But there is another aspect that also needs emphasizing.

If such a representation is to work well, then it must be complete and exclusive in the following sense. The quantities stored should be determined by all the important parts of the object represented, not just by some of them, and they should be determined only by the object, and not by the rest of the visual scene. Considering the range of forms an object can assume in its image, this is clearly a difficult requirement to meet, but its desirability is quite easy to see: if parts of the object do not influence the representation, then those parts could be removed or changed without affecting the representation, and this obviously will not do if the parts concerned are important; and if the rest of the visual scene can influence the representation, then it will be unnecessarily susceptible to outside factors.

The requirements for invariance, completeness and exclusiveness mean that object recognition is a horrendously complicated matter. To do the job right the system has got to implement just the right logical function of a very large number of input variables, and the number of possible logical functions from which it must pick the right one is unimaginably large. It is inconceivable that a brain can always find the right logical function under all circumstances, but we do not know at all what restrictions apply; all we can say is that the logical complexity of the object recognition problem is so great that it surely limits what the brain can do, and we should be aware of this fact.

It is more useful to look at the requirements for the lower-level descriptors out of which high-level representations are made, for here we know better how they must be implemented. I think we can already discern two serious shortcomings in the neural equipment available to do this, namely the limited dynamic range of nerve fibres as communication channels and the limited range of connections any one cell can establish with others. In both cases a sceptic might say, 'If this property is so important it must be under intense evolutionary pressure, so why has it not been improved?'. Perhaps they have been evolved as far as is allowed by some biophysical factors we do not know about, but a discussion of such possibilities would lead us too far from the point, and they will simply be taken as limits that cannot be improved upon.

To demonstrate their importance let us suppose that the system needs to know the numbers of quanta absorbed in blobs and patches of a great variety of sizes and shapes in a vast number of different positions and orientations in the visual field. Experience of image-processing tells us that obtaining this type of information is extraordinarily time-consuming, even on a modern computer. In such a machine the quantities would normally be represented in a word of at least 16 bits, whereas it is hard to believe (see below) that a nerve fibre can transmit more than about 6 bits, and it would require 100 ms to do even that. This is the reason for believing that the dynamic range of nerve fibres is an important limitation.

A computer suitable for image-processing would also have direct access to a wide range of addresses, perhaps using 20 bits or more for this purpose. The range of connectivity of an individual nerve cell limits how many other cells it receives information from and can transmit to, and it is hard to believe that this even approaches the 1 million implied by the 20-bit address range of the computer, hence the importance of connectivity as a limitation of neural computing. There are obvious problems with the analogies I have drawn, especially because of the parallel nature of neural computing, but the test of the case is to go ahead and look at the physiology of the visual pathway with these limitations in mind, and to see if they help to make sense of the arrangements we find. The limited dynamic range of nerve fibres seems to throw more light on retinal coding, whereas limited connectivity seems important for the cortex, so they will be considered separately, in that order.

Retinal coding and the dynamic range of nerves

We do not know if dynamic range is a problem in receptors, bipolars, horizontal and amacrine cells, but it must be serious for the ganglion cells. They characteristically pick up information from many receptors, and would require a greater dynamic range on that account, but they also have to signal by means of discrete all-or-nothing impulses rather than by graded potentials. It is hard to make a firm estimate of the dynamic range of nerve messages, mainly because
we are ignorant about the possible importance of the precise timing of individual impulses, but the generally accepted view is that the important variable is the number of impulses in some short interval of time. If that interval was as long as 100 ms, the maximum number of impulses would be about 64, so the dynamic range would be equivalent to 6 bits. If the interval was only 10 ms, the dynamic range of each message would be equivalent to 3 bits or less, but there would of course be 10 times as many messages. In any case the dynamic range seems extremely small, compared with the range of quantal absorption rates in the receptors that have to be signalled, for this range must regularly exceed $10^{12}$ or 40 bits. The problem posed by this discrepancy between the wide range of the input and the narrow range of the output makes sense of many features of the retina.

Adaptation

Craik (1938) looked upon adaptation as a mechanism for adjusting the response range of the retina in accordance with the prevailing light level, and it has subsequently been shown how this is done by those who have recorded the response of individual nerve cells (Werblin, 1971). It was the wide range of the input, combined with the high incremental sensitivity of the eye, that led Craik to this view, but the poor dynamic range of nerve fibres clearly makes adjustments of sensitivity even more important. Fig. 1.10 shows very clearly how the range of a ganglion cell's response shifts with adaptation level.

Spatial opponency and the on-off system

To use a limited output range effectively it is obviously advantageous to remove any constant component. Lateral inhibition can be regarded as a means of finding the local mean value of luminance and subtracting it from the signal to be transmitted, thus conserving the output range for the useful signal (Barlow & Levick, 1966). Nerve signals cannot assume negative values, so subtracting from the mean entails signalling the upward and downward deviations from the mean separately; however, this brings the additional advantage of doubling the available dynamic range.

Colour opponency

The extensive overlap of the spectral sensitivities of the different primary colour mechanisms means that signals in them are necessarily strongly correlated. Buchsbaum & Gottschalk (1983) have worked out the implications of this, and shown that the colour opponent arrangements that are found to approximate quite closely to those that would be recommended for optimal coding of the information. Their argument gains strength from the limited range of the output, for this tells one why optimal arrangements are so important.

Signalling probabilities

It was suggested earlier that, ideally, nerve signals should inform their destinations of the exact numbers of quanta absorbed in particular blobs and patches of the image, but this is not strictly true. The aim of most central information-processing must be to build up the probability of making a correct deduction about the environment, and for this purpose what is needed are probability statements about the constituent parts, such as 'This part of the image is unusually bright (or unusually dim, or unusually red or blue in spectral content).' In the retina the unusualness could be assessed from what has occurred locally in the image.
in the fairly recent past, and Fig. 1.10 shows the responses of an ideal device, signalling the log of the improbability of the Null hypothesis that there was no increase, or no decrease, in the quantum flux. It seems to me that the messages provided by retinal ganglion cells, as shown in Fig. 1.10, fit rather well the idea that they are signalling improbabilities (see Fig. 1.11) (Barlow, 1969).

There are two further comments to make about the limited dynamic range of the optic nerve fibres. First, it must be admitted that there are many aspects of retinal coding which it does not help us to understand, such as the distinction between X and Y units (Enroth-Cugell & Robson, 1966), or the reasons for coding more complex features such as direction of movement or oriented edges (Levick, 1967). Secondly, there is the interesting question of how this explanation of the selective advantage of adaptation, lateral inhibition, and colour opponency fits in with other explanations. It is sometimes suggested that these mechanisms exist because they are useful for the performance of some higher-level task, such as detecting edges or aiding image segregation, but this is a tautology: higher mechanisms have no alternative but to use the information provided by the retina. What one needs is an explanation in terms of image properties; the redundancy reduction hypothesis (Barlow, 1961, 1985) and the recent ideas of Srinivasan, Laughlin & Dubbs (1982) on predictive coding may provide this.

Fig. 1.11. Idealized ON and OFF units that signal \(-\log p\), where \(p\) is the probability of the number of quanta shown on the abscissa being absorbed when the mean rate of absorption is the adaptation level: it is supposed that, as in Fig. 1.10, the retina is adapted to three different levels, then stepped to another level. These curves have points in common with the experimental results shown in Fig. 1.10. (From Barlow, 1969.)

Attaching importance to the limited dynamic range of optic nerve fibres certainly does not conflict with these ideas, for it gives an added reason for the importance of predictive coding or redundancy reduction. These other ideas may, however, add a broader perspective to the problem.

Cortex and connectivity

The role of the cortex must be to derive knowledge of objects in the environment from the image it receives in the optic radiation from the lateral geniculate nucleus. Cortical neurons receive connections on their spines, dendrites, and cell bodies in one or other restricted region of the cortex, and transmit impulses down their axons to other parts of the brain, often to other restricted parts of the cortex. Because of their diversity it is a bit absurd to talk of a typical cortical neuron, but the above description is so general that no one could quarrel with it seriously, and the important point to note is that both the neuron's sources of information and the destinations of its axon are restricted, when compared with the address range available in an image-processing computer. A cortical neuron may be unusual in being able to interconnect a very large number of other cells, but this large number is only a very small fraction of all the cells in the brain which it might be advantageous for it to connect with. As an exercise I shall assume that this limited capacity to interconnect is an overriding factor in the design of the cortex, and make some predictions about how it would have to set about doing its job if that were the case.

In discussing object recognition it has already been emphasized that the information does not arrive in neatly delineated orderly packages, but is liable to be disordered by changes of perspective, size, illumination, background, and so on. The amazing thing about perception is that, in spite of these disorganizing factors, it gives us an organized and orderly view of the objects around us; hence what the nerve cells in the visual system must do above all else is to find order and preserve it. The first requirement must surely be to preserve whatever order is already to be found in the image and prevent things that belong together being further separated; hence the requirement for an accurate topographical map in VI, for although information about objects is not well segregated in the image, there is certainly a higher prior probability of neighbouring points belonging to the same object.
than would be the case if the image had been fragmented by a disordered projection.

Since the cortex has to deal with objects, and since its neurons pick up information from regions of only a few square millimetres, it must find ways of bringing together all the information relevant to an object within small cortical areas. We have seen that neighbouring points in the image are likely to belong to the same object, but they are not certain to, because objects have edges and holes or may be partially obscured by other objects. In the same way, regions of the image of similar colour, texture, disparity, or direction of motion do not necessarily belong to the same object, but they are more likely than area regions of the image that differ in these characteristics. Hence, in its efforts to organize the image into a representation of objects, the cortex needs first to detect these characteristics, because they are the local features which can link together the parts of the image that belong to the same object (Barlow, 1981).

This has been presented as a prediction from the limited ability of neurons to interconnect with other neurons, but I think it fits what we know of the physiology of V1, as well as much previous work on the psychological factors that cause segregation of the image into figure and ground. I will not repeat the evidence that neurons in V1 are selective for colour, motion and disparity, but the role of the simple cells as orientation or texture detectors needs amplifying. The orientation selectivity of cortical neurons was the aspect to which Hubel & Wiesel (1962) originally gave most emphasis, and it certainly seems to play a very important part in the organization of the primary visual cortex, yet at first sight it may not be obvious that orientation selectivity plays a key role in image segregation. This was, however, the first characteristic to be recognized as important in artificial intelligence. Guzman (1968), working on the problems of deriving the three-dimensional form and position of a number of blocks from a line drawing of them, came across the problem of identifying the two parts of a block whose contour had been obscured by another one. The co-linearity of the two separated parts of the edges was what he called the 'linking feature' that enabled them to be reunited in the final representation; it was therefore the orientation and position of the edge-segments that enabled this to be done. Notice that in a computer program the importance of a linking feature is that it enables parts to be reunited after a limited, instead of an exhaustive, search; if the address range of the computer was small, this would obviously be even more important, and that is the situation in the neural computer, whose elements can only interconnect a tiny proportion of the other cells in the brain.

A more up-to-date view of cortical neurons regards them as providing a representation of a local part of the image in terms of spatial frequencies (Gieizer, Ivanoff & Tscherbach, 1973; Robson, 1975). It has long been recognized that cortical neurons vary in the size as well as the orientation of the stimulus they prefer, and this idea of local Fourier-type analysis incorporates this fact into a systematic model. How does this fit in? Because of its limited connectivity, a given neuron can only respond to properties of a small patch of the image, yet the role of a group of such neurons must be to provide a description of the local pattern of light and shade in that small patch that will be helpful, together with the descriptions of other patches, in recognizing objects. There are innumerable possible ways of describing such patches, but one in terms of spatial frequency has merit for several reasons. First, it is potentially a complete description of the patch (Sakitt & Barlow, 1982). Secondly, each element, which we suppose signals the coefficient of one of the frequency and orientation components, is derived from the whole of the patch so that all such messages have nearly equal signal/noise ratios. Thirdly, the components have generality so that the similarity or dissimilarity of descriptions from different parts can be compared. Finally, the search for similarities can be much simplified by proper arrangement of the destinations of the information about each component. For example, Marr & Hildreth's (1980) program for edge detection uses knowledge of edges of different degrees of blur but similar orientation, and it would be possible to bring this information together by the appropriate targeting of the destinations of the simple cells in the primary cortex. It should be added that information from continuations of the edge in either direction would also be useful, and this too could be achieved by the appropriate pattern of connections from the primary cortex. We obviously do not know whether or how this is done, but I think recognizing the connectivity problem makes us look at the operation of the visual cortex in a new light.

The next prediction is that V1 will not be the only visual area (Barlow, 1981): multiple visual areas will be necessary, and have of course been found
In these secondary and tertiary areas, the magnification factor should be less if there is a well-defined topographical map, so that one cell can look at a larger region of the visual field; furthermore, one would expect to find evidence of information being brought together on non-topographical principles, for example according to orientation or colour or direction of motion. It is too early to say how far this prediction is correct, but some of the evidence is promising (Zeki, 1980).

It is apparent that what has so far been described will not take the cortex very far towards object recognition. The main hindrance is that each of the cues for segregating the information about the object from irrelevant information is uncertain and unreliable. An object cannot be reliably defined by its position in the image, by its texture, by its colour, by its direction of motion, or by its disparity. To have any chance of working, the system must use all these cues in combination, but the combination required will not be fixed: sometimes colour will be reliable by itself, sometimes it will not help, and so on for the other possible linking characteristics. It is crucially important to take account of the reliability of each message when combining them, if the messages have differing degrees of reliability, so it is highly desirable for each message to convey its own reliability. This is the reason why the signalling of improbabilities that was suggested for the retina seems an even better arrangement in the cortex. If the cortex uses such a probabilistic language, one can perhaps see why it is so adept at combining weak cues and recognizing which are strong cues.

Summary and conclusions

In the early stages of vision, physical factors set firm limits to what can be achieved, and the eye seems to approach them quite closely. The evolution of the eye has clearly been moulded to a great extent by these factors, the most important of which appear to be: diffraction as a limit to acuity; quantum fluctuations and the thermal stability of photopigments as limits to sensitivity; propagation down the waveguide-like outer segment of photoreceptors as a limit to their size and separation; and the narrowness of spectral sensitivity curves as a limit to the number of worthwhile colour channels. There are probably other important limiting factors lying behind the loss of quantum efficiency at high background levels, and the slowness of the receptor's response.

It is much harder to spot the limiting factors for the higher functions of vision, but it is certainly important to realize the tremendous difficulty of the task of object recognition set by the requirements for invariance in the face of perturbing factors, and for completeness and exclusiveness in the representation of the objects to be recognized. There are also two characteristics of neurons as computing elements that must create difficulties, namely the limited dynamic range of neural messages, and the limited range of connectivity of any given nerve cell.

Asking 'Why doesn't the eye see better?' will not answer all problems of vision, but it does direct our attention to interesting aspects, and it may help us to understand why the visual system is organized in the way it is.

References

Section I: The retina